Validity Testing in Child and Adolescent Assessment

Evaluating Exaggeration, Feigning, and Noncredible Effort

edited by
Michael W. Kirkwood

ebook

THE GUILFORD PRESS

VALIDITY TESTING
IN CHILD AND ADOLESCENT ASSESSMENT

EVIDENCE-BASED PRACTICE IN NEUROPSYCHOLOGY

Kyle Brauer Boone, *Series Editor*

Clinical Practice of Forensic Neuropsychology: An Evidence-Based Approach
Kyle Brauer Boone

Psychological Assessment: A Problem-Solving Approach
Julie A. Suhr

Validity Testing in Child and Adolescent Assessment: Evaluating Exaggeration, Feigning, and Noncredible Effort
Michael W. Kirkwood

Neuropsychological Report Writing
Jacobus Donders

Validity Testing in Child and Adolescent Assessment

Evaluating Exaggeration, Feigning, and Noncredible Effort

edited by
MICHAEL W. KIRKWOOD

Series Editor's Note by Kyle Brauer Boone

THE GUILFORD PRESS
New York London

A Division of Guilford Publications, Inc.
370 Seventh Avenue, Suite 1200, New York, NY 10001
www.guilford.com

Printed in the United States of America

This book is printed on acid-free paper.

Last digit is print number: 9 8 7 6 5 4 3 2 1

Library of Congress Cataloging-in-Publication Data is available from the publisher.

ISBN 978-1-4625-2185-2

To Jen—your love and support keep me going

To Finn and Jack—your zeal and joy for life
provide delight and perspective in mine

About the Editor

Michael W. Kirkwood, PhD, ABPP-CN, is Associate Clinical Professor in the Department of Physical Medicine and Rehabilitation at the University of Colorado School of Medicine. He is also Director of the Psychology/Neuropsychology Program in the Department of Rehabilitation Medicine at Children's Hospital Colorado and founder and Codirector of the hospital's Concussion Program. Dr. Kirkwood is board certified in Clinical Neuropsychology and board certified as a subspecialist in Pediatric Neuropsychology by the American Board of Professional Psychology, and he has served on multiple boards of directors and expert committees. He has conducted extensive research on traumatic brain injury and pediatric validity testing in children and teens, has lectured widely, and has published more than 50 journal articles and 10 book chapters. He is coeditor of the books *Mild Traumatic Brain Injury in Children and Adolescents* and *Clinical Neuropsychology Study Guide and Board Review* and coauthor of the book *Board Certification in Clinical Neuropsychology.*

Contributors

David A. Baker, PsyD, Department of Physical Medicine and Rehabilitation, University of Colorado School of Medicine and Children's Hospital Colorado, Aurora, Colorado

Brian L. Brooks, PhD, Neurosciences Program (Brain Injury and Rehabilitation), Alberta Children's Hospital, and Departments of Paediatrics, Clinical Neurosciences, and Psychology, University of Calgary, Calgary, Alberta, Canada

Dominic A. Carone, PhD, ABPP-CN, Neuropsychology Assessment Program, SUNY Upstate Medical University, Syracuse, New York

Michael D. Chafetz, PhD, ABPP-CN, Algiers Neurobehavioral Resource, LLC, New Orleans, Louisiana

Amy K. Connery, PsyD, ABPP-CN, Department of Physical Medicine and Rehabilitation, University of Colorado School of Medicine and Children's Hospital Colorado, Aurora, Colorado

Jacobus Donders, PhD, ABPP-CN, Mary Free Bed Rehabilitation Hospital, Grand Rapids, Michigan

Allyson G. Harrison, PhD, CPsych, Regional Assessment and Resource Centre, Queen's University, Kingston, Ontario, Canada

Michael W. Kirkwood, PhD, ABPP-CN, Department of Physical Medicine and Rehabilitation, University of Colorado School of Medicine and Children's Hospital Colorado, Aurora, Colorado

Jennifer Langhinrichsen-Rohling, PhD, Department of Psychology, University of South Alabama, Mobile, Alabama

Glenn J. Larrabee, PhD, ABPP-CN, independent practice, Sarasota, Florida

William S. MacAllister, PhD, ABPP-CN, Department of Neurology, New York University, New York, New York

Eric Peterson, PhD, School of Psychological Sciences, University of Northern Colorado, Greeley, Colorado

Robin L. Peterson, PhD, ABPP-CN, Department of Physical Medicine and Rehabilitation, University of Colorado School of Medicine and Children's Hospital Colorado, Aurora, Colorado

Martin L. Rohling, PhD, Department of Psychology, University of South Alabama, Mobile, Alabama

Elisabeth M. S. Sherman, PhD, RPsych, Copeman Healthcare Centre, Faculty of Medicine, University of Calgary, Calgary, Alberta, Canada

Yana Suchy, PhD, ABPP-CN, Department of Psychology, University of Utah, Salt Lake City, Utah

Marsha Vasserman, PsyD, ABPP-CN, Department of Child and Adolescent Psychiatry, New York University, and NYU Child Study Center, New York, New York

Melissa M. Womble, PhD, Department of Orthopaedic Surgery, UPMC Center for Sports Medicine, Pittsburgh, Pennsylvania

Series Editor's Note

Prior to the 1990s, research on techniques to identify noncredible performance on neurocognitive testing was sparse, and training programs in clinical neuropsychology rarely addressed the importance of, and methods for, identifying noncredible test results. However, in the 1990s a rapidly accumulating literature describing the development and validation of performance validity techniques began to appear, but these studies have, in large part, been confined to the identification of noncredible performance in adults. In fact, the prevalent view in the field before 2000 was that children do not feign deficits or malinger, and prominent child neuropsychology researchers and clinicians reported that they had never seen such cases. In 2000 I evaluated a 9-year-old litigant who had sustained a moderate traumatic brain injury. Based on my examination, I judged his very low neurocognitive scores as due to his injury until I was provided with school records that showed that he was performing above average (72nd–94th percentiles) on academic testing during the same time period in which I evaluated him. Puzzled, I requested a follow-up testing session in which I administered multiple performance validity tests (PVTs), all of which the child failed (including a significantly below-chance performance on the forced-choice measure) and which were discrepant with his intact academic skills. Records subsequently came to light showing that the plaintiff's counsel was also the child's stepfather. This case was a very important learning experience for me, and it was subsequently published as an illustration of a unique type of malingering, that is, children who malinger at the direction of adults (Lu & Boone, 2002).

Once it became apparent in the field of neuropsychology that children and adolescents can and do feign neurocognitive deficits, the question arose as to whether existing PVTs developed for adults could be imported for use in evaluations of children. The answering of this question has required a concerted research effort, with Michael W. Kirkwood and colleagues shouldering a primary role. Books addressing PVT performance in adults were published nearly 10 years ago (Boone, 2007; Larrabee, 2007), but the field has been without a comprehensive resource on performance validity testing in children until the arrival of this current volume.

In *Validity Testing in Child and Adolescent Assessment,* Michael W. Kirkwood provides a very carefully considered and thorough compilation of chapters from authoritative contributors. In the introductory chapter to Part I, he sets the stage with a well-developed rationale for the need for performance validity testing in children and adolescents. In Chapter 2, Elisabeth M. S. Sherman summarizes relevant terminology and diagnostic criteria, including those specific to child and adolescent examinees, such as differentiation of malingering by proxy (parental claim of symptoms in a child without the child malingering) versus "secondary" malingering (deliberately feigned presentations orchestrated by adults, typically parents). In Chapter 3, Eric Peterson and Robin L. Peterson summarize the research on the developmental underpinnings of deception, including the role of maturing executive functions (including working memory and inhibitory control) and theory of mind (which allows an understanding of others' private mental states and perspectives). In Chapter 4, Glenn J. Larrabee provides a concise review of the history and evolution of performance and symptom validity testing in adults, including an overview of research designs, classification statistics, and interpretation of scores from multiple PVTs.

In Part II, on detection methods, Michael W. Kirkwood (Chapter 5) summarizes current findings regarding the specificity and sensitivity of dedicated and embedded PVT cutoffs in children and adolescents and provides recommendations regarding their clinical use. He also includes the very small literature on symptom validity tests (SVTs; symptom report validity scales contained in personality inventories) in children and adolescents. In Chapter 6, Dominic A. Carone describes the importance of understanding the social contexts that can lead children and adolescents to fail to perform to true ability on testing. He then discusses the types of test-taking behaviors, unique to children and adolescents, that can signal a lack of engagement in the testing procedures and how this information, as well as data from classroom observations and PVTs, is integrated. David A. Baker and Michael W. Kirkwood (Chapter 7) then summarize available literature on the types of external and internal incentives associated with noncredible test performance, such as personal injury litigation, anxiety and poor coping skills, attention seeking, diagnosis threat, and somatization and conversion disorder, as well as some motives specific to children and adolescents (e.g., avoidance of school and sports participation). They outline a careful approach to identifying the operational motives in a child's or adolescent's failure to perform to true ability. In Chapter 8, Amy K. Connery and Yana Suchy discuss how to manage issues of noncredible test performance during an exam, including how to introduce the importance of performing to true ability at the beginning of the testing and providing feedback regarding performance invalidity subsequent to the evaluation. In Chapter 9, William S. MacAllister and Marsha Vasserman address professional standards and ethical considerations in the use of performance validity testing in children and adolescents.

Part III addresses performance validity testing in various evaluative settings, including psychoeducational (Allyson G. Harrison, Chapter 10), medical (Brian L. Brooks, Chapter 11), sports and athletic (Martin L. Rohling, Jennifer Langhinrichsen-Rohling, and Melissa M. Womble, Chapter 12), forensic (Jacobus Donders, Chapter 13), and disability (Michael D. Chafetz, Chapter 14) contexts.

Validity Testing in Child and Adolescent Assessment is an important addition to Guilford's Evidence-Based Practice in Neuropsychology series, and it provides clinicians and researchers with up-to-date, practical, authoritative, and comprehensive knowledge in this critical area.

KYLE BRAUER BOONE, PHD, ABPP, ABCN
Alliant International University Los Angeles
Los Angeles, California

REFERENCES

Boone, K. B. (2007). *Assessment of feigned cognitive impairment: A neuropsychological perspective* New York: Guilford Press.

Larrabee, G. J. (2007). *Assessment of malingered neuropsychological deficits*. New York: Oxford University Press.

Lu, P., & Boone, K. (2002). Suspect cognitive symptoms in a 9-year-old child: Malingering by proxy? *The Clinical Neuropsychologist, 16*, 90–96.

Preface

Over the last several decades, few topics in adult psychology, and especially in adult neuropsychology, have received as much attention as validity testing. The rationale for utilizing validity tests with adults has been clear for years now, and the evidentiary support behind such testing is impressively strong.

Yet, when I started practicing in the early 2000s, I, like many other pediatric neuropsychologists, child clinical psychologists, and school psychologists, had only a vague notion of what validity testing was. At that time, I evaluated several children clinically, and I was left wondering: Did these kids really try their best to do well on testing? Their performances were not egregiously suspicious, but their efforts seemed suspect. The problem was that I had no way to confidently determine whether their efforts were genuine or not. To this day, I still remember the uncertainty I felt. Was I supposed to confront the child during testing to talk about effort? Should I interpret the test data typically? What should I say to the parents and in my report about the data's validity?

My feelings of uncertainty then spurred me to look into objective validity tests that could be used reasonably with children. I've been using validity tests in my clinical assessment practice ever since, and our group at Children's Hospital Colorado has been conducting research on the topic for nearly 10 years. These experiences have left me convinced that validity testing adds value to ability-based assessments with school-age children and teens, even in nonforensic settings in which there is no obvious secondary gain apparent at the outset of the evaluation.

Although the literature on validity testing in children and teens is less well developed than the adult literature on the subject, the number of studies devoted to validity testing in child populations has increased steadily over the last decade, with exponential growth and wider clinical and mainstream recognition over the last few years. The current trend is unmistakable and, from where I sit today, I could not be more confident that the future of child and adolescent assessment will include a much greater emphasis on objectively measuring validity during both testing and self-report than has been apparent historically.

With this in mind, the time seemed right for a volume that could provide a state-of-the science synthesis of validity testing with children to guide practice and to set the stage for future research and test development. In order for the book to be worthwhile, I knew it needed a group of contributors who could not only summarize the literature but also appreciate the benefits and limitations of validity testing with children through work in their own practices. Fortunately, a veritable who's who of preeminent pediatric validity-testing clinicians and researchers and several authors renowned for their contributions to the adult literature agreed to contribute. The resulting volume thoroughly covers the "why" and "how" of validity testing in child and adolescent neuropsychological and psychoeducational assessment.

The intended audience for the book includes practitioners, researchers, and students in neuropsychology, clinical psychology, and school psychology. Educators, allied health providers, and policymakers may also find the book useful, as the chapters are written by leading experts who provide the latest scientific information about a topic that will undoubtedly grow in importance in the cognitive and psychoeducational assessment fields in the years ahead.

An edited volume is only as valuable as the individual chapters, so I first want to extend my sincere thanks to the chapter authors, who generously devoted time amid their hectic schedules to skillfully summarize the current literature and add an impressive amount of new scholarship to the field. I also want to thank my editor at The Guilford Press, Rochelle Serwator; without her gentle but persuasive nudging and continual support, this book would still be sitting in the "to-do" pile. Louise Farkas and the rest of the production team at Guilford were nothing short of superb, so I am indebted to them as well.

I additionally want to extend my appreciation to the many colleagues whose work and perspectives have shaped my thinking on the topic for the better, including Drs. David Baker, Kyle Brauer Boone, Brian Brooks, Dominic Carone, Michael Chafetz, Amy Connery, Jacobus Donders, Lloyd Flaro, Paul Green, Allyson Harrison, John Kirk, Glenn Larrabee, William MacAllister, Joel Morgan, Robin Peterson, Martin Rohling, Elisabeth Sherman, and Jerry Sweet. Last, but definitely not least, I want to thank my wife, Dr. Jennifer Janusz, who supported me while I worked on this project with her usual generous blend of tolerance, encouragement, and cheer.

Contents

Part I

Introduction

1

A Rationale for Performance Validity Testing in Child and Adolescent Assessment

MICHAEL W. KIRKWOOD

The importance of objective validity testing in adult neuropsychological and psychological assessment has long been recognized. Over the past 30 years, more than 1,000 scientific articles, 20 comprehensive reviews, a dozen meta-analytic studies, and a dozen textbooks have appeared in the adult literature (Carone & Bush, 2013; Sweet & Guidotti Breting, 2013). Practice organizations, focusing primarily on adults, have also emphasized the importance of validity testing in both clinical and independent evaluations. The National Academy of Neuropsychology published a position paper on the topic in 2005 (Bush et al., 2005), and the American Academy of Clinical Neuropsychology (AACN) published a consensus statement in 2009 (Heilbronner, Sweet, Morgan, Larrabee, & Millis, 2009). Practice guidelines from both the AACN (American Academy of Clinical Neuropsychology Board of Directors, 2007) and the British Psychological Society (2009) also highlight the need to routinely include validity testing in assessments.

Attention to validity testing in pediatric assessment pales in comparison. Select empirical work appeared in the 1980s, and a book on adolescent malingering was published in the 1990s (McCann, 1998), but serious interest was not shown until the 2000s. The first neuropsychological case report was published in 2002 (Lu & Boone, 2002), the first pediatric case series in 2003 (Constantinou & McCaffrey, 2003; Courtney, Dinkins, Allen, & Kuroski, 2003; Green & Flaro, 2003), and the first review from a neuropsychological perspective in 2004 (Rohling, 2004). Over the last decade, dozens of articles have focused on validity testing in cognitive or neuropsychological assessment with children. Although no meta-analyses have been conducted, two reviews summarizing the child literature have appeared in the last few years (DeRight & Carone, 2013; Kirkwood, 2012).

Despite the growing interest in the topic, many pediatric neuropsychologists, child clinical psychologists, and school psychologists still view the use of validity tests during their evaluations as less necessary than do adult-focused practitioners. This chapter provides a rationale for including performance validity tests (PVTs) in neuropsychological and psychoeducational batteries with school-age children.

PVTs are objective measures intended to evaluate validity during performance-based tests (Larrabee, 2012a). They are designed to be relatively insensitive to ability-based problems and to instead detect noncredible effort. Most school-age children, even those with bona fide developmental and neurological conditions, can readily pass the most well-established PVTs (see Kirkwood, Chapter 5, this volume). Of note, evaluating the veracity of self-report data during an evaluation through the use of symptom validity tests (SVTs) is also important, but as discussed in Chapter 5 of this volume, pediatric practitioners do not yet have independently validated measures to use for this purpose, at least when attempting to detect feigned or exaggerated health-related and cognitive symptomatology.

NONCREDIBLE PRESENTATIONS HAPPEN IN CHILDREN

The child literature has likely lagged so far behind the adult literature in part because many practitioners believed historically that children could not, or would not, feign or exaggerate in an assessment setting. However, as summarized by Peterson and Peterson (Chapter 3, this volume), a sizable developmental psychology literature has demonstrated that children are capable of deception by the preschool years and engage in deceptive acts quite frequently under the right circumstances. Increasingly sophisticated deceptive behavior occurs throughout childhood and into the adolescent years, secondary to the development of the underlying psychological abilities necessary for successful deception (e.g., theory of mind, working memory, inhibitory control).

Thus research documents that children and adolescents *can* deceive. A more important question in justifying the use of validity testing in children is whether or not they actually *do* deceive in health care and other assessment settings. Although deception by children is not a well-studied area in medicine, multiple case reports have appeared in the medical and psychiatric literature establishing that noncredible presentations occur, for both conscious and unconscious reasons and with and without parental influence (Kirkwood, 2012). Children have been found to feign many types of physical and psychiatric difficulties, including motor disturbance, vision and other sensory problems, seizures, psychosis, fever, skin conditions, respiratory problems, gastrointestinal upset, and orthopedic injury (Enzenauer, Morris, O'Donnell, & Montrey, 2014; Feldman, Stout, & Inglis, 2002; Greenfield, 1987; Kozlowska et al., 2007; Libow, 2000; Peebles, Sabella, Franco, & Goldfarb, 2005; Reilly, Menlove, Fenton, & Das, 2013).

As covered in more detail in subsequent chapters, multiple lines of evidence now indicate that children also feign cognitive problems. Most of the evidence so far has been accumulated in clinical, not forensic, settings. Table 1.1 highlights some of the individual case reports that have appeared in the literature. These reports offer rich descriptions of individual children providing noncredible cognitive data, with presentations varying significantly in terms of symptomatology and the underlying reasons for the distortion.

As illustrated in Table 1.2, a number of larger case series have also documented how often children provide noncredible effort during cognitive or neuropsychological evaluations. In most outpatient clinical settings,

TABLE 1.1. Individual Case Reports of Children Providing Noncredible Effort during Cognitive or Neuropsychological Evaluations

Source	Reason for referral	Age (years)
Lu & Boone (2002)	Moderate TBI	9
Henry (2005)	Mild TBI	8
Flaro, Green, & Blaskewitz (2007)	Psychoeducational evaluation	8
	Criminal charges	12
	Learning disability	16
	Autism spectrum	7
Flaro & Boone (2009)	Mild TBI	16
McCaffrey & Lynch (2009)	TBI	13
Kirkwood, Kirk, Blaha, & Wilson (2010)	Mild TBI	16
	Mild TBI	8
	Mild TBI	15
	Mild TBI	13
	Medically unexplained symptoms	16
	Medically unexplained symptoms	11
Chafetz & Prentkowski (2011)	Social Security Disability determination	9
Harrison, Green, & Flaro (2012)	Learning disability	17

Note. TBI, traumatic brain injury.

TABLE 1.2. Case Series Estimating the Base Rate of Noncredible Performance during Cognitive Evaluations

Source	Population	*N*	Age range (years)	Primary performance validity test	% of cases deemed noncredible
Donders (2005)	Mixed clinical	100	6–16	TOMM	2%
Chafetz, Abrahams, & Kohlmaier (2007); Chafetz (2008)	Social Security Disability claimants	123	6–16	TOMM MSVT	26–60%
Carone (2008)	Moderate to severe brain injury or dysfunction	38	(*M* age 11.8)	MSVT	5%
MacAllister, Nakhutina, Bender, Karantzoulis, & Carlson (2009)	Epilepsy	60	6–17	TOMM	3%
Kirkwood & Kirk (2010); Kirkwood, Hargrave, & Kirk (2011); Baker, Connery, Kirk, & Kirkwood (2014); Kirkwood, Connery, Kirk, & Baker (2014)	Mild (TBI)	~500 (independent patients)	8–17	MSVT	12–17%
Kirk, Harris, Hutaff-Lee, Koelmay, Dinkins, & Kirkwood (2011)	Mixed clinical	101	5–16	TOMM	4%
Larochette & Harrison (2012)	Learning disability	63	11–14	WMT	1%
Green, Flaro, Brockhaus, & Montijo (2012)	Mixed clinical	380	7–18	WMT	5%
Green, Flaro, Brockhaus, & Montijo (2012)	Mixed clinical	265	7–18	MSVT	3%
Green, Flaro, Brockhaus, & Montijo (2012)	Mixed clinical	217	7–18	NV-MSVT	4%
Ploetz, Mosiewicz, Kirkwood, Sherman, & Brooks (2014)	Mixed clinical	266	5–18	TOMM	3%

Note. All studies involved clinical samples except those from Chafetz. Includes only studies that provide clinical or actuarial means to estimate true positives for noncredible effort (e.g., study reported true vs. false positives or used more than one PVT). TOMM, Test of Memory Malingering; MSVT, Medical Symptom Validity Test; WMT, Word Memory Test; NV-MSVT, Nonverbal Medical Symptom Validity Test.

noncredible presentations do not occur frequently, but they do happen consistently, with at least a small percentage of children documented in every case series published to date. By comparison, 8% of adults in general medical/psychiatric clinical settings are estimated to feign or exaggerate symptomatology, with higher rates seen in forensic and other compensation-seeking contexts (Mittenberg, Patton, Canyock, & Condit, 2002).

Not unlike what is seen in adults, certain pediatric conditions and settings have been found to be associated with more frequent noncredible presentations. The clinical population found to date to display the highest rate of noncredible effort is children with persistent problems following mild head injury (for further discussion, see Brooks, Chapter 11, this volume). Noncredible presentations in this population have been documented to occur 12–20% of the time, considerably more often than other investigated clinical conditions, and have been determined in a case series described in multiple studies by our group at Children's Hospital Colorado (Baker, Connery, Kirk, & Kirkwood, 2014; Green, Kirk, Connery, Baker, & Kirkwood, 2014; Kirk, Hutaff-Lee, Connery, Baker, & Kirkwood, 2014; Kirkwood, Connery, Kirk, & Baker, 2014; Kirkwood, Hargrave, & Kirk, 2011; Kirkwood & Kirk, 2010; Kirkwood, Peterson, Connery, Baker, & Grubenhoff, 2014; Kirkwood, Yeates, Randolph, & Kirk, 2012) and also in a study by the neuropsychology group at Nationwide Children's Hospital in Ohio (Araujo et al., 2014).

As discussed by Chafetz (Chapter 14, this volume), children undergoing independent evaluation for Social Security Disability benefits display even higher rates of noncredible data. Remarkably, upward of 60% of children seen for psychological consultative examinations for Social Security Disability display some evidence of malingering, which is thought to be driven by the parents in most cases and so would be considered "malingering by proxy" (Chafetz, 2008; Chafetz, Abrahams, & Kohlmaier, 2007).

Other pediatric evaluation settings and/or conditions are also likely to be associated with elevated rates of invalid data. However, at this point, we do not yet know what these settings or conditions are, because they have not been investigated adequately. For example, little systematic examination has focused on how frequently noncredible presentations happen in child or adolescent psychoeducational settings, but, as reviewed by Harrison (Chapter 10, this volume), ample reason for concern exists given the secondary gain that is often present (e.g., accommodations for high-stakes testing; stimulant medication prescription). The concern here is significant enough that it has begun to receive mainstream recognition. For example, the College Board, which administers the SAT, has essentially taken the position that PVTs should be used during psychological–neuropsychological evaluations conducted for disability accommodations (College Board, 2014), a position that can be expected to be the norm in the not too distant future for all major national testing services.

Children with conditions that by definition involve increased rates of noncompliance (e.g., oppositional defiant disorder, conduct disorder) seem as though they might also be at higher risk for putting forth noncredible effort during assessments, and indeed there are some cases in which this happens (Carone, 2008; Chafetz, 2008; Donders, 2005); on the other hand, a study published with juvenile offenders did not actually find elevated PVT failure (Gast & Hart, 2010). Innumerable adult studies have found that the presence of external incentive for financial gain increases the chance of failed PVTs after mild head injury and other conditions (Boone, 2007; Carone & Bush, 2013; Larrabee, 2012b). In contrast, neither the pediatric group in Colorado (e.g., Kirkwood & Kirk, 2010), nor the group in Ohio (Araujo et al., 2014) has found PVT failure to be associated with family litigation in children seen clinically after a mild head injury, suggesting that the circumstances that elicit increased rates of PVT failure in children may differ to some extent from those in adults. Additional research will be required to more clearly elucidate which children under which conditions are most at risk of displaying noncredible presentations.

INADEQUACY OF SUBJECTIVE JUDGMENT IN DETECTING NONCREDIBLE DATA

The research discussed previously is focused on children who are likely engaged in outright deception. In a pediatric assessment setting, noncredible effort can also be produced for a host of other reasons that anyone working with children likely naturally appreciates, including initial separation anxiety and state-dependent fatigue, hunger, or noncompliance (see Carone, Chapter 6, this volume). Identifying the underlying reasons for invalid effort is crucial in determining the most appropriate practitioner response (as discussed by Baker and Kirkwood, Chapter 7, and Connery & Suchy, Chapter 8, in this volume). Regardless of the underlying motivation, however, as a first step, invalid data need to be recognized as invalid. All practitioners conducting assessments use a process to make determinations about whether they think the examinee exerted sufficient effort to consider the data valid or not. Historically, in child assessments, practitioners have relied on subjective judgment to make such determinations.

Clinical judgment is obviously a crucial component of any psychological or neuropsychological assessment. Nevertheless, an extensive literature has documented the potential drawbacks of relying exclusively on subjective judgment (e.g., Garb, 1998; Hastie & Dawes, 2010; Kahneman, Slovic, & Tversky, 1982; Meehl, 1954). Aspects of this research are controversial, but few would argue with the fact that the literature demonstrates that clinician judgment can be frequently flawed. Errors occur for many reasons, including lack of direct feedback about the correctness of judgments, ignoring base rates and normative data, and failure to properly assess covariation. A variety of cognitive heuristics (e.g., availability, affect, representativeness) and biases (e.g., confirmatory, hindsight) also negatively affect the accuracy of judgments. As such, objective instrumentation is widely recognized as having the potential to improve clinical decision making.

Guilmette (2013) summarized the studies that have examined how effective neuropsychologists are when making judgments about whether test data may have been feigned. A few of these studies have found support for the idea that neuropsychologists are adequate judges of malingering (e.g., Trueblood & Binder, 1997; Bruhn & Reed, 1975). However, these studies have some significant methodological problems, and a number of other studies have found considerably less support for the idea that noncredible data produced by adults can be identified effectively without validity test results (e.g., Heaton, Smith, Lehman, & Vogt, 1978; van Gorp et al., 1999).

Two studies from the late 1980s evaluated whether neuropsychologists were able to detect invalid data produced by children and adolescents asked to "fake bad" on a test battery (Faust, Hart, & Guilmette, 1988; Faust, Hart, Guilmette, & Arkes, 1988). The children and teens were instructed to perform less well than usual but not so badly that it would be obvious that they were faking. The data were then sent to neuropsychologists for review. The majority of the respondents viewed the results as abnormal. None of the respondents in either study identified noncredible responding as a possible explanation for the results, even though they were quite confident in the accuracy of their ratings. The studies have been criticized (Bigler, 1990), but even after 25 years they serve as a reminder that practitioners are apt to be less accurate than they think they are when attempting to subjectively identify invalid data.

Recognition of the many potential flaws in clinical judgment is one of the reasons most adult-focused neuropsychologists and practice organizations recommend so strongly that examiners incorporate objective validity tests into their evaluations. When examinees engage in more sophisticated deception (e.g., seemingly compliant with a plausible presentation), PVTs may be the only sign that invalid data were produced.

Adult practitioners seem to readily appreciate that PVTs, like any other tool in the testing toolbox, are not intended to replace clinical judgment. Rather, they are intended to supplement and improve it by allowing the objective measurement of a measurable behavior. In contrast, the adoption of objective validity testing in pediatric assessments has been much slower, with statements such as the following continuing to predominate in child clinical, school psychology, and some pediatric neuropsychological reports:

> "Mary *appeared* to put forth her best effort on all tasks, so the results are judged to be reliable and valid."

Imagine a similar statement being made about a child's intelligence without the use of any objective testing:

> "Mary *appeared* to have below average intelligence so she was judged to be functioning in the intellectually disabled range."

Psychologists long ago moved away from relying on such gross appearance and subjective judgment when measuring intelligence and just about every other performance-based domain we evaluate (e.g., language, memory, attention, executive functioning). Why have so many practitioners continued to rely solely on judgment to determine whether a child or teen exerted valid effort during an exam? Until fairly recently, the simple answer was that pediatric practitioners did not have access to any empirically supported objective tools to determine noncredible effort. However, as detailed by Kirkwood (Chapter 5, this volume), a growing number of PVTs now have adequate evidence to justify their inclusion in batteries with school-age children.

VALIDITY TEST RESULTS MATTER

Child practitioners less familiar with PVTs often question the added value of such tests given their financial costs and administration time. Available research with children, as well as more extensive work with adults, indicates that PVT performance likely has not only substantial implications for how providers should interpret evaluation data but broader implications as well.

Clinical Implications for Ability-Based Test Interpretation

Numerous studies in adults have demonstrated clearly that PVT failure is associated with significantly worse performance on a wide variety of neuropsychological tests. In essence, as performance on PVTs diminishes, examinee scores on neuropsychological tests decline dramatically as well (Green, Rohling, Lees-Haley, & Allen, 2001). Despite the fact that PVTs are relatively insensitive to ability-based deficits, PVT failure in adults accounts for approximately 50% of the variance on ability-based tests, far more than that explained by educational level, age, neurological condition, and neuroimaging results (e.g., Constantinou, Bauer, Ashendorf, Fisher, & McCaffrey, 2005; Green et al., 2001; Lange, Iverson, Brooks, & Rennison, 2010; Meyers, Volbrecht, Axelrod, & Reinsch-Boothby, 2011).

Only a few studies have investigated the relationship between PVT performance and ability-based tests in children. Nonetheless, available work suggests that similar relationships may exist, at least in pediatric samples with relatively high rates of noncredible effort. In a sample of 123 child Social Security Disability claimants, Chafetz (2008) classified participants according to the likelihood of malingering based on their performance on a variety of PVTs. IQ scores differed significantly among the various groups in a linear fashion, such that the worse the child performed on the PVTs, the lower his or her IQ score was.

Two studies have also focused on the relationship between PVT performance and ability-based tests in samples of children with lingering problems after mild traumatic brain injury (mild TBI). In a group of 276 school-age children referred clinically to the Children's Hospital Colorado Concussion Program, we found that performance on the Medical Symptom Validity Test (MSVT; see Kirkwood, Chapter 5, this volume, for a description) correlated significantly with performance on all ability-based tests and explained more than a third (38%) of the variance across ability tests (Kirkwood et al., 2012). Even after controlling for premorbid and injury factors that could have influenced test performance (e.g., age, history of ADHD/learning disability/special education, injury severity, time since injury), MSVT performance remained a robust unique predictor of ability-based test performance. Participants failing the MSVT also performed significantly worse on nearly all neuropsychological tests, with large effect sizes seen across most standardized tests (see Table 1.3). In comparison with children who passed the MSVT, those who failed were also at least twice as likely to perform poorly across ability-based tests (Table 1.4). The group at Nationwide Children's Hospital also recently found a similar relationship between PVT performance and the Trail Making Test in 382 children referred clinically after mild TBI (Araujo et al., 2014).

In brief, available studies indicate that noncredible effort can have a dramatic effect across most cognitive domains, not only in adult but also in child evaluations. Given the size of the effects, interpreting data without accounting for invalid effort could lead to gross interpretive errors, inaccurate diagnostic and etiological conclusions, ineffective treatment recommendations, and inappropriate health care, educational, and governmental resource utilization. Any of these errors could result in iatrogenic harm to the child and raise serious questions about a provider's competence.

Clinical Implications for Interpreting Self-Reported Data

In adults, a voluminous literature has documented that PVTs relate strongly to validity indices on personality scales, as well as to self-reported emotional, cognitive, and health-related complaints (Gervais, Wygant, Sellbom, & Ben-Porath, 2011; Greiffenstein, 2010; Jones, Ingram, & Ben-Porath, 2012; Tarescavage, Wygant, Gervais, & Ben-Porath, 2013). In child samples, minimal work has examined the relationship between PVT

TABLE 1.3. Descriptive Statistics and Comparisons between MSVT Pass and MSVT Fail Groups on Ability-Based Tests in the Kirkwood, Yeates, Randolph, and Kirk (2012) Study

	MSVT Pass			MSVT Fail				Effect size
	n	M	SD	n	M	SD	*p*	(Cohen's *d*)
WASI								
Estimated IQ	215	105.5	11.6	48	94.5	13.4	< .001*	0.9
Vocabulary T-score	215	53.6	8.6	48	50.7	10.9	.045	0.3
Matrix Reasoning T-score	215	52.4	7.2	50	41.0	10.6	< .001*	1.4
CVLT-C								
Total Learning Trials 1–5 T-score	186	53.0	8.4	40	46.6	11.4	.002*	0.7
Long Delay Free Recall *z*-score	186	0.34	0.8	40	–0.48	1.3	< .001*	0.9
Recognition Discriminability *z*-score	186	0.18	0.6	40	–1.29	1.8	< .001*	1.6
WISC-IV								
Digit Span scaled score	224	9.7	2.9	51	6.4	3.2	< .001*	1.2
Coding scaled score	207	9.7	5.3	45	6.4	3.1	< .001*	0.6
Grooved Pegboard								
Dominant hand *z*-score	213	–0.25	1.4	45	–1.7	2.5	.001*	0.9
Nondominant hand *z*-score	215	–0.41	1.5	45	–1.6	2.2	.001*	0.7
Woodcock–Johnson III								
Letter–Word ID standard score	191	100.2	9.7	45	97.0	22.0	.347	0.3
Automatized Sequencing (time in seconds)								
Alphabet	216	5.6	6.1	50	11.4	10.9	.001*	0.8
Counting 1 to 20	172	4.7	1.4	44	9.6	12.5	.013	0.9
Days of week	209	2.5	1.2	47	5.4	5.1	< .001*	1.2
Months of year	214	6.1	4.4	47	12.0	6.8	< .001*	1.2

Note. MSVT, Medical Symptom Validity Test; WASI, Wechsler Abbreviated Scale of Intelligence; CVLT-C, California Verbal Learning Test—Children's Version; WISC-IV, Wechsler Intelligence Scale for Children—Fourth Edition. From Kirkwood, Yeates, Randolph, and Kirk (2012). Copyright 2012 by the American Psychological Association. Reprinted by permission.

* significant at $p < .003$ (Bonferroni corrected value)

TABLE 1.4. Percentage of Participants in MSVT Pass and MSVT Fail Groups Performing below 1 Standard Deviation of the Normative Mean on Ability-Based Tests and Associated Odds Ratios in the Kirkwood, Yeates, Randolph, & Kirk (2012) Study

	MSVT Pass	MSVT Fail	Odds ratio (95% CI)
WASI			
Estimated IQ	5%	23%	5.5 (2.2–13.6)
Vocabulary T-score	5%	10%	2.4 (0.8–7.3)
Matrix Reasoning T-score	6%	44%	13.3 (5.9–29.8)
CVLT-C			
Total Learning Trials 1–5 T-score	5%	25%	6.6 (2.5–17.5)
Long Delay Free Recall *z*-score	3%	28%	11.4 (3.9–33.2)
Recognition Discriminability *z*-score	2%	40%	40.7 (11.0–149.9)
WISC-IV			
Digit Span scaled score	11%	59%	11.9 (5.9–24.0)
Coding scaled score	16%	49%	5.2 (2.6–10.5)
Grooved Pegboard			
Preferred hand	16%	58%	7.5 (3.7–15.0)
Nonpreferred hand	25%	44%	2.4 (1.2–4.6)
Woodcock–Johnson III			
Letter–Word Identification standard score	4%	9%	2.2 (0.6–7.8)

Note. MSVT, Medical Symptom Validity Test; WASI, Wechsler Abbreviated Scale of Intelligence; CVLT-C, California Verbal Learning Test—Children's Version; WISC-IV, Wechsler Intelligence Scale for Children—Fourth Edition; CI, confidence interval. From Kirkwood, Yeates, Randolph, and Kirk (2012). Copyright 2012 by the American Psychological Association. Reprinted by permission.

performance and self-reported data. However, two studies suggest that there is likely to be a strong relationship between PVTs and the number of health-related complaints reported, at least in the context of mild TBI. Not unlike what is seen in adult populations with mild TBI (Iverson, Lange, Brooks, & Rennison, 2010; Lange, Iverson, Brooks, & Rennison, 2010; Tsanadis et al., 2008), both our group at Children's Hospital Colorado (Kirkwood, Peterson, et al., 2014) and the group at Children's Nationwide (Araujo et al., 2014) found that patients who failed PVTs endorsed significantly more postconcussive symptoms than those who passed, even after controlling for other factors that influenced symptom reporting in the clinical samples (e.g., preinjury symptoms, female gender, premorbid anxiety/depression, time since injury).

PVT failure in adults raises suspicions about the veracity of all collected data, not just data from performance-based tests. Further work in children will be required to definitively understand the relationship between PVT performance and subjectively reported data, but existing work is consistent with the idea that those children who fail PVTs are apt to be engaging in misrepresentation during self-report as well.

Clinical Implications Case Example

The studies discussed here illustrate the potential clinical implications of PVT failure at a group level. The following case example provides an illustration of what PVT use, or lack thereof, can mean at the level of the individual child.

> Joe was a 15-year-old male who suffered a concussion in a football game. He was seen soon after injury through an emergency department, where he had a normal neurological exam and normal head computerized tomography (CT) scan. In the first days after injury, he was managed through the primary care office. Because of persistent symptomatology, he was seen at 2 weeks postinjury by a clinical psychologist, who administered a 20-minute computerized cognitive test battery and a postconcussive symptom scale. The psychologist documented "severe deficits in memory and response speed" and "an alarming number of postconcussive symptoms." Recommendations included that the teen stop going to school and "rest" his brain. He was seen again 1 week later and another time 3 weeks after that, with no change in the test results or recommendations. No validity testing was included during any of the three evaluations.
>
> Joe was then seen for neuropsychological consultation at 13 weeks postinjury, at which time he was still not back in school. He failed multiple PVTs and presented as clinically depressed. A combination of factors was thought to be contributing to the noncredible data (e.g., dislike of football and school, family dysfunction). Regardless, the lack of PVT use during the previous evaluations and the failure to detect invalid data when it was likely present almost certainly contributed to substantial errors in test interpretation and inappropriate clinical management. A number of iatrogenic effects resulted, including an exacerbation of the teen's mood due to being away from friends, school, and family routines; academic stress due to missing more than 3 months of school; and unnecessary parental alarm about what they understood to be a "severe" brain injury.

Broader Systemic Implications

Not only does PVT performance have the potential to alter the understanding and care of individual examinees, but such performance also likely has broader implications. From a research perspective, virtually no pediatric outcome studies have included validity testing as part of the test battery, which raises interpretive questions for studies focused on conditions with relatively high rates of noncredible presentations.

At a more fundamental level, the idea that all children participating in research-based cognitive testing exert adequate effort needs to be questioned. Researchers have traditionally assumed that performance-based tests primarily measure ability. The fact that performance may also reflect different levels of effort during testing has been nearly completely ignored. In a thought-provoking study by Duckworth and colleagues (Duckworth, Quinn, Lynam, Loeber, & Stouthamer-Loeber, 2011), this assumption was critically examined for intelligence testing. The researchers conducted a meta-analysis of random-assignment experiments looking at the effect of material incentives on IQ test performance in a total of 2,008 children. Incentives increased IQ scores by an average of 0.64 of a standard deviation (~10 IQ points), with larger effects apparent for individuals with lower baseline IQ scores. These results suggest that it may be as important to objectively measure effort during research settings as it is during clinical settings. If effort is not examined and controlled for explicitly, it may significantly confound the association between test results and whatever outcome is being evaluated, perhaps particularly among children who may not have the inherent motivation to perform well in a low-stakes research environment (e.g., children who are lower functioning or those with conditions that undermine motivation).

Inadequate effort may also have public health implications. This was illustrated in a recent study with adults by Horner, VanKirk, Dismuke, Turner, and Muzzy (2014), who found that PVT failure was associated with increased emergency department visits and more inpatient service use in a Veterans Affairs sample. No pediatric study has yet examined the relationship between PVT performance and health care utilization. Clinically, though, we not infrequently see cases like the one of Joe, described earlier, in which children providing noncredible data are not properly identified, in turn resulting in the inappropriate utilization of health care and educational resources.

Finally, noncredible effort during childhood assessments is also likely to be associated with an unnecessary cost to society. Chafetz and Underhill (2013) estimated the financial costs of malingered mental disorders in adult Social Security Disability evaluations to be $20.02 billion during a single year, 2011. Chafetz (Chapter 14, this volume) reports that malingered disability in youth during 2011 cost the Social Security Administration more than $2.13 billion. These amounts seem staggeringly high, but given that most pediatric practitioners do not systematically examine performance validity, they actually likely underestimate the ultimate costs to governmental, legal, school, and health care systems of children and adolescents providing noncredible data during evaluations.

CONCLUSION

Psychological and neuropsychological test interpretation rests on the assumption that the examinee responded in a credible fashion during the exam. If a child provides noncredible effort, the resulting data will represent an inaccurate representation of the child's true abilities and/or difficulties. Reliance on such data can lead to a host of problems, including errors in diagnosis, conceptualization, and management, any of which could result in potential harm to the child and questions about an examiner's competence.

Childhood noncredible presentations likely happen consistently, if not commonly, in both practice and research assessment settings, whether recognized or not. Subjective judgment alone is unlikely to be optimally effective in detecting many of these presentations. Given that several PVTs have been well validated in school-age children, the decision to *not* include PVTs now needs to be justified by the child practitioner. Some reasonable justifications for such decisions still exist. The two clearest are that the evaluation is of a preschool or younger child (for whom there has been a paucity of research) or that the evaluation is of a child who is extremely impaired (for which more work still needs to be done to confidently interpret PVTs). Nevertheless, the historic reasons that many child practitioners have provided to justify not using PVTs in their batteries (e.g., "I don't have time to include them"; "My clinical judgment is good enough") have become much less defensible. Investing a few dollars and some minutes on PVTs to help ensure that a large financial investment and data from hours of testing are interpreted accurately certainly seems worth it. A case can even be made that child or pediatric examiners choosing to not include PVTs during assessments may be acting unethically (see MacAllister & Vasserman, Chapter 9, this volume).

Pediatric specialists have decades of work ahead of them to amass a literature that even begins to approach that available right now to practitioners working with adults (see Larrabee, Chapter 4, this volume). Even so, the rationale for PVT use is convincing enough, and the extant evidence base strong enough, to justify the incorporation of PVTs in the vast majority of school-age evaluations, be they clinical, psychoeducational, forensic, or research oriented. Therefore, the default position in a pediatric or child test battery needs to move away from justifying when to include PVTs to including them routinely during evaluations, unless there is very strong justification to do otherwise.

ACKNOWLEDGMENTS

I am grateful to Drs. Robin L. Peterson, Amy K. Connery, and David A. Baker for their valuable suggestions to improve this chapter and for more generally sharpening my thinking on this topic (and many other related topics) on a near daily basis.

REFERENCES

American Academy of Clinical Neuropsychology Board of Directors. (2007). American Academy of Clinical Neuropsychology (AACN) practice guidelines for neuropsychological assessment and consultation. *The Clinical Neuropsychologist, 21*, 209–231.

Araujo, G. C., Antonini, T. N., Monahan, K., Gelfius, C., Klamar, K., Potts, M., et al. (2014). The relationship between suboptimal effort and post-concussion symptoms in children and adolescents with mild traumatic brain injury. *The Clinical Neuropsychologist, 28*, 786–801.

Baker, D. A., Connery, A. K., Kirk, J. W., & Kirkwood, M. W. (2014). Embedded performance validity indicators within the California Verbal Learning Test, Children's Version. *The Clinical Neuropsychologist, 28*, 116–127.

Bigler, E. D. (1990). Neuropsychology and malingering: Comment on Faust, Hart, and Guilmette (1988). *Journal of Consulting and Clinical Psychology, 58*, 244–247.

Boone, K. B. (2007). *Assessment of feigned cognitive impairment: A neuropsychological perspective*. New York: Guilford Press.

British Psychological Society Professional Practice Board. (2009). Assessment of effort in clinical testing of cognitive functioning for adults. Retrieved from *www.bps.org.uk/sites/default/files/documents/assessment_of_effort_in_clinical_testing_of_cognitive_functioning_for_adults.pdf.*

Bruhn, A. R., & Reed, M. R. (1975). Simulation of brain damage on the Bender-Gestalt Test by college students. *Journal of Personality Assessment, 39*, 244–255.

Bush, S. S., Ruff, R. M., Troster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., et al. (2005). Symptom validity assessment: Practice issues and medical necessity: NAN Policy and Planning Committee. *Archives of Clinical Neuropsychology, 20*, 419–426.

Carone, D. A. (2008). Children with moderate/severe brain damage/dysfunction outperform adults with mild to no brain damage on the Medical Symptom Validity Test. *Brain Injury, 22*, 960–971.

Carone, D. A., & Bush, S. S. (2013). *Mild traumatic brain injury: Symptom validity assessment and malingering*. New York: Springer.

Chafetz, M. D. (2008). Malingering on the Social Security Disability consultative exam: Predictors and base rates. *The Clinical Neuropsychologist, 22, 529–546.*

Chafetz, M. D., Abrahams, J. P., & Kohlmaier, J. (2007). Malingering on the Social Security Disability consultative exam: A new rating scale. *Archives of Clinical Neuropsychology, 22*, 1–14.

Chafetz, M. D., & Prentkowski, E. (2011). A case of malingering by proxy in a Social Security disability psychological consultative examination. *Applied Neuropsychology, 18*, 143–149.

Chafetz, M., & Underhill, J. (2013). Estimated costs of malingered disability. *Archives of Clinical Neuropsychology, 28*, 633–639.

College Board. (2014). Common diagnostic tests. Retrieved from *www.collegeboard.org/students-with- disabilities/documentation-guidelines/common-diagnostic-tests*.

Constantinou, M., Bauer, L., Ashendorf, L., Fisher, J. M., & McCaffrey, R. J. (2005). Is poor performance on recognition memory effort measures indicative of generalized poor performance on neuropsychological tests? *Archives of Clinical Neuropsychology, 20,* 191–198.

Constantinou, M., & McCaffrey, R. J. (2003). Using the TOMM for evaluating children's effort to perform optimally on neuropsychological measures. *Child Neuropsychology, 9*, 81–90.

Courtney, J. C., Dinkins, J. P., Allen, L. M., III, & Kuroski, K. (2003). Age-related effects in children taking the Computerized Assessment of Response Bias and Word Memory Test. *Child Neuropsychology, 9*, 109–116.

DeRight, J., & Carone, D. A. (2013). Assessment of effort in children: A systematic review. *Child Neuropsychology, 21*(1), 1–24.

Donders, J. (2005). Performance on the Test Of Memory Malingering in a mixed pediatric sample. *Child Neuropsychology, 11*, 221–227.

Duckworth, A. L., Quinn, P. D., Lynam, D. R., Loeber, R., & Stouthamer-Loeber, M. (2011). Role of test motivation in intelligence testing. *Proceedings of the National Academy of Sciences of the USA, 108*, 7716–7720.

Enzenauer, R., Morris, W., O'Donnell, T., & Montrey, J. (2014). *Functional ophthalmic disorders: Ocular malingering and visual hysteria*. New York: Springer.

Faust, D., Hart, K., & Guilmette, T. J. (1988). Pediatric malingering: The capacity of children to fake believable deficits on neuropsychological testing. *Journal of Consulting and Clinical Psychology, 56*, 578–582.

Faust, D., Hart, K. J., Guilmette, T. J., & Arkes, H. R. (1988). Neuropsychologists' capacity to detect adolescent malingerers. *Professional Psychology: Research and Practice, 19*, 508–515.

Feldman, K. W., Stout, J. W., & Inglis, A. F., Jr. (2002). Asthma, allergy, and sinopulmonary disease in pediatric condition falsification. *Child Maltreatment, 7*, 125–131.

Flaro, L., & Boone, K. (2009). Using objective effort measures to detect noncredible cognitive test performance in children and adolescents. In J. E. Morgan & J. J. Sweet (Eds.), *Neuropsychology of malingering casebook* (pp. 369–376). New York: Psychology Press.

Flaro, L., Green, P., & Blaskewitz, N. (2007). The importance of symptom validity testing with children: WMT and MSVT. *Praxis der Rechtspsychologie*, *17*, 125–139.

Garb, H. N. (1998). *Studying the clinician: Judgment research and psychological assessment*. Washington, DC: American Psychological Association.

Gast, J., & Hart, K. J. (2010). The performance of juvenile offenders on the Test of Memory Malingering. *Journal of Forensic Psychology Practice, 10,* 53–68.

Gervais, R., Wygant, D., Sellbom, M., & Ben-Porath, Y. (2011). Association between symptom validity test failure and scores on the MMPI-2-RF validity and substantive scales. *Journal of Personality Assessment, 93*, 508–517.

Green, C. M., Kirk, J. W., Connery, A. K., Baker, D. A., & Kirkwood, M. W. (2014). The use of the Rey 15-Item Test and recognition trial to evaluate noncredible effort after pediatric mild traumatic brain injury. *Journal of Clinical and Experimental Neuropsychology, 36*, 261–267.

Green, P., & Flaro, L. (2003). Word Memory Test performance in children. *Child Neuropsychology, 9*, 189–207.

Green, P., Flaro, L., Brockhaus, R., & Montijo, J. (2012). Performance on the WMT, MSVT, & NV-MSVT in children with developmental disabilities and in adults with mild traumatic brain injury. In C.R. Reynolds & A. Horton (Eds.), *Detection of malingering during head injury litigation* (2nd ed., pp. 201–219). New York: Plenum Press.

Green, P., Rohling, M., Lees-Haley, P., & Allen, L.M. (2001). Effort has a greater effect on test scores than severe brain injury in compensation claimants. *Brain Injury, 15,* 1045–1060.

Greenfield, D. (1987). Feigned psychosis in a 14-year-old girl. *Hospital and Community Psychiatry, 38,* 73–75.

Greiffenstein, M. (2010). The MMPI-2 symptom validity scale (FBS) not influenced by medical impairment: A large sleep center investigation. *Assessment, 17*, 269–277.

Guilmette, T. J. (2013). The role of clinical judgment in symptom validity assessment.

In D. Carone & S. Bush (Eds.), *Mild traumatic brain injury: Symptom validity assessment and malingering* (pp. 31–45). New York: Springer.

Harrison, A. G., Green, P., & Flaro, L. (2012). The importance of symptom validity testing in adolescents and young adults undergoing assessments for learning or attention difficulties. *Canadian Journal of School Psychology, 27,* 98–113.

Hastie, R., & Dawes, R. M. (2010). *Rational choice in an uncertain world: The psychology of judgment and decision making.* Thousand Oaks, CA: Sage.

Heaton, R. K., Smith, H. H., Jr., Lehman, R. A., & Vogt, A. T. (1978). Prospects for faking believable deficits on neuropsychological testing. *Journal of Consulting and Clinical Psychology, 46*, 892–900.

Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., Millis, S. R., & Conference Participants. (2009). American Academy of Clinical Neuropsychology Consensus Conference Statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist, 23*, 1093–1129.

Henry, G. K. (2005). Childhood malingering: Faking neuropsychological impairment in an 8-year old. In R. L. Heilbronner (Ed.), *Forensic neuropsychology casebook* (pp. 205–217). New York: Guilford Press.

Horner, M. D., VanKirk, K. K., Dismuke, C. E., Turner, T. H., & Muzzy, W. (2014). Inadequate effort on neuropsychological evaluation is associated with increased healthcare utilization. *The Clinical Neuropsychologist, 28*, 703–713.

Iverson, G. L., Lange, R. T., Brooks, B. L., & Rennison, V. L. (2010). "Good old days" bias following mild traumatic brain injury. *The Clinical Neuropsychologist, 24*, 17–37.

Jones, A., Ingram, M., & Ben-Porath, Y. (2012). Scores on the MMPI-2-RF scales as a function of increasing levels of failure on cognitive symptom validity tests in a military sample. *The Clinical Neuropsychologist, 26*, 790–815.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases.* New York: Cambridge University Press.

Kirk, J. W., Harris, B., Hutaff-Lee, C. F., Koelmay, S. W., Dinkins, J. P., & Kirkwood, M. W. (2011). Performance on the Test of Memory Malingering (TOMM) among a large clinic-referred pediatric sample. *Child Neuropsychology, 17*, 242–254.

Kirk, J. W., Hutaff-Lee, C. F., Connery, A. K., Baker, D. A., & Kirkwood, M. W. (2014). The relationship between the self-report BASC-2 validity indicators and performance validity test failure after pediatric mild traumatic brain injury. *Assessment, 21*, 562–569.

Kirkwood, M. W. (2012). Overview of tests and techniques to detect negative response bias in children. In E. M. S. Sherman & B. L. Brooks (Eds.), *Pediatric forensic neuropsychology* (pp. 136–161). New York: Oxford University Press.

Kirkwood, M. W., Connery, A. K., Kirk, J. W., & Baker, D. A. (2014). Detecting performance invalidity in children: Not quite as easy as A, B, C, 1, 2, 3 but automatized sequences appears promising. *Child Neuropsychology*, *20*, 245–252.

Kirkwood, M. W., Hargrave, D. D., & Kirk, J. W. (2011). The value of the WISC-IV Digit Span subtest in detecting noncredible performance during pediatric neuropsychological examinations. *Archives of Clinical Neuropsychology, 26*, 377–384.

Kirkwood, M. W., & Kirk, J. W. (2010). The base rate of suboptimal effort in a pediatric mild TBI sample: Performance on the Medical Symptom Validity Test. *The Clinical Neuropsychologist, 24,* 860–872.

Kirkwood, M. W., Kirk, J. W., Blaha, R. Z., & Wilson, P. (2010). Noncredible effort during pediatric neuropsychological exam: A case series and literature review. *Child Neuropsychology, 16*, 604–618.

Kirkwood, M. W., Peterson, R. L., Connery, A. K., Baker, D. A., & Grubenhoff, J. A. (2014). Postconcussive symptom exaggeration after pediatric mild traumatic brain injury. *Pediatrics, 133*, 643–650.

Kirkwood, M. W., Yeates, K. O., Randolph, C., & Kirk, J. W. (2012). The implications of symptom validity test failure for ability-based test performance in a pediatric sample. *Psychological Assessment, 24*, 36–45.

Kozlowska, K., Nunn, K. P., Rose, D., Morris, A., Ouvrier, R. A., & Varghese, J. (2007). Conversion disorder in Australian pediatric practice. *Journal of the American Academy of Child and Adolescent Psychiatry, 46,* 68–75.

Lange, R.T., Iverson, G. L., Brooks, B. L., & Rennison, L. A. (2010). Influence of poor effort on self-reported symptoms and neurocognitive test performance following mild traumatic brain injury. *Journal of Clinical and Experimental Neuropsychology, 32*, 961–972.

Larochette, A. C., & Harrison, A. G. (2012). Word Memory Test performance in Canadian adolescents with learning disabilities: A preliminary study. *Applied Neuropsychology—Child, 1*, 38–47.

Larrabee, G. J. (2012a). Performance validity and symptom validity in neuropsychological assessment. *Journal of the International Neuropsychological Society*, *18,* 625–631.

Larrabee, G. J. (2012b). *Forensic neuropsychology: A scientific approach* (2nd ed.). New York: Oxford University Press.

Libow, J. A. (2000). Child and adolescent illness falsification. *Pediatrics, 105,* 336–342.

Lu, P. H., & Boone, K. B. (2002). Suspect cognitive symptoms in a 9-year-old child: Malingering by proxy? *Clinical Neuropsychology, 16*, 90–96.

MacAllister, W. S., Nakhutina, L., Bender, H. A., Karantzoulis, S., & Carlson, C. (2009). Assessing effort during neuropsychological evaluation with TOMM in children and adolescents with epilepsy. *Child Neuropsychology, 15*, 521–531.

McCaffrey, R. J., & Lynch, J. K. (2009). Malingering following documented brain injury: Neuropsychological evaluation of children in a forensic setting. In J. E. Morgan & J. J. Sweet (Eds.), *Neuropsychology of malingering casebook* (pp. 377–385). New York: Psychology Press.

McCann, J. T. (1998). *Malingering and deception in adolescents: Assessing credibility in clinical and forensic settings*. Washington, DC: American Psychological Association.

Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis: University of Minnesota Press.

Meyers, J. E., Volbrecht, M., Axelrod, B. N., & Reinsch-Boothby, L. (2011). Embedded symptom validity tests and overall neuropsychological test performance. *Archives of Clinical Neuropsychology, 26*, 8–15.

Mittenberg, W., Patton, C., Canyock, E. M., & Condit, D. C. (2002). Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology, 24,* 1094–1102.

Peebles, R., Sabella, C., Franco, K., & Goldfarb, J. (2005). Factitious disorder and malingering in adolescent girls: Case series and literature review. *Clinical Pediatrics, 44,* 237–243.

Ploetz, D. M., Mosiewicz, A. M., Kirkwood, M. W., Sherman, E. M. S., & Brooks, B. L. (2014). Performance on the Test of Memory Malingering in children with neurological conditions. *Child Neuropsychology*. Available at *www.tandfonline.com/doi/abs/10.1080/09297049.2014.986446#preview.*

Reilly, C., Menlove, L., Fenton, V., & Das, K. B. (2013). Psychogenic nonepileptic seizures in children: A review. *Epilepsia, 54*, 1715–1724.

Rohling, M. L. (2004). Who do they think they're kidding: Review of the use of symptom validity tests with children. *Division of Clinical Neuropsychology Newsletter, 22,* 1–7, 21–26.

Sweet, J. J., & Guidotti Breting, L. M. (2013). Symptom Validity Test Research: Status and clinical implication. *Journal of Experimental Psychopathology, 4,* 6–19.

Tarescavage, A. M., Wygant, D. B., Gervais, R. O., & Ben-Porath, Y. S. (2013). Association between the MMPI-2 restructured form (MMPI-2-RF) and malingered neurocognitive dysfunction among non-head-injury disability claimants. *The Clinical Neuropsychologist, 27*, 313–335.

Trueblood, W., & Binder, L. M. (1997). Psychologists' accuracy in identifying neuropsychological test protocols of clinical malingerers. *Archives of Clinical Neuropsychology, 12*, 13–27.

Tsanadis, J., Montoya, E., Hanks, R. A., Millis, S. R., Fichtenberg, N. L., & Axelrod, B. N. (2008). Brain injury severity, litigation status, and self-report of postconcussive symptoms. *The Clinical Neuropsychologist, 22*, 1080–1092.

van Gorp, W. G., Humphrey, L. A., Kalechstein, A., Brumm, V. L., McMullen, W. J., Stoddard, M., et al. (1999). How well do standard clinical neuropsychological tests identify malingering?: A preliminary analysis. *Journal of Clinical and Experimental Neuropsychology, 21*, 245–250.

2

Terminology and Diagnostic Concepts

ELISABETH M. S. SHERMAN

GENERAL CONCEPTS AND TERMINOLOGY IN VALIDITY TESTING

Terminology

Validity testing matters. Regardless of work setting, psychologists and neuropsychologists need to be able to detect when examinees produce invalid test results. As highlighted elsewhere in this book, the percentage of children providing invalid data may range from 5% in pediatric medical settings to 30% or more in disability or personal injury settings. Validity testing is designed to identify scores that are not credible or reflective of the examinee's actual skill level, as well as symptoms or behaviors that are not attributable to a bona fide medical or neurological condition. Validity testing tools allow practitioners to make decisions about examinee intents or behaviors (i.e., is the examinee feigning or exaggerating symptoms?) and about the accurateness of scores (i.e., is a score a good reflection of ability, or is it an underestimate?).

The terms *feigning, exaggeration,* and *fabrication* are sometimes used interchangeably in the literature to describe examinee behaviors. However, each of these terms designates distinct, specific entities (see Table 2.1). *Feigning* is pretending there is a condition or symptom that is not in fact present (e.g., simulating forgetting test items when no memory problems exist). *Exaggeration* consists of embellishing or amplifying existing symptoms stemming from a condition that is present (e.g., exaggerating the impact of a head injury on daily function). *Fabrication* involves creating symptoms de novo for a condition that is not present (e.g., claiming long-term effects from a work injury that never occurred; falsifying health records). Last, the term *induction*, seen more in the

TABLE 2.1. Proposed Terminology and Definitions for Describing Invalid Results

Definition	Example
Feigning	
Pretending there is a condition or symptom that is not present	An examinee simulates forgetting test items when no memory problem exists. A healthy young adult knowingly endorses items suggestive of severe executive dysfunction on ADHD rating scales when no symptoms are present.
Exaggeration	
Embellishing or amplifying existing symptoms stemming from a bona fide condition	An adolescent reports that a confirmed mild TBI has caused him to be unable to study or go to school. A parent reports severe delays on an adaptive behavior scale in a child with mild ADHD.
Fabrication	
Creating evidence for a condition that one does not have	A worker reports long-term effects from a work injury that never actually occurred. A parent falsifies health records in order to get accommodations at school for her child.
Induction	
Deliberately creating a condition or symptoms	An adolescent creates a skin wound by manual irritation in order to avoid hospital discharge.

Note. ADHD, attention-deficit/hyperactivity disorder; TBI, traumatic brain injury.

medical than neuropsychological literature, refers to deliberately creating a condition in oneself or someone else (e.g., injecting oneself with a toxic substance).

There are several terms for describing invalid test results. In neuropsychology, examinees are sometimes said to have shown "poor effort" when test scores are invalid, a term that implies that the examinee purposefully intended to not fully engage in the testing process. Although it has face validity, the term *effort* is problematic because high effort can be expended by examinees toward producing invalid but believable results. Other terms used in neuropsychology focus on describing facts and do not necessarily imply intent, including the terms *suboptimal performance* and *noncredible performance*. In the medical domain, purely descriptive terms such as *abnormal illness behavior* and *medically unexplained symptoms* have replaced older terms such as *psychogenic symptoms* (in this case, implying a psychological cause).

Although the field has not yet reached consensus on precise terms, when referring to examinee behaviors and test performance relating to invalid results, it is preferable to use clear and specific terminology (e.g., *exaggeration* or *fabrication*) and to opt for a more behavioral or observational description (e.g., *noncredible performance*) rather than a term such as *poor effort* or *psychogenic symptoms*, terms that imply knowledge of the examinee's internal states and motivations.

Measuring Validity

An examinee's performance is deemed to be valid when it provides an accurate estimate of that examinee's actual skill level, as measured for a specific purpose and circumstance. The same applies to self- or other-rated symptom reporting on standardized questionnaires, as well as to other aspects of the assessment, such as behavior. Thus assessing validity applies to several components of the neuropsychological assessment: neuropsychological performance, behavioral presentation, self-reported symptoms, and other-reported symptoms (Bass & Halligan, 2014). Validity testing of cognition is accomplished using performance validity tests (PVTs), whereas validity testing of self-reported symptoms is accomplished via symptom validity tests (SVTs; Larrabee, 2012).

Clinical Diagnoses and Invalid Performance

Some clinical conditions are diagnostically defined by the presence of invalid performance or symptoms that are incompatible with medical history; others are associated with high rates of invalid performance or symptoms but are not defined by deception. This is a critical distinction.

Neuropsychologists need to be familiar with current diagnostic models and terminologies to identify these conditions. Clinicians are often reluctant to diagnose conditions involving unexplained symptoms attributed to deception or psychological factors; these include malingering, factitious disorder, and conversion disorder. In clinical practice, these disorders are rare, in part because they are often unrecognized by clinicians but also because of clinician reluctance to diagnose (Hamilton et al., 2012). It is certainly true that correctly identifying disorders involving deception is a diagnostic challenge, made more complicated by the fact that these can co-occur in the presence of bona fide medical conditions, as well as co-occurring with unexplained medical symptoms that are not being feigned. Unlike most disciplines, however, neuropsychology has distinguished itself by the creation of specific diagnostic guidelines to aid in identifying deception, as well as an impressive array of clinical tools to help identify invalid test results. Most important, because early detection is critical in avoiding harm to children, particularly in the case of severe factitious disorders (Bass & Halligan, 2014; Ferrara et al., 2012), a clear understanding of the distinguishing features of these disorders and a proactive attitude to assessment is recommended.

This chapter presents a review of pediatric disorders defined primarily by deception (malingering and factitious disorder) and reviews clinical disorders in which the incidence of malingering may be elevated (e.g., attention-deficit/hyperactivity disorder [ADHD], posttraumatic stress disorder [PTSD]). The presence of these disorders should cue the practitioner to carefully assess validity. Last, conditions associated with unexplained or invalid test results that do not involve deception (i.e., conversion disorder, illness anxiety disorder, and dissociative amnesia) are reviewed. See Table 2.2 for a list of the disorders discussed in the chapter.

MALINGERING

The two main classes of conditions defined by deceptive behavior are malingering and factitious disorder. The main difference between them is the presence of secondary gain in the former and the absence of obvious external rewards in the latter. Both are important to understand in the context of neuropsychological assessment.

Malingering in DSM-5

In the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5; American Psychiatric Association, 2013), malingering is currently conceptualized as a condition that may be the focus of clinical attention but that is not a mental health diagnosis. In DSM-5, malingering is defined as the intentional production of false or exaggerated symptoms for external incentives (avoiding work or military duty, obtaining financial compensation, obtaining drugs, evading criminal prosecution).

Malingering is classified in DSM-5 under "Non-adherence to Medical Treatment." Conceptualizing malingering in neuropsychological assessment as a form of *nonadherence to neuropsychological assessment* is also a useful way of helping neuropsychologists maintain a behavioral, nonjudgmental approach to detection and interpretation.

Malingered Neurocognitive Dysfunction

Over the last two decades, various definitions of and criteria for malingering of cognitive problems have been proposed for application in neuropsychology. Among these, the Slick, Sherman, and Iverson (1999) criteria continue to stand the test of time as the most commonly used definition and the definition with the strongest empirical basis (e.g., Larrabee, 2012). In the Slick et al. (1999) framework, the term *malingered neurocognitive dysfunction* (MND) is defined as "the volitional exaggeration or fabrication of cognitive dysfunction for the purpose of obtaining substantial material gain, or avoiding or escaping formal duty or responsibility" (Slick et al., 1999, p. 552).

TABLE 2.2. Conditions and Diagnoses to Consider in Cases of Exaggerated, Feigned, or Unexplained Symptoms

- Malingered neurocognitive dysfunction (MND)
- Malingered neurocognitive dysfunction by proxy (MND by proxy)
- Factitious disorder imposed on self
- Factitious disorder imposed on another
- Conversion disorder (functional neurological symptom disorder)
- Illness anxiety disorder
- Dissociative amnesia

These criteria have become known as the "malingering criteria" or "Slick criteria." In this framework, definite MND is defined by the presence of (1) below-chance performance on forced-choice measures, (2) high posterior probability (≥ .95 that performance is substantially below actual ability level) on one or more well-validated psychometric indices, or (3) self-reported symptoms that are unambiguously incompatible or directly contradicted by observed behavior and/or test performance. Any compelling evidence for malingering that falls short of these possibilities is therefore defined as *probable malingering*.

Other authors have conceptualized malingering as falling into three categories: (1) nonexisting symptoms that are completely feigned (pure), (2) actual symptoms that are exaggerated (partial), or (3) deliberate misattribution of actual symptoms to a compensable event (false imputation; Bass & Halligan, 2014). According to Bass and Halligan (2014), exaggeration of symptoms or partial malingering is expected to be more common than pure malingering. MND diagnostic criteria do not differentiate between these categories, but they could be a consideration in individual cases.

Malingered Neurocognitive Dysfunction in Children ("Secondary MND")

All parents know that children are perfectly capable of dissimulation, exaggeration, or outright fabrication and that deception may be employed for secondary gain by children, even very young children (see Peterson & Peterson, Chapter 3, this volume). According to the strictest definition of malingering, deceiving for purposes of external incentives is equivalent to malingering. But can children really malinger? Should the term *malingering* be reserved for adults?

The original Slick et al. (1999) criteria made reference to malingering in vulnerable examinees such as children without providing clear guidelines on what to do when malingering occurs or whether malingering could even be invoked in minors. Young examinees were thought to be capable of exaggerating symptoms for material gain but were not deemed to have the ability to appreciate the consequences of their actions. Malingering through coercion (e.g., by a child under the influence of an adult) was also noted but not clearly defined.

Slick and Sherman (2012) therefore proposed an update of these criteria (Table 2.3). The main changes were to divide malingering into three subcategories: (1) *primary MND*, which essentially reflects the original Slick et al. (1999) definition of MND; (2) *secondary MND*, for children (or those with developmental, neurological, or psychiatric disorders); and (3) *MND by proxy*, in which MND is attributable to coercion, such as the case of a parent influencing the test performance of a child. Importantly, although this is not noted in the criteria, a diagnosis of MND by proxy should be applied to the parent, not to the vulnerable child.

Kirkwood, Kirk, Blaha, and Wilson (2010) describe a case series of six children with noncredible aspects to their presentation on neuropsychological examination. In four of these children, there was evidence of clear external incentives, yet few were felt to reach the Slick et al. (1999) criteria for MND. External incentives included typical childhood-type external rewards, such as getting to stay home from school, avoiding a high-pressure sports situation, and delaying or stopping a parental separation. Unlike the majority of adults who malinger, compensation, litigation, disability, and criminal charges were not primary factors in this clinical case series nor in a larger pediatric case series published by the same authors (Kirkwood & Kirk, 2010).

In contrast, in other non-neuropsychology case series, malingering has been diagnosed in children based on very similar external incentives. For example, Peebles, Sabella, Franco, and Goldfarb (2005) describe two adolescent girls who malingered by simulating skin and wound problems sufficiently severe to require hospitalization. Both had clear external incentives involving avoiding school, one due to harassment at school and one due to too-high academic expectations. Both confessed when confronted and had excellent outcomes once their fears of attending school were addressed and they were given support (Peebles et al., 2005). School phobia, school refusal, severe test anxiety, and a history of being bullied or harassed at school may therefore be triggers for malingering in children.

The decision as to whether children can malinger may relate to their capacity to understand consequences, but it also relates to the definition of malingering and the types of incentives included in the definitions. As noted before, examples of external incentives in DSM-5 and the Slick criteria include avoiding military duty, avoiding work, obtaining financial compensation, evading criminal prosecution, or obtaining drugs, all of which are incentives for adults, not children. Notably, the DSM-5 criteria specify that malingering can be adaptive—as in the case of being captured by the enemy during wartime. Children are also capable of using deception in an adaptive manner—for example, to help them evade aversive situations

TABLE 2.3. Proposed Diagnostic Criteria for Malingered Neuropsychological Dysfunction (MND): A Revision and Extension of the Slick et al. (1999) Criteria for Malingered Neurocognitive Dysfunction

Primary MND

Definite

1. Presence of a substantial external incentive for exaggeration/fabrication of symptoms (Criterion 1)
2. One or more very strong indicators of exaggeration/fabrication of neuropsychological problems or deficits (one or more of Criteria 2.1–2.3)
3. Behaviors meeting necessary criteria are not substantially accounted for by psychiatric, neurological, or developmental factors

Probable

1. Presence of a substantial external incentive for exaggeration/fabrication of symptoms (medical–legal secondary gain)
2. Three or more indicators of possible exaggeration/fabrication of neuropsychological problems or deficits (three or more of Criteria 3.1–3.7)

Secondary MND (definite and probable)

Criteria for definite or probable MND are otherwise met, but there are compelling grounds to believe that at the time of assessment the examinee did not have the cognitive capacity to understand the moral–ethical–legal implications of his or her behavior, and/or was unable to control his or her behavior, secondary to immaturity (i.e., in childhood) or bona fide developmental, psychiatric, or neurological disorders or injuries of *at least* moderate severity.

MND by Proxy (definite and probable)

Criteria for definite or probable MND are otherwise met, but there are compelling grounds to believe that a vulnerable examinee acted primarily under the guidance, direction, influence, or control of another individual. Examinees may be vulnerable to the influence of others by virtue of immaturity, neurodevelopmental and cognitive disabilities, and psychiatric illness, or by perceived inability to escape or avoid substantial coercion such as threats of physical harm for failure to behave as directed.

Specific Criteria

1) Presence of a substantial external incentive for exaggeration/fabrication of symptoms (medical–legal secondary gain)
2) Very strong indicators of exaggeration/fabrication of neuropsychological problems or deficits
    - 2.1) Below chance performance (≤ .05) on one or more forced choice measures
    - 2.2) High posterior probability (≥ .95) that performance is substantially below actual ability level) on one or more well-validated psychometric indices
    - 2.3) Self-reported symptoms are unambiguously incompatible with or directly contradicted by directly observed behavior and/or test performance
3) Possible indicators of exaggeration/fabrication of neuropsychological problems or deficits
    - 3.1) Data from one or more well-validated psychometric measures, while not sufficient to meet Criterion 2a or 2b, are on balance more consistent with noncompliance than compliance
    - 3.2) Marked and implausible discrepancy between test performance and level of function expected based on developmental and medical history

(*continued*)

TABLE 2.3. (*continued*)

3.3) Marked and implausible discrepancy between test performance and directly observed behavior and capabilities
3.4) Marked and implausible discrepancy between test performance and reliable collateral reports concerning behavior and capabilities
3.5) Marked and implausible discrepancy between self-reported and documented history, consistent with exaggeration of preinjury level of function and capabilities, minimization of preexisting injuries or neuropsychological problems, and/or exaggeration of the severity of new injuries
3.6) Marked and implausible discrepancy between self-reported symptoms and level of function expected based on developmental and medical history
3.7) Marked and implausible discrepancy between self-reported symptoms and information obtained from reliable collateral informants

Note. From Slick & Sherman (2012). Copyright 2012 by Oxford University Press. Reprinted by permission.

involving trauma or pain. Examples might be malingering to be allowed to stay longer in a hospital to escape abuse at home or to escape being returned to a school setting in which severe bullying has occurred. For now, the Slick and Sherman (2012) criteria address the difference between adult and child malingering by deeming any malingering by minors as "secondary malingering."

The lack of identification of MND in cases with clear external incentives in some studies suggests that a revision of the MND criteria is in order, at least in children. Similar concerns have been raised in adults. Additional diagnostic categories, namely cogniform disorder and its milder variant, cogniform condition, were proposed by Delis and Wetter (2007) to "encompass cases of excessive cognitive complaints and inadequate test-taking effort in the absence of sufficient evidence to diagnose malingering" (p. 589). The need for creating additional categories does raise the question as to whether the external incentives defined in the Slick criteria are too biased toward medicolegal settings, making them more difficult to apply in nonforensic cases in children.

The literature on prognosis of child malingering is very scant. One case series suggests good prognosis (Peebles et al., 2005), whereas others indicate poor outcome (Chafetz & Prentkowski, 2011). Context may play a role in outcomes, and will vary depending on the setting (e.g., pediatric hospital vs. medicolegal vs. social services).

Regardless of the cause, it is important to look carefully for external incentives that matter to children (not necessarily adults) in cases of PVT or SVT failure during neuropsychological evaluation. Doing this may help identify children who are employing malingering behaviors to avoid common childhood responsibilities, such as school, or rarer but much more severe situations involving abuse, situations that require comprehensive and prompt intervention.

Malingered Neurocognitive Dysfunction by Proxy (MND by Proxy)

As noted by Chafetz and Prentkowski (2011), a child need not conceive of and develop a strategy to malinger for external reward; simple coaching by an adult, typically a parent, would suffice. To date, only a small number of papers describing cases of malingering of cognitive problems by proxy have been reported in the literature, two involving litigation and one seen as part of a Social Security Disability evaluation (Chafetz & Prentkowski, 2011; Lu & Boone, 2002; McCaffrey & Lynch, 2009).

Chafetz (2008) reported that a large proportion of children (20–26%) met criteria for malingering in a Social Security Disability setting and noted that some of these children may not have appreciated the consequences of their actions and may have been coached by parents. Chafetz and Prentkowski (2011) also reported on a clear case of malingering by proxy in the context of a Social Security Disability evaluation that involved deliberate deception on the parent's part, as well as failure by the child on PVTs and other indices of deliberate feigning and exaggeration of symptoms.

MND by proxy can also include intentional production of fabricated behaviors rather than cognitive deficits, as in the case of fabricated ADHD symptoms, so that a parent may receive disability benefits on the child's behalf. Walker (2011) described a case in which a mother instructed her son to show disruptive behavior during an evaluation so that she could obtain Social Security benefits on his behalf for learning problems and ADHD. A false history was provided by the mother, and the child confessed to the coaching. This case would meet Slick criteria for malingering by proxy.

Differentiating Malingering from Malingering by Proxy in Child Examinees

Differentiating MND from MND by proxy involves a careful review of evidence because of the different ways in which deception can manifest in neuropsychological assessment involving children and parents. Three scenarios illustrate the complexity inherent in identifying deception in child evaluations. In the first, the parent originates the plan to deceive and carries the plan out independently. The parent produces SVT invalidity on standardized parent scales by exaggerating child symptoms and provides a falsified history. If the child is not colluding, this represents MND by proxy. A second scenario involves a parent instructing a child to do poorly on tests; the child fails PVTs, the parent provides invalid SVTs. In this case, both secondary malingering (by the child) and malingering by proxy (by the parent) are at play. In a third scenario, the child acts alone to obtain an external reward by failing memory tests and exaggerating problems on standardized questionnaires after a minor concussion to avoid going back to school; the parent is unaware of the deceptive behavior. This would be a case of malingering in a child (i.e., secondary malingering, according to Slick and Sherman criteria).

CLINICAL DISORDERS WITH A HIGH INCIDENCE OF MALINGERING

Several disorders may raise the likelihood of deception and malingering. These are reviewed in this section, as well as in other chapters in this book.

Traumatic Brain Injury

Pediatric traumatic brain injury (TBI) is the most common condition associated with deception in the neuropsychological literature. In adults, the base rates of malingering in TBI depend to a large degree on whether the individual is seen in a clinical or medicolegal setting. Although there exist as yet no precise data on the percentage of children who meet criteria for MND in medicolegal evaluations for TBI, evidence suggests that these rates are likely substantial, based on the base rate of invalid results found in nonforensic settings. For example, Kirkwood and Kirk (2010) reported that 17% of a pediatric clinical sample failed at least one PVT. Actual estimates in forensic settings are likely closer to the estimated 40% or higher of adults who malinger during medicolegal neuropsychological evaluations. In the detection of invalid test results, objective tests appear to have superior detection compared with subjective tests; 8% of pediatric patients with mild TBI failed SVTs (i.e., the Behavior Assessment System for Children, Second Edition [BASC-2]), even though 18% failed at least one PVT (Kirk, Hutaff-Lee, Connery, Baker, & Kirkwood, 2014). TBI is discussed in more detail elsewhere in this volume (see Kirkwood, Chapters 1 and 5, and Brooks, Chapter 11).

ADHD and Learning Disability

Malingered ADHD is a common problem in settings in which clinicians evaluate adolescents, college students, and adults for ADHD. In older adolescents and young adults, estimates range from 10 to 17% in psychoeducational or neuropsychological assessment referral settings (Harrison, Rosenblum, & Currie, 2010; Marshall et al., 2010), but it was estimated as high as 50% in one campus study (Sullivan et al., 2007). The Slick criteria have been used successfully in identifying malingered ADHD (Harrison et al., 2010; Marshall et al., 2010). In studies reporting very high rates of feigning (e.g., Sullivan et al., 2007), the use of the Slick criteria would likely adjust malingering rates downward to a more realistic base rate closer to 20% (Booksh, Pella, Singh, & Gouvier, 2010).

A diagnosis of ADHD entails a number of tangible incentives: accommodations for testing, extra assistance for schoolwork that might improve grades, and, most important, access to stimulants—drugs that have a significant street value (Rabiner, 2013). ADHD symptoms are easy to fabricate on most commonly used rating tools (Musso & Gouvier, 2014). Musso and Gouvier (2014) found that failure on three or more PVTs was most useful at detecting malingered ADHD, but good identification has also been reported with two PVTs (Jasinski et al., 2011).

Despite these high rates, a review of the existing scientific literature on ADHD feigning, exaggeration, and malingering in children 16 and under produced no studies. Nevertheless, it is likely that malingered ADHD occurs in younger adolescents and possibly children. Parents may also engage in MND by proxy by exaggerating ADHD symptoms on standardized rating scales; coaching children is also not unheard of. As noted previously, Walker's (2011) case of coaching of ADHD symptoms by a mother certainly meets criteria for malingered ADHD by proxy. See Harrison (Chapter 10, this volume) for an in-depth discussion of ADHD.

Reports on malingering in young adults with learning disabilities such as reading disorders are more rare, but estimates are still substantial, ranging up to 15% (see Harrison, Chapter 10, this volume). Including PVTs and SVTs in child ADHD and learning disability evaluations is prudent; including them in evaluations of older adolescents and college-age students is necessary.

Conduct Disorder and Oppositional Defiant Disorder

Conduct disorder and oppositional defiant disorder may be conceptualized as precursors or childhood variants of antisocial personality disorder, which can be diagnosed only in adults. The DSM-5 indicates that malingering should be suspected if antisocial personality disorder is present, along with other flags (e.g., medicolegal evaluation). The association between conduct disorder, oppositional defiant disorder, and validity testing in children has not been well studied. However, Chafetz (2008) concluded that conduct disorder symptoms may increase the severity of PVT failure; in his study, conduct disorder symptoms were associated with PVT performance, and symptoms of conduct disorder were highest in children meeting Slick criteria for malingering.

Posttraumatic Stress Disorder

Up to half of children may meet criteria for PTSD after road traffic accidents (Mehta & Ameratunga, 2012). However, no identified studies in children use methodology to screen for exaggeration and feigning, even though PTSD is associated with a high rate of malingering in adults across personal injury, military, and criminal settings. Gaughwin (2008) notes that PTSD has become ubiquitous in the forensic setting and that PTSD symptoms can be readily simulated when the examinee is seeking a benefit. Benefits for PTSD include disability, compensation for victims of crime, and civil settlement, but criminal or civil responsibility may also be reduced if PTSD can be proven.

Lubit, Hartwell, van Gorp, and Eth (2002) caution that parents may coach children to create or exaggerate complaints and that children may be more likely to feign for secondary gain as they enter adolescence. However, no published instances of parental coaching of PTSD symptoms could be found in the literature as of this writing.

Notably, one questionnaire for screening for PTSD in children, the Trauma Symptom Checklist for Children (Briere, 1996), does include validity scales. One study in adults showed that these scales might be useful in detecting malingered PTSD (Gray, Elhai & Briere, 2010). Although the incidence of exaggeration and malingering in child PTSD is unknown, it would be prudent to include both SVTs and PVTs in the assessment of children and adolescents with presumed PTSD.

FACTITIOUS DISORDER

Factitious disorder is the fabrication of physical or psychological symptoms in order to deceive. DSM-IV did not include the requirement for falsification; it required that the motivation was to assume the sick role. DMS-5 requires *evidence of falsification or deception*. DSM also stipulates more simply that the behavior should exist in the absence of external rewards, instead of the requirement for a psychological incentive such as the sick role. Conversely, the presence of material gain or external incentives does not rule out factitious disorder, and so this may be a diagnosis to consider in cases in which both internal and external incentives are involved.

In factitious disorder, deception is accomplished by exaggeration, fabrication, feigning, and induction of a medical condition. This may take the form of false reporting of symptoms, fabricated history, deliberate tampering with lab tests, or deliberate physical injury. The deliberate deception and fabrication of symptoms differentiates factitious disorder from other somatic symptom disorders, discussed later in this chapter.

The DSM-5 formulation divides factitious disorder into two separate subcategories: *factitious disorder imposed on self* and *factitious disorder imposed on another*. The latter replaces "factitious disorder by proxy." According to some authors, *Munchausen disorder* and *Munchausen disorder by proxy* are terms used to describe the severest forms of the disorder. DSM-5 does not include these terms in its formulation, but they are nevertheless covered here due to their clinical and legal importance.

In children, both factitious disorder and factitious disorder imposed on another can be encountered. These disorders describe the separate but related scenarios of a child (or more likely an adolescent) deliberately exaggerating symptoms or creating illness and of a parent fabricating or exaggerating a child's symptoms (factitious disorder by proxy). Factitious disorder occurring by itself in minors may be more common in older children and adolescents, whereas factitious disorder imposed on another is more common in very young, dependent children. For example, in one study of mothers who fabricated illness, 75% of children were under the age of 5 years (Bass & Jones, 2011). However, both can be present in children old enough to deceive—that is, when both the child and the parent collude to falsify or deceive. Like MND and MND by proxy, both child and parent behaviors, intentions, and test results must be reviewed to establish the diagnosis.

Factitious disorder, although rare, should be considered in the differential diagnosis of children presenting with medically and neurologically unexplained symptoms. In one hospital series, prevalence was 1.8% (Ferrara et al., 2012). In its mildest form, the child reports disease symptoms in the absence of any identifiable signs on observation or examination. In its severest form, factitious disorder involves the deliberate induction of injury (Peebles et al., 2005). Neurological symptoms are common, and seizures are the most commonly reported symptom in fabricated illness in children (Barber & Davis, 2002). Other common symptoms involve dermatological symptoms (e.g., unexplained skin rashes) and pain (abdominal or headache). Self-inflicted or aggravated wounds and rashes were common in one pediatric series (Peebles et al., 2005). Children may have low school attendance, low involvement in sports, visible aids for the sick role such as wheelchairs, and be socially isolated (Bass & Glaser, 2014). Prompt identification is critical and may prevent adult factitious disorder, which is a much more intractable and difficult disorder to treat effectively. In adults with factitious disorder, symptoms tend to begin in adolescence, again highlighting both the chronicity of the condition into adulthood and the need for early identification to prevent generational transmission of factitious behaviors from vulnerable adolescents to their future children.

Munchausen Disorder and Munchausen Disorder by Proxy

Munchausen disorder by proxy is a term referring to an extreme form of factitious disorder occurring in about 20% of pediatric factitious disorder cases and in less than 1% of hospital patients (Ehrlich, Pfeiffer, Salbach, Lenz, & Lehmkuhl, 2008; Ferrara et al., 2012). Munchausen by proxy is a form of child abuse in which caregivers produce or feign signs of illness, requiring unnecessary and invasive medical examinations and treatments. Alerting signs include multiple visits and admissions for medical care, improvement of symptoms with separation from the caregiver, and multiple moves from hospital to hospital to avoid detection or treatment (Ferrara et al., 2012, Rosenberg, 2003). Child victims of fabricated or induced illness have a high base rate of neurological diagnoses, including epilepsy, but other common presenting conditions such as anoxia or ataxia have been reported (Bass & Jones, 2011). Simulation that includes false reporting of symptoms or tampering with objective results, such as thermometer readings, is common.

Bass and Jones (2011) and Ferrara et al. (2012) provide compelling profiles of caregivers who engage in this form of child abuse. They are typically female with a high prevalence of personality disorders and histories of early abuse (sexual and physical), predisposing them to feigning symptoms to avoid abuse, and who have experienced caregiver loss or foster care and have somatoform illnesses themselves. As adolescents, many had been referred for disruptive behavior disorders, self-harm (including cutting), anxiety, depression, school refusal, eating disorders, abuse, and somatoform disorders. Pathological lying was a common finding, with antecedents in childhood or adolescence (Bass & Jones, 2011; Ferrara et al., 2012). Lying was described as habitual, compulsive, and self-aggrandizing (Bass & Jones, 2011).

Causes and Treatment of Factitious Disorder

Factitious disorder is a puzzling condition for most clinicians, and its causes are poorly understood, which probably contributes to its underdetection in clinical settings. Kozlowska (2014) posited that factitious disorders stem from early disruptions in the body's stress system in response to maltreatment, generating somatic symptoms. Difficulties with emotional regulation due to deprivation of nurturing attachments then arise, along with disturbance in the development of the self, leading to personality disorders and the enactment of unmet needs. She notes that factitious behaviors may be an extension of childhood behaviors whose original function was to secure comfort and protection from others. In typically developing children, rewards for the sick role are common, such as being allowed to stay home from school and getting more parental attention. Maladaptive parental reinforcement of the sick role may thus predispose to development of factitious disorder in vulnerable children or in children exposed to abuse or to caregivers with personality disorders. Last, because of a cultural stigma against mental illness, it is possible that factitious disorders provide an accepted and face-saving way of manifesting pathology that is reinforced by a medicalized health care system based on biological conceptualizations of illness. Kozlowska (2014) notes that drawing a clear line between volitional and unconscious behavior may be impossible in most cases.

Management of pediatric factitious disorder is optimized by a multidisciplinary team approach and a close collaboration with primary care physicians, using principles of constructive and supportive confrontation aimed at saving face and minimizing harm; many patients nevertheless are lost to follow-up (Bass & Halligan, 2014). In severe cases, removal from the home is necessary (Bass & Glaser, 2014). In one small case series of adolescent girls, the malingering cases had much better outcomes than those with factitious disorder (Peebles et al., 2005).

DISORDERS NOT DEFINED BY DECEPTION: CONDITIONS INVOLVING NEUROLOGICALLY OR MEDICALLY UNEXPLAINED SYMPTOMS

Conversion Disorder (Functional Neurological Symptom Disorder)

Conversion disorder, also known in DSM-5 as "functional neurological symptom disorder," is included under the broad DSM-5 category of somatic symptom and related disorders. Conversion disorder involves *voluntary motor* or *sensory* symptoms that are incompatible with recognized neurological or medical conditions, that are not better explained by another condition, and that cause impairment and distress. In pediatrics, cases of conversion disorder in children may be encountered by neuropsychologists who work in movement disorder clinics or epilepsy monitoring units, where a subgroup of children may present with unexplained paralysis or nonepileptic seizures. Conversion disorders may also be overrepresented in pain clinics. In children, conversion disorders tend to include only one symptom, are typically of abrupt onset, and are often associated with a precipitant event (Mink, 2013).

Cognitive symptoms are *not* included under the diagnosis of conversion disorder in DSM-5, although some experts believe they should have been (Stone et al., 2011). However, neuropsychologists may be calledon to assess symptoms relating to conversion disorder because the disorder can include a broad variety of pseudo-neurological symptoms that may be represented in children, including diminished consciousness or loss of speech, which occur in 20–30% of cases (Ani, Reading, Lynn, Forlee, & Garralda, 2013). Other common conversion symptoms in children include motor weakness, abnormal movements, nonepileptic seizures, and sensory problems such as loss of sight or hearing. Pain is also a common symptom, occurring in over 50% of cases (Ani et al., 2013). Notably, there may be a clear difference between nonepileptic seizures secondary to conversion disorder and fabricated seizures due to factitious disorder. Some authors recommend specifically assessing for deception and fabrication at the very beginning when nonepileptic seizures in children are suspected (Barber & Davis, 2002).

Childhood conversion disorder is very rare, with incidence rates of 1–4 cases per 100,000, with most cases being female, presenting with multiple symptoms, and with antecedent stressors such as school bullying, separation, and family conflict; median age is 11–12 (Ani et al., 2013; Kozlowska et al., 2007). The incidence of conversion disorder increases with age, with no children younger than 7 diagnosed in one large study (Ani et al., 2013).

Illness Anxiety Disorder (i.e., Hypochondriasis)

Persons presenting with illness anxiety disorder have an intense fear about having a medical disorder. DSM-5 defines illness anxiety disorder as a preoccupation with having or acquiring a serious illness, without any physical symptoms, or else preoccupation that is in excess of the symptoms that are present. The term *illness anxiety disorder* replaces the older term *hypochondriasis*. Illness anxiety disorder may occur in asymptomatic individuals without illness or may co-occur with a diagnosable medical condition. In one series, the majority of children with health anxiety did not have a medical condition (Rask, Elberling, Skovgaard, Thomsen, & Fink, 2012). Overreporting of symptoms may be due to attention seeking, hypervigilance, or catastrophic thinking.

The main symptom is excessive preoccupation with health that is not accounted for by other disorders, such as anxiety disorders or obsessive–compulsive disorder. However, illness anxiety disorder is thought to be a form of anxiety, with overfocus on physical symptoms and fears about disease. Like anxiety, it tends to follow a fluctuating course and is exacerbated by stress and difficult life events. In children, health anxiety is associated with more general anxiety symptoms, co-occurs with somatoform symptoms, and is likely a precursor of adult somatoform disorders (Rask et al., 2012; Wright & Asmundson, 2003). Health anxiety symptoms have been reported in children as young as 5 years of age, and severe health anxiety occurs in about 2% of young children (Rask et al., 2012). Like other disorders reviewed here, early identification in children may prevent escalation of hypochondriasis in adulthood (Wright & Asmundson, 2003).

In the neuropsychology realm, few studies of illness anxiety disorder exist (but see Boone, 2009). Illness anxiety disorder may present with an intense fear associated with cognitive symptoms. For example, after a minor concussion, an adolescent may have intense anxiety about having permanent brain damage and may overinterpret normal memory slips as evidence of memory loss. Illness anxiety disorder may also predispose to protracted or difficult recovery in patients with mild TBI who continue to report cognitive symptoms in the absence of any evidence of cognitive problems on objective evaluation. In the context of neuropsychological assessment, a review of preexisting history may reveal a long-standing anxiety disorder.

Illness anxiety disorder has relevance for SVT interpretation. Individuals with this condition may produce invalid, elevated, or questionable results on SVTs due to excessive symptom reporting in the absence of deception or external incentives.

Dissociative Amnesia

Dissociative amnesia is the inability to recall basic autobiographical information, most often either in the form of memory loss for a specific event or generalized amnesia for one's life history. In dissociative amnesia, the amnesia itself does not fit with test results or history and is therefore itself a medically or neurologically unexplained symptom. Dissociative amnesia is thought to fall under the larger umbrella of dissociative disorders in DSM-5, which places it apart from the other disorders discussed here.

According to the DSM-5, dissociative amnesia is associated with a history of trauma, abuse, self-harm, suicide, and high-risk behaviors. It may therefore involve overlap with personality disorders, particularly those of cluster B. However, dissociative amnesia is very rare and has not been well researched. It most often occurs in the setting of other psychiatric disorders and is reportedly more common in females and adolescents, with most individuals regaining memory soon after the precipitating event (Granacher, 2014).

Dissociative amnesia may be more commonly reported in criminal settings, in which it is invoked to reduce criminal responsibility, and therefore careful assessment must follow to differentiate it from malingering. However, dissociative amnesia also occurs outside of forensic settings. Kirkwood et al. (2010) describe one of the few pediatric cases of dissociative amnesia in the neuropsychological literature. Their case was a 16-year-old boy seen for neuropsychological assessment after unexplained medical symptoms and dramatic transient memory loss who passed all PVTs and did adequately on all aspects of the evaluation, including memory tests. Notably, cases in the literature are vanishingly small, and pediatric cases under 16 years of age are virtually unknown. There currently exist no conceptual models, diagnostic tests, or standard clinical approaches to properly diagnose dissociative amnesia (Granacher, 2014).

In most cases of bona fide brain injury, the likelihood of severe transient memory loss or loss of autobiographical memory is implausible. For this reason, unsophisticated attempts to malinger can mimic dissociative amnesia. As well, it is important not to confuse dissociative amnesia with retrograde amnesia and posttraumatic amnesia or with PTSD. For example, a motor vehicle collision might cause memory loss restricted to a specific event (i.e., the collision) because of trauma to the brain, but the inability to recall would not be due to dissociation. When the event includes very traumatic content (e.g., witnessing the death of a loved one or sustaining exceedingly severe injuries), it may be difficult to disentangle these factors.

In most assessments, the differential diagnosis is PTSD versus injury-related memory loss, and not dissociative amnesia. In the current DSM-5 nosology, dissociative amnesia should not be diagnosed when PTSD better accounts for symptoms.

TREATING CONDITIONS INVOLVING NEUROLOGICALLY OR MEDICALLY UNEXPLAINED SYMPTOMS

Injury, stress, hospitalization, and even litigation itself may exacerbate or trigger psychiatric symptoms in vulnerable examinees with subclinical mental health issues, including somatic symptom disorders such as factitious disorder or illness anxiety disorder, even when there is no brain injury per se. These disorders are generally poorly understood by health professionals (Hamilton et al., 2012) and carry pejorative connotations. However, effective behavioral approaches to treatment are available (Sharma & Manjula, 2013), so proper identification is important in neuropsychological assessment, ideally following evidence-based practices based on established definitions and conceptual models. For example, several approaches are available for treatment of childhood conversion disorder, including cognitive-behavioral therapy, family therapy, and physical therapy (Campo & Fritz, 2001; Mink, 2013).

A FINAL NOTE ON INTERPRETING INVALID TEST RESULTS

Invalid test results must always be interpreted in light of the examinee's unique individual, clinical, and environmental situation. The practitioner's role is to maintain professionalism and objectivity at all times and to refrain from negative judgment even in cases of deliberate fabrication. The use of correct terminology and precise diagnostic models is critical to advancing the assessment and treatment of these conditions.

REFERENCES

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: Author.

Ani, C., Reading, R., Lynn, R., Forlee, S., & Garralda, E. (2013). Incidence and 12-month outcome of non-transient childhood conversion disorder in the UK and Ireland. *British Journal of Psychiatry, 202*, 413–418.

Barber, M. A., & Davis, P. M. (2014. Fits, faints, or fatal fantasy?: Fabricated seizures and child abuse. *Archives of Diseases in Childhood, 86*, 230–233.

Bass, C., & Glaser, D. (2014). Early recognition and management of fabricated or induced illness in children. *Lancet, 383*, 1412–1421.

Bass, C., & Halligan, P. (2014). Factitious disorder and malingering: Challenges for clinical assessment and management. *Lancet, 383*, 1422–1432.

Bass, C., & Jones, D. (2011). Psychopathology of perpetrators of fabricated or induced illness in children: Case series. *British Journal of Psychiatry, 199*, 113–118.

Booksh, R. L., Pella, R. D., Singh, A. N., & Gouvier, W. D. (2010). Ability of college students to simulate ADHD on objective measures of attention. *Journal of Attention Disorders, 13*, 325–338.

Boone, K. B. (2009). Fixed belief in cognitive dysfunction despite normal neuropsychological scores: Neurocognitive hypochondriasis? *The Clinical Neuropsychologist, 23,* 1016–1036.

Briere, J. (1996). *Trauma Symptom Checklist for Children.* Odessa, FL: Psychological Assessment Resources.

Campo, J. V., & Fritz, G. (2001). A management model for pediatric somatization. *Psychosomatics, 42*, 467–476.

Chafetz, M. D. (2008). Malingering on the Social Security Disability consultative exam: Predictors and base rates. *The Clinical Neuropsychologist, 22, 529–546.*

Chafetz, M., & Prentkowski, E. (2011). A case of malingering by proxy in a Social Security Disability psychological consultative examination. *Applied Neuropsychology, 18,* 143–149.

Delis, D. C., & Wetter, S. R. (2007). Cogniform disorder and cogniform condition: Proposed diagnoses for excessive cognitive symptoms. *Archives of Clinical Neuropsychology, 22,* 589–604.

Ehrlich S., Pfeiffer, E., Salbach, H., Lenz, K., & Lehmkuhl, U. (2008). Factitious disorder in children and adolescents: A retrospective study. *Psychosomatics, 49,* 392–398.

Ferrara, P., Vitelli, O., Bottaro, G., Gatto, A., Liberatore, P., Binetti, P., et al. (2012). Factitious disorders and Munchausen syndrome: The tip of the iceberg. *Journal of Child Health Care, 17,* 366–374.

Gaughwin, P. (2008). Psychiatry's problem child: PTSD in the forensic context: Part 1. *Australian Psychiatry, 16*, 104–108.

Granacher, R. P. (2014). Commentary: Dissociative amnesia and the future of forensic psychiatric assessment. *Journal of the American Academy of Psychiatry and the Law, 42,* 14–18.

Gray, M. J., Elhai, J. D., & Briere, J. (2010). Evaluation of the Atypical Response scale of the Trauma Symptom Inventory—2 in detecting simulated posttraumatic stress disorder. *Journal of Anxiety Disorders, 24,* 447–451.

Hamilton, J. C., Eger, M., Razzak, S., Feldman, M. D., Hallmark, N., & Cheek, S. (2012). Somatoform, factitious, and related diagnoses in the National Hospital Discharge Survey: Addressing the proposed DSM-5 revision. *Psychosomatics, 54,* 142–148.

Harrison, A. G., Rosenblum, Y., & Currie, S. (2010). Examining unusual digit span performance in a population of postsecondary students assessed for academic difficulties. *Assessment, 17,* 283–293.

Jasinski, L. J., Harp, J. P., Berry, D. T., Shandera-Ochsner, A. L., Mason, L. H., & Ranseen, J. D. (2011). Using symptom validity tests to detect malingered ADHD in college students. *Clinical Neuropsychologist, 25*, 1415–1428.

Kirk, J. W., Hutaff-Lee, C. F., Connery, A. K., Baker, D. A., & Kirkwood, M. W. (2014). The relationship between the self-report BASC-2 validity indicators and performance validity failure after pediatric mild traumatic brain injury. *Assessment, 21*(5), *562–569.*

Kirkwood, M. W., & Kirk, J. W. (2010). The base rate of suboptimal effort in a pediatric mild TBI sample: Performance on the Medical Symptom Validity Test. *The Clinical Neuropsychologist, 24*, 860–872.

Kirkwood, M. W., Kirk, J. W., Blaha, R. Z., & Wilson, P. (2010). Noncredible effort during pediatric neuropsychological exam: A case series and literature review. *Child Neuropsychology, 16*, 604–618.

Kozlowska, K. (2014). Abnormal illness behaviors: A developmental perspective. *Lancet, 383*, 1368–1369.

Larrabee, G. J. (2012). Performance validity and symptom validity in neuropsychological assessment. *Journal of the International Neuropsychological Society, 18,* 625–631.

Lu, P. H., & Boone, K. B. (2002). Suspect cognitive symptoms in a 9-year-old child: Malingering by proxy. *The Clinical Neuropsychologist, 16,* 90–96.

Lubit, R., Hartwell, N., van Gorp, W., & Eth, E. (2002). Forensic evaluation of trauma syndromes in children. *Child and Adolescent Psychiatric Clinics of North America, 11*, 823–857.

McCaffrey, R. J., & Lynch, J. K. (2009). Malingering following a documented brain injury: Neuropsychological evaluation of children in a forensic setting. In J. E. Morgan & J. J. Sweet (Eds.), *Neuropsychology of malingering casebook* (pp. 77–385). New York: Psychology Press.

Marshall, P. S., Schroeder, R., O'Brien, J., Fischer, R., Ries, A., Blesi, B., et al. (2010). Effectiveness of symptom validity measures in identifying cognitive and behavioral symptom exaggeration in adult attention deficit hyperactivity disorder. *The Clinical Neuropsychologist, 24*, 1204–1237.

Mehta, S., & Ameratunga, S. N. (2012). Prevalence of post-traumatic stress disorder among children and adolescents who survive road traffic crashes: A systematic review of the international literature. *Journal of Paediatric Child Health, 48,* 876–885.

Mink, J. W. (2013). Conversion disorder and mass psychogenic illness in child neurology. *Annals of the New York Academy of Sciences, 1304*, 40–44.

Musso, M. W., & Gouvier, W. D. (2014). "Why is this so hard?": A review of detection of malingered ADHD in college students. *Journal of Attention Disorders, 18*, 186–201.

Peebles, R., Sabella, C., Franco, K. F., & Goldfarb, J. (2005). Factitious disorder and malingering in adolescent girls: Case series and literature review. *Clinical Pediatrics, 44*, 237–243.

Rabiner, D. L. (2013). Stimulant prescription cautions: Addressing misuse, diversion and malingering. *Current Psychiatry Reports, 15,* 375.

Rask, C. U., Elberling, H., Skovgaard, A. M., Thomsen, P. H., & Fink, P. (2012). Parental-reported health anxiety symptoms in 5- to 7-year-old children: The Copenhagen Child Cohort CCC 2000. *Academy of Psychosomatic Medicine, 53*, 58–67.

Rosenberg, D. A. (2003). Munchausen syndrome by proxy: Medical diagnostic criteria. *Child Abuse & Neglect, 27*, 421–430.

Sharma, M. P., & Manjula, M. (2013). Behavioral and psychological management of somatic symptom disorders: An overview. *International Review of Psychiatry, 25,* 116–124.

Slick, D. J., & Sherman, E. M. S. (2012). Differential diagnosis of malingering and related clinical presentations. In E.M.S. Sherman & B.L. Brooks (Eds.), *Pediatric forensic neuropsychology* (pp. 113–135). New York: Oxford University Press.

Slick, D. J., Sherman, E., & Iverson, G. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *Clinical Neuropsychologist, 13,* 545–561.

Stone, J., LaFrance, W. C., Brown, R., Spiegel, D., Levenson, J. L., & Sharpe, M. (2011). Conversion disorder: Current problems and potential solutions for DSM-5. *Journal of Psychosomatic Research, 71,* 369–376.

Sullivan, B. K., May, K., & Galbally, L. (2007). Symptom exaggeration by college adults in attention-deficit hyperactivity disorder and learning disorder assessments. *Applied Neuropsychology, 14*, 189–207.

Walker, J. S. (2011). Malingering in children: Fibs and faking. *Child Adolescent Psychiatric Clinics of North American, 20,* 547–556.

Wright, K. D., & Asmundson, G. J. G. (2003). Health anxiety in children: Development and psychometric properties of the Childhood Illness Attitude Scales. *Cognitive Behavioral Therapy, 32*, 194–202.

3

Understanding Deception from a Developmental Perspective

ERIC PETERSON
ROBIN L. PETERSON

From the earliest scientific study of the child (e.g., Darwin, 1877; Hartshorne & May, 1928) through the work of Piaget (e.g., 1965), the development of deception has been recognized as an important lens through which to examine social-cognitive development. In the developmental literature, the study of deception has largely focused on verbal deception, that is, making verbal statements with the intentional goal of instilling in another person a false sense of reality or, more simply put, lying (e.g., Lewis, 1993; Lee, 2013).[1] Across the past two decades or so, there has been a resurgence of research on deception motivated by both practical and theoretical interests. The rise of child witnesses in the legal system created an urgent demand for insight into children's potential for misrepresenting their understanding of reality (Malloy, Johnson, & Goodman, 2013). Clinicians who seek verbal reports from the child as an index of the child's neurocognitive status have also recognized the importance of a richer theoretical understanding of the child's capacity for deception. As some clinical scientists (DeRight & Carone, 2013; Kirkwood, Kirk, Blaha, & Wilson,

[1] An important aspect of the development of deception concerns lying for prosocial reasons (e.g., to avoid harming an individual or group). Prosocial lying has been well studied in the deception literature (e.g., Bussy, 1992, 1999; Fu, Evans, Wang, & Lee, 2008; Poplinger, Talwar, & Crossman, 2011; Williams, Kirmayer, Simon, & Talwar, 2013). However, such research is tangential to the primary goals of the current chapter and so is not reviewed here.

2010) have suggested, our theoretical understanding of deception and its early emergence may not be widely appreciated by clinicians who must rely on the child's best truthful effort. The resurgence of developmental investigations of deception also reflects a growing interest in two fundamental developments of early childhood: theory of mind (for a review, see Stone, 2006) and executive functions (for a review, see Garon, Bryson, & Smith, 2008).

Developmental scientists have extensively studied both theory of mind and executive functioning. The degree to which the two cognitive domains are intertwined has been the subject of debate, with some arguing that theory of mind represents a highly specialized conceptual development (e.g., Leslie & Thaiss, 1992; Scholl & Leslie, 1999) and others arguing that executive development is centrally involved in the emergence of theory of mind (e.g., Carlson, Claxton, & Moses, 2015). As should be clear throughout this chapter, the child's growing capacity to lie effectively surely involves both theory of mind and executive functioning. Thus we begin with a brief overview of development in these two domains. Following this background review, we turn to our primary topic: the development of deception. Today, solid empirical research provides a nuanced view of the young child's ability to deceive into the early school years and beyond (e.g., Lee, 2013; Talwar & Lee, 2008). Finally, we consider some questions of practical importance: How well can adults detect children's deceit, and what factors influence the ability to know when children are lying?

THEORY OF MIND AND EXECUTIVE FUNCTIONS: CRITICAL DEVELOPMENTS FOR SUCCESSFUL DECEPTION

Evidence emphasizing the mechanistic roles of both theory of mind and executive functions in the development of deception has burgeoned in the last 20 years. Theory of mind refers to the appreciation that others have mental states that can be different from one's own (Stone, 2006). People perform actions based on their own beliefs about reality even when their beliefs are false. Thus, if I believe that some coffee remains in the pot even when there is no more coffee, I may reach for the pot to pour some. At the same time, my spouse may look on, understanding both my goal and why I was disappointed. Executive functions refers to a suite of cognitive processes (such as planning, working memory, mental flexibility, and inhibition) that support goal-directed behavior (e.g., Welsh, Pennington, & Grossier, 1991; Welsh & Pennington, 1988; Zelazo & Müller, 2002; Zelazo, Qu, & Müller, 2005). For the purpose of understanding the development of deception, inhibition and working memory are of particular relevance and are emphasized in this chapter.

Some have argued that children's capacity for deception *is* their emerging theory of mind at work (e.g., Chandler, Fritz, & Hala, 1989; Peskin, 1992; Sodian, Taylor, Harris, & Perner, 1991). Others (e.g., Carlson, Moses, & Hix, 1998) have argued that the role of executive functions in the development of deception is much more primary. Given the broader disagreement about the relationship between executive functions and theory of mind, we will not argue that one or the other competence is primary. Our view is that these two developments together are central to understanding the emergence of deception and its increasing sophistication with age. Clearly, a child's ability to willfully present a false story, however flimsy, must be related to an appreciation that mental states are private and not necessarily tied to observable behavior. The ability to maintain the lie, resisting the urge to blurt out incriminating information, must involve executive processes. We look first at the development of theory of mind and then turn to executive functions.

Theory of Mind

The study of theory of mind began with the goal of precisely comparing the relative social-cognitive understanding of human and nonhuman primates (Premack & Woodruff, 1978). Dennett (1978) argued for understanding "false belief" as a gold-standard criterion for defining and assessing the human capacity for theory of mind. In the classic false-belief paradigm (Wimmer & Perner, 1983), a young child watches a puppet show in which a character places an object in a hidden location (e.g., a cupboard or a box) and then leaves. While this first character remains out of sight of the object, a second, devious character arrives and moves the object to a new location. After the object is relocated, the first character returns to look for the object. To pass the task, the child must appreciate that the first character's action will be motivated by a false belief rather than by the true state of affairs. The first character believes the object remains in the first location and will initially look there. In essence, the child must demonstrate metarepresentation—the capacity to represent another person's mental representation of the world. Analyses of hundreds of false-belief tasks have yielded a remarkably consistent description of the development of theory of mind by about 4 years of age (Wellman, Cross, & Watson, 2001).

Since the advent of the seminal false-belief task, many different conceptually related paradigms have been developed, yielding a coherent pattern (for a review, see Stone, 2006). By about age 2, children are able to understand the psychology of desire. So, for example, a toddler who prefers trucks to dolls can appreciate that another child with the opposite preference will choose the doll. Understanding desire (Jane likes the doll and she will choose what she wants) clearly requires a degree of mental state understanding. However, this achievement falls short of the more cognitively demanding appreciation that others can hold false beliefs. Consistent with the notion that desire is less demanding than belief, in the next section we show that the toddler's first false statements may reflect awareness of the child's own desire rather than anyone's belief status.

Perhaps the most striking evidence suggesting that children younger than the age of about 4 struggle to understand false beliefs can be observed in a task that examines awareness of a child's *own* false beliefs (Wimmer & Hartl, 1991). In the unexpected-contents paradigm, the child is presented with a known container (e.g., a bag of M&Ms) and is asked what is in the bag. Invariably, the child will make the correct inference. Then the child is allowed to spill out the contents, only to discover unexpected contents (e.g., pennies rather than M&Ms). Now the child is asked what she or he thought was in the bag a moment ago. The preschooler is likely to report, "pennies," reflecting a difficulty uncoupling belief from reality. From this evidence, it would seem that children who cannot yet pass false belief have no insight into the notion that minds—others' or one's own—can have an incorrect understanding of the world (a prerequisite for deception). However, clever research paradigms leave no doubt that this conclusion is too simplistic. Tasks that do not involve explicit verbal responses yield evidence that some understanding of false belief can be demonstrated behaviorally (Garnham & Perner, 2001). Indeed, Onishi and Baillargeon (2005) adapted a classic false-belief task in order to examine whether 15-month-old toddlers' nonverbal behavior reveals understanding. Even at this young age, evidence is consistent with some understanding of false belief that is not accessible to explicit verbal response. It would seem reasonable to guess that acts of deception that do not require verbal expression (e.g., deceptive behavior rather than lying) may be apparent at an earlier age than that at which children can pass the false-belief task. Indeed, in the next section, we show that choice of response system in a deception paradigm plays an important role in the outcome.

Since the creation of the original false-belief paradigm, researchers have adapted the task to examine "second order" false-belief understanding. Second-order false-belief tasks require the child to make a judgment about how yet another observer will evaluate the mental status of an individual (e.g., John believes that Mary believes that the cookies are in the box). The more advanced theory of mind ability necessary for these tasks is first evident by about 6 years of age and continues to develop through adolescence (Perner & Wimmer, 1985; Sullivan, Zaitchik, & Tager-Flusberg, 1994).

To summarize, with the use of many different but closely related paradigms, the child's gradual acquisition of theory of mind (e.g., desire- vs. belief-based reasoning, first- vs. second-order theory of mind) has been well described. Some degree of theory of mind can be observed behaviorally well before the preschool milestone passing of the gold-standard false-belief task. A few years after passing the first-order false-belief task, children are able to demonstrate second-order false-belief understanding. As is evidenced later, each of these sequential steps in theory of mind development has been associated with an increasingly sophisticated capacity for deception. However, across every stage of theory of mind development, it is difficult to disambiguate the role of conceptual change in mental state understanding from that of improved executive functions. For example, success in the classic false-belief task requires the child to update the location of the hidden object in working memory, as well as to inhibit a prepotent response (the response that the child knows to reflect the true state of affairs) when asked where the character will search. We turn now to this second development that is central to the child's growing capacity to deceive.

Executive Functions

The term *executive functions* refers to a set of cognitive processes that together support goal-directed behavior and that are particularly important in performing novel tasks that require the ability to override routinized, automatic responses (for reviews, see Best & Miller, 2010; Diamond, 2013; Garon et al., 2008). Although the nature of executive processes has been debated (e.g., Miyake & Friedman, 2012; O'Reilly, 2011; Peterson & Welsh, 2014), there is general agreement that working memory, inhibition, and shifting are core executive processes (Miyake et al., 2000). When considering the development of deception, a role for both inhibition and working memory has clear face validity (for review, see Gombos, 2006). Lying involves generating a plausible but false statement and maintaining this line of thought in working memory separate from the true state of affairs. Once the lie has been initiated, successful deception may involve controlling later nonverbal and verbal communication, particularly inhibiting a prepotent response that would take the form of subsequent "semantic leakage" (revealing information consistent with the true state of affairs rather than with the lie that has been told).

The relationship between working memory and inhibition has been the subject of some disagreement in the literature. Some researchers (e.g., Munakata et al., 2011) have argued that inhibition is itself a product of the working memory system, whereas others have argued that the constructs are distinct (Miyake et al., 2000). For our purposes, we consider each of these executive constructs as separate while acknowledging that they typically operate together and are mutually supportive (Diamond, 2013). Working memory can be defined generally as the ability to hold and manipulate information in conscious awareness. Inhibition encompasses a number of separate inhibitory processes. Complex response inhibition, which involves generating a response while inhibiting a conflicting prepotent response, is of particular importance for deception.

Both working memory and inhibitory control can be observed in infancy and continue to mature over many years. Garon and colleagues (2008) reviewed evidence supporting significant maturational change in both working memory and complex response inhibition across the preschool years, which aligns nicely with the period of greatest change in the development of deception. Using age-appropriate tasks, investigations have identified the ongoing development of both working memory and inhibition into adolescence (Diamond, 2013). In the case of working memory, tasks that require either holding many items online or performing mental manipulation (as might be most relevant in the case of lying) show the most protracted development. Inhibitory development has been demonstrated beyond early childhood and into adolescence using a range of tasks that emphasize different aspects of inhibition. Notably, no single task measuring inhibition is developmentally appropriate for use across this wide age span, so researchers have often been challenged to disambiguate true developmental change from task-specific effects (Huizinga, Dolan, & van der Molen, 2006). Table 3.1 provides a brief overview of tasks that have been developed or adapted to examine working memory and inhibition from preschool onward.

To summarize, both working memory and inhibition emerge in infancy and develop significantly throughout the preschool years and into adolescence. The steady improvement of executive functioning across this age span parallels the earlier discussion of maturation of theory of mind. However, unlike the case of measuring theory of mind development, no tasks can be identified for which performance can be related to a specific change in deception ability (as was the case in the review of first- and second-order theory of mind). This may reflect the general limitations of the executive-functions literature. As discussed in many reviews (e.g., Huizinga et al., 2006; Miyake & Friedman, 2012), available executive tasks likely involve numerous nonexecutive cognitive processes (the "task impurity" problem), and the specific cognitive skills emphasized vary with tasks appropriate for different developmental levels (Garon et al., 2008). Thus evidence garnered with a particular task does not always generalize to other tasks that purportedly tap the same underlying executive construct.

Although executive functions have historically been studied in decontextualized, nonemotional (i.e., "cool") settings, researchers have recently begun to explore the possibility of "hot" executive processes that may mediate performance in contexts characterized by greater motivational or emotional significance. Although the hot versus cool executive functioning framework remains unresolved (for reviews, see Peterson & Welsh, 2014; Welsh & Peterson, 2014), the consideration of context is intriguing for the case of deception. It makes sense to posit that generating lies (presumably in the presence of adults) involves managing an emotional context (see Carlson et al., 1998, for a discussion of the "social intimidation" that may

TABLE 3.1. A Sampling of Tasks Used to Examine Working Memory and Inhibition in Children

Working memory

Working memory involves maintaining either verbal or spatial information in awareness in the absence of perceptual input and manipulating the information. All reasoning must involve working memory.

- ***Backward Digit Span.*** Children listen to a list of digits and then repeat back the list in reverse order.
- ***Reordering Span Tasks.*** Children listen to a list of items and must reorder them (e.g., in numerical order).
- ***Corsi Block Span.*** Children observe an adult tap a series of blocks and then try to tap the blocks, reproducing the sequence.
- ***Backward Corsi Span.*** Follows the same procedure at the Corsi Span except that children reproduce the sequence in reverse order.

Inhibition

Inhibition involves controlling behavior or mental processes so as to avoid making a habitual or impulsive response (i.e., focusing on one aspect of the environment while ignoring task-irrelevant distractions, choosing a thoughtful and appropriate response rather than blurting out whatever “comes to mind”).

- ***Delay of Gratification.*** Children are presented with a treat and asked to wait to take it in exchange for a larger amount of the treat.
- ***Day/Night.*** Children see either a sun or a moon and must generate a mismatched response (saying “night” to the sun or “day” to the moon).
- ***Dimensional Change Card Sort.*** Children sort cards that vary on two dimensions (shape and color). After sorting by one dimension (e.g., color), the child is asked to switch sorting rules (e.g., to shape).
- ***Shape Stroop.*** In this adaptation of the classic Stroop interference paradigm, children see pictures of small fruit embedded within larger fruit and must point to the smaller rather than the larger fruit.
- ***Antisaccade.*** Children see a stimulus presented peripherally and must make a saccade away from the stimulus.
- ***Go/no-go.*** In most trials children see a stimulus that indicates a button press response (*go*). However, in some trials, a stimulus is presented that requires the inhibition of any response (*no-go*).
- ***Flanker task.*** Children see a target (e.g., an arrow pointing left or right) that is “flanked” by distractors on each side and are asked to identify the target. Response conflict occurs when the distracting flanker stimuli are mapped to another response.

Note. For a more in-depth review of working memory and inhibition, as well as the relevant tasks, see Best and Miller (2010), Diamond (2013), and Garon, Bryson, and Smith (2008).

affect lying performance). Although many fundamental questions remain about the possibility of both hot and cool executive systems, it may be worthwhile to consider both the context (i.e., situational factors) and the individual (i.e., individual-specific factors such as temperament and personality that affect the degree to which an individual is more or less vulnerable to a hot context) in future child deception research.

From False Statements to Unsophisticated Lies to Successful Deception

As mentioned previously, the literature on deception in early child development has focused largely on evidence of deception in the child's explicit verbal report (e.g., Ahern, Lyon, & Quas, 2011; Evans & Lee, 2013; Lee, 2013; Polak & Harris, 1999; Talwar & Lee, 2002, 2008). In experimental settings, successful lying appears to emerge in the preschool years, around the same time that children pass classic false-belief tasks. However, consistent with the evidence that false-belief understanding can be observed with implicit behavioral measures much earlier in toddlerhood, some studies have demonstrated that children between 2 and 3 years of age, like nonhuman primates (Hare, Call, & Tomasello, 2006), will perform deceptive actions (e.g., concealing the location of a toy; Chandler et al., 1989; Sodian et al., 1991). Further, some researchers have detailed their own children's ability to knowingly produce untruthful statements before the age of 3 (e.g., Darwin, 1877; Newton, Reddy, & Bull, 2000). We imagine that many parents appreciate the fun young children can experience in making consciously false statements and acting is if they are true ("Buses drive on the ceiling!") Recently, our younger child, not yet 3, demonstrated what seemed to be clear lying behavior—denying pulling a plant out of the garden while holding the evidence—albeit without sufficient guile to avoid revealing incriminating information (where in the garden the plant came from). "Semantic leakage," the subsequent revelations of inculpatory information following the initial untrue statement, is a marker of early lies (e.g., Talwar & Lee, 2008). Our own parental observations are consistent with at least one home observational study (Wilson, Smith, & Ross, 2003), in which 65% of 2-year-olds demonstrated the capacity to lie.

Of course, observations of children in natural contexts offer a limited source of evidence, with little opportunity to either verify the original judgment of the behavior or probe the possible constraints on the child's ability to produce false statements. In an early classic observational study, Stern and Stern (1909, reviewed in Ahern et al., 2011) referred to "pseudolies," which are essentially inaccurate statements that seem to be a temporary error in the child's thinking, perhaps reflecting either an instant of confusion about imagination versus reality or a failure of impulse control. Thus anecdotal observations of children in natural contexts, ours and others', are very difficult to interpret with confidence. To date, we are aware of just two studies (Ahern et al., 2011; Evans & Lee, 2013) that have explored whether 2-year-olds will produce lies using a well-controlled paradigm (reviewed subsequently). However, even methodologically sound studies designed to observe lying behavior are limited with children this young. Below the age of 3, children are susceptible to "desire-based" responses in which their statements reflect how they wish they had performed (e.g., "I wish I hadn't pulled this plant from the garden") rather than an intentional deception (Ahern et al., 2011).

Although very little evidence has been accrued with children below age 3, many well-controlled studies (e.g., Evans & Lee, 2011; Lewis, Stanger, & Sullivan, 1989; Polak & Harris, 1999; Talwar, Gordon, & Lee, 2007; Talwar & Lee, 2002, 2008, 2011) have explored the propensity to lie beginning in preschool. These studies have relied on the temptation resistance paradigm, originally developed by Sears, Rau, and Alpert (1965). In this procedure, children are typically told by an adult not to peek at or play with a toy, and then they are left alone in the room with a video recorder running. Upon return, the adult asks the children whether they either peeked at or played with the toy. This paradigm has several strengths (Talwar & Lee, 2008). It reliably yields responses and associated behaviors that can reasonably be judged as lies, and it approximates the real-world conditions under which children make false statements. Next we consider the current evidence concerning developmental changes in lying behavior beginning at age 2, most of which comes from the temptation resistance paradigm.

Talwar and Lee (2008) outlined a framework for understanding the development of lying (see Table 3.2). In their model, acts of verbal deception can be segregated into three developmental levels: primary, secondary, and tertiary lies. Between the ages of 2 and 3, primary lying behavior emerges. At this point, children intentionally make untrue statements. However, their verbal deceptions may not reflect an awareness of the listener's mental state but a wish regarding their behavior (i.e., a desire-based lie; Ahern et al., 2011). To date, just one study (Evans & Lee, 2013) has examined lying behavior among 2-year-olds using a temptation resistance paradigm. Following the general procedure, children had an opportunity to lie about peeking behavior (providing a denial) while the experimenter was absent. Then they were asked a follow-up question about the toy's location to assess their ability to avoid semantic leakage in order to maintain the lie. Finally, children completed a battery of executive functioning tasks and a language measure. Lying behavior was observed in about 25% of the 2-year-old children, whereas a majority of 3-year-olds produced a lie to conceal their peeking behavior. Further, overall executive function skill was significantly related to lying. Among the subset of children who lied (n = 21), only three children avoided semantic leakage. It should be noted that this study did not enable disambiguation of primary lies (desire based) from secondary lies, which reflect intentional deception. Several other studies have also demonstrated that by 3 years a substantial portion of children will lie to conceal a transgression (Evans & Lee, 2013; Lewis et al., 1989; Fu, Evans, Xu, & Lee, 2012; Talwar & Lee, 2002). Throughout the preschool years, the ability to deceive becomes more prevalent and more sophisticated (for a review, see Lee, 2013).

TABLE 3.2. The Development of Verbal Deception within the Talwar and Lee Three-Stage Framework

Stage 1: Primary lies (2–3 years)	Stage 2: Secondary lies (3–5 years)	Stage 3: Tertiary lies (beyond age 5)
	Description of Lying Ability	
A minority of 2-year-olds and a majority of 3-year-olds are capable of lying. However, at this stage the child's lie is hypothesized to reflect a desire ("I wish I hadn't peeked") rather than a belief.	Children produce lies with the intent to deceive. However, in spite of this intention, they still fail to prevent semantic leak age (i.e., revealing incriminating information).	Children are now capable of lying more effectively by maintaining a consistent account and preventing semantic leakage.
	Theory of Mind Development	
Children are capable of desire-based reasoning but not explicit belief-based reasoning. They can pass implicit but not explicit tests of false belief.	Between about 3 and 4 years of age, children acquire a "theory of mind" as defined by the ability to pass an explicit false-belief task.	Between about ages 5 and 7, the child develops the ability to pass "second order" theory of mind tests. Second-order belief performance improves into adolescence.
	Key Studies of Verbal Deception	
From 2 to 3 years of age: • Ahern, Lyon, & Quas (2011) • Evans & Lee (2013)	From 3 to 5 years of age: • Lewis, Stanger, & Sullivan (1989) • Polak & Harris (1999) • Talwar & Lee (2002, 2008, 2011) • Fu, Evans, Xu, & Lee (2012)	Beyond age 5: • Talwar & Lee (2002, 2008) • Talwar, Gordon, & Lee (2007) • Evans & Lee (2011)

The second stage of Talwar and Lee's (2008) model, secondary lies, emerges at approximately age 4 and shows steady development over the following years (see Table 3.2). At this developmental level, the child's false statements reflect an understanding of the listener's private mental state. Clearly, it makes sense to hypothesize that the emergence of secondary lies coincides with the acquisition of full theory of mind as defined by the ability to pass explicit false-belief tasks. Further, given that these tasks also tap working memory and inhibition, we might expect that executive functioning is also mechanistically involved in the changes that are apparent in the early preschool years. Polak and Harris (1999) implemented a modified temptation resistance paradigm that involved two conditions: one in which peeking was permitted and another in which it was prohibited. Comparison of rates of honest versus deceptive answers across the two conditions supported the conclusion that children's denials of peeking were intentional deceptive acts reflecting an awareness of the adult's mental state (see also Chandler et al., 1989). This investigation included theory of mind measures in order to formally assess the relationship between lying behavior and mental state understanding. Consistent with the hypothesized demand characteristics of secondary lying, children with higher theory of mind scores were significantly more likely to lie than those with lower theory of mind scores.

For children younger than age 5 who have achieved theory of mind, lying behavior typically remains unsophisticated. In the Polak and Harris (1999) study, even children with high theory of mind scores failed to maintain consistency following the lie (i.e., by feigning ignorance of knowledge gained while peeking). In a later study, Talwar and Lee (2002) demonstrated that although naïve adults were not able to differentiate liars from nonliars on the basis of nonverbal behavior, the children's explicit verbal responses to subsequent questions provided semantic leakage. Interestingly, development of deception in young children is influenced not only by child-level cognitive factors but also by social context. Talwar and Lee (2011) compared 3- and 4-year-old West African children from the same neighborhoods who attended either a school that utilized punitive discipline or one that used nonpunitive discipline. The preschoolers who experienced a punitive environment were much more likely to lie to conceal their own transgressions. Further, the lies of the children in the punitive environment were more sophisticated, with some ability to prevent semantic leakage. Thus exposure to a punitive environment appears to accelerate this aspect of social-cognitive development, although it also is associated with more negative psychosocial outcomes (Gershoff, 2002).

In Talwar and Lee's (2008) model, the ability to control behavior after lying and to minimize leakage is the hallmark of the third stage, tertiary lies (see Table 3.2). Throughout the school-age years, children improve considerably in their deceptive skill as they become better able to maintain consistency following a lie so as to avoid self-incrimination (Talwar & Lee, 2008). Polak and Harris (1999) hypothesized that, whereas first-order theory of mind skill would predict the likelihood of lying, second-order theory of mind ability would be associated with a decline in semantic leakage. Using a temptation resistance paradigm, Talwar and colleagues (Talwar & Lee, 2008; Talwar et al., 2007) explored this hypothesis in children ranging from 3 to 11 years. Several key findings emerged from these studies. First, although most preschool children peeked at the hidden object, only about half of school-age children did so, probably reflecting development in inhibitory control. Second, across age groups, most of those who peeked lied about doing so, consistent with the notion that lying to conceal a transgression represents a normative behavior through most of childhood. Comparison of results across these and other studies suggests that asking children to promise to tell the truth reduces lying behavior, though it remains prevalent. Third, better semantic leakage control was associated both with age and with second-order theory of mind ability. When older children named the hidden toy (a tactical error in an attempt at deception), they were more likely than younger children to provide a plausible explanation for their knowledge. Finally, performance in the Stroop interference task (which taps both working memory and inhibition) correlated with the child's lying behavior, though two other executive tasks did not.

Limited methodologically rigorous work has examined the development of deception into adolescence. Most extant studies that focused on this developmental period have utilized self-report methods, which have obvious limitations for studying dishonest behavior! Available evidence supports the conclusion that the sophistication of deception continues to increase into adolescence, consistent with the ongoing maturation of executive functioning and higher order mental state understanding skills during this developmental period. Evans and Lee (2011) used a modified version of the temptation resistance paradigm to examine the development of lying from age 8 to age 16. Participants were asked to answer a number of trivia questions for a reward. They were left alone in a room with a booklet containing the answers and asked not to peek. Consistent with the age trend described earlier in school-age children, the prevalence of peeking behavior declined with age, to a low of 40% in 14- to 16-year-olds. Among those who peeked, tendency to lie also decreased with age, from about 70% in the youngest age group to about 25% in the oldest age group. Tendency to lie was not associated with executive functioning skill, and overall, this study did not elucidate the reason for age-related decreases in verbal deception. The authors suggested that the change might reflect either moral development or an increased awareness of the likelihood of being caught. Adults lie frequently in everyday situations (e.g., DePaulo, Kashy, Kirkendol, Wyer, & Epstein, 1996), and conceptual and moral understanding of lies appears similar between adolescents and adults (Talwar & Crossman, 2012). Thus the observed decrease in lying behavior with age in this specific experimental setting may not carry over to naturalistic settings. A final important finding from this study was that performance on some executive functioning measures (including a Stroop task and the Tower of London task, which emphasizes planning, problem solving, and working memory) was associated with increased sophistication of lying in the form of decreased semantic leakage, again consistent with results from younger children.

In addition to sophistication and consistency of lies, other aspects of deceptive behavior continue to develop through the school years and into adolescence. These include the ability to control nonverbal behavior (e.g., facial expression, gaze cues) while lying (McCarthy & Lee, 2009; Talwar & Lee, 2002) and to make strategic or tactical decisions about lying (Smith & LaFreniere, 2011). The extent to which these developments represent relatively specific advances in deceptive skills rather than broad shifts in conceptual ability remains unclear.

It also appears likely that the primary motivation behind lying behavior shifts with age. Preschool children frequently lie to conceal their own transgressions (Lee, 2013). Under some circumstances, children will lie to conceal the transgression of another. Talwar, Lee, Bala, and Lindsay (2004) examined 3- to 11-year-old children's willingness to lie to conceal a parent's transgression and found no relationship with age. However, an early study reported that 12-year-olds were more likely than 8-year-olds to lie to conceal the transgression of a peer (Greenglass, 1972). Further, over the elementary school years, children become more willing to tell prosocial or "white" lies at some personal cost (Poplinger, Talwar, & Crossman, 2011).

In summary, by age 2, a subset of children is capable of lying to conceal a transgression. Lying behavior increases significantly over the preschool years, a development that relates to growth in executive functions and theory of mind. Lying to conceal a transgression remains normative through the school years. At least in experimental settings, this behavior decreases somewhat in adolescence, though the reason for this trend is not known. Throughout middle childhood and adolescence, children demonstrate an increasingly sophisticated control of verbal and nonverbal deceptive behaviors, and these developments appear linked to improved inhibition and working memory.

ADULTS' ABILITY TO DETECT DECEPTION: THE EXPERIMENTAL EVIDENCE

The degree to which adults can accurately detect children's lies is of great practical importance, as children's verbal reports represent a key source of evidence in legal, educational, and—of most relevance to this book—clinical settings. Lay adults are quite poor at detecting lies in other adults, typically performing around chance levels (Bond & DePaulo, 2006). Based on the evidence reviewed herein that children's sophistication in lying steadily improves with age, we might predict that adults can readily detect the lies of young children. However, empirical work paints a more complex picture. Generally speaking, children's lies do become less transparent with age, but under some circumstances, even young children's lies can be difficult to detect (Crossman & Lewis, 2006). Child, observer, and situational factors all influence the likelihood that adults will be able to know when children are lying and when they are telling the truth.

Early research reported that adults detected younger children's lies better than those of adolescents or other adults (e.g., Feldman, Jenkins, & Popoola, 1979). However, this work explicitly instructed participants to make up false stories. It is easy to imagine that in comparison to naturalistic settings, this paradigm might introduce increased cognitive load for young children and thus diminish the effectiveness of their deceit. Recent research has shown that in more ecologically valid settings (e.g., variations on the temptation resistance paradigm), adults have difficulty discriminating lies from truth in children as young as 3 (Talwar & Lee, 2002; Crossman & Lewis, 2006). However, likelihood of detection is moderated by the type of information available to adult observers. Most adults cannot reliably discriminate preschool lying from truth telling based on nonverbal behavior alone (Lewis et al., 1989; Talwar & Lee, 2002; though see Talwar, Crossman, Gulmi, Renaud, & Williams, 2009), nor on nonverbal behavior in conjunction with simple yes/no responses (e.g., "Did you peek?"; Crossman & Lewis, 2006). On the other hand, consistent with the preceding evidence for semantic leakage, adults are better able to detect deception in transcripts of longer narratives from young children or when children are asked probing or unexpected follow-up questions. These circumstances yield the expected age-of-child effects, in which adults easily detect lies made by preschool children with poor leakage control but not those made by school-age children (Talwar & Crossman, 2012). Semantic leakage control emerges around age 6 and improves steadily through early adolescence (Talwar et al., 2007).

Not all adults are equally capable of detecting children's deception. For example, teachers and other professionals who work with children were more accurate than other adults in detecting lies of 3- to 6-year-old children (Talwar, Crossman, Williams, & Muir, 2011; Crossman & Lewis, 2006), suggesting that increased experience with children is an advantage. As children get older and lying behavior becomes more sophisticated, even professionals can be expected to have difficulty discriminating accurate from deceitful statements, similar to what has been reported for detecting lies made by adults (e.g., Vrij, 2008). Parents are generally better at detecting deception in children than nonparents but may have a truth bias when it comes to their own children (Talwar & Crossman, 2012). Age of the observer can also influence the pattern of results. With age, adults become more accurate and have a more conservative response bias when detecting children's deceit (Block et al., 2011) and are better overall at discriminating truth and lie tellers (Talwar et al., 2011). Again, this ability suggests a relationship between the amount of experience with children and the ability to detect their deception. Interestingly, older children were more accurate than both younger children and young adults at detecting deception in one study (Talwar et al., 2009). Perhaps this edge was the result of children spending a significant amount of time with other children their age in school, as well as their own emerging ability to tell more sophisticated lies.

Results regarding gender effects for the child and observer have been mixed. Some studies found no differences in likelihood of detection according to gender (Crossman & Lewis, 2006). However, some gender differences have emerged on specific measures. Girls' antisocial lies were more easily discriminated than their prosocial lies (Talwar et al., 2011). Observers were more accurate at detecting boys' lies than girls' lies but were less accurate at detecting boys' truths compared to girls' truths (Talwar et al., 2009). Despite most adults' poor ability to discriminate children's lies from truth telling, adults' overall confidence in their judgment of the veracity of children's statements is high (Saykaly et al., 2013). Men generally have higher confidence ratings in their judgments than women (Block et al., 2011), and confidence does not appear to be related to accuracy in judgments (Stromwall, Granhag, & Landstrom, 2007).

To summarize, adults often think they can readily discriminate children's lies from their true statements, but in reality, this task is frequently difficult, even in very young children. Successful detection is more likely when cognitive load is increased, presumably by taxing underlying executive skills. For example, the child may be asked a series of follow-up questions, or working memory demands may be manipulated (e.g., by asking the child to tell the story in reverse order; Talwar & Crossman, 2012). Eliciting a promise from the child to tell the truth leads to both an increase in the likelihood of a truthful report and greater accuracy in detection of lies (Talwar & Crossman, 2012).

CONCLUSION

Throughout this chapter, we have focused on the development of verbal deception with the primary goals of describing both the overall age-related changes in deceptive ability and the cognitive underpinnings that support that developmental change. It is clear that lying is a normative behavior that emerges, on average, at about 3 years of age. Across the preschool years lying behavior becomes more prevalent. Both theory of mind and executive development have been shown to support increasing sophistication of lying behavior, although it may be simplistic to argue that one or the other plays the more primary role given our limited understanding. Although there are fewer methodologically rigorous studies of lying behavior in adolescence, the evidence suggests that the tendency to lie declines during this period but that lies that are made are likely as sophisticated as those of adults.

A secondary goal of the chapter was to examine the degree to which adults can accurately detect deception. This question has important consequences in a number of settings. Of particular interest to the readers of this volume, it is important to appreciate the degree to which a clinician might be duped by a child's attempt to provide misleading verbal responses. Overall, the evidence for poor deception detection may be surprising to some. Although there are factors associated with better detection (e.g., experience with children and increased cognitive load for the child), adults are, on average, poor detectors, and their feelings of confidence are unrelated to accuracy in knowing when children are telling the truth.

Perhaps one of the biggest limitations in the current literature reflects the contexts in which lying behavior has been explored. The temptation resistance paradigm enables researchers to probe lying behavior in a tightly controlled environment. However, there is an obvious problem with generalizability. It may well be that lying behavior in some real-world contexts reveals important differences. For example, in the temptation resistance paradigm, adolescents who peek are less likely than younger children to produce a lie when confronted. Lying may be less prevalent at this age due to moral development or to an increased awareness of the chances of being caught, as the proponents of this paradigm have suggested. However, it may also be true that adolescents simply understand the low-stakes nature of this experimental setting well and are less concerned than younger children about the need to cover their transgressions. Although the experimental context may reasonably approximate the settings in which young children frequently lie, it likely has poorer external validity for the contexts that at least some teens face (e.g., those surrounding legal or high-stakes educational decisions).

Thus a number of important theoretical and applied issues surrounding the development of deception and its cognitive underpinnings remain to be resolved by future research. What is not up for debate, however, is whether children are capable of deceiving adults. Lying is a normative childhood behavior, and even young children frequently dupe adults in roles of authority. In fact, the ability to deceive is a key milestone in early social-cognitive development and part of what separates the preschooler from nonhuman primates. As will be clear throughout the remainder of this volume, these results highlight the need for clinical evaluation of children to include objective measures of performance and response validity.

REFERENCES

Ahern, E. C., Lyon, T. D., & Quas, J. A. (2011). Young children's emerging ability to make false statements. *Developmental Psychology*, *47*, 61–66.

Best, J. R., & Miller, P. H. (2010). A developmental perspective on executive function. *Child Development, 81*, 1641–1660.

Block, S. D., Shestowsky, D., Segovia, D. A., Goodman, G. S., Schaaf, J. M., & Alexander, K. W. (2011). "That never happened": Adults' discernment of children's true and false memory reports. *Law and Human Behavior*, *36*, 365–374.

Bond, C. F., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review, 10*, 214–234.

Bussey, K. (1992). Lying and truthfulness: Children's definitions, standards, and evaluative reaction. *Child Development*, *63*, 129 –137.

Bussey, K. (1999). Children's categorization and evaluation of different types of lies and truths. *Child Development, 70*, 1338 –1347.

Carlson, S. M., Claxton, L. J., & Moses, L. J. (2015). The relation between executive function and theory of mind is more than skin deep. *Journal of Cognition and Development, 16,* 186–197.

Carlson, S. M., Moses, L. J., & Hix, H. R. (1998). The role of inhibitory processes in young children's difficulties with deception and false belief. *Child Development, 69*, 672–691.

Chandler, M. J, Fritz, A. S., & Hala, S. M. (1989). Small-scale deceit: Deception as a marker of two-, three-, and four-year-olds' early theories of mind. *Child Development, 60*, 1263–1277.

Crossman, A. M., & Lewis, M. (2006). Adults' ability to detect children's lying. *Behavioral Sciences and the Law, 24*, 703–715.

Darwin, C. (1877). A biographical sketch of an infant. *Mind*, *2*, 285–294.

Dennett, D. (1978). Beliefs about beliefs. *Behavioral Brain Sciences, 1*, 568–570.

DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., & Epstein, J. A. (1996). Lying in everyday life. *Journal of Personality and Social Psychology, 70*, 979–995.

DeRight, J., & Carone, D. A. (2013). Assessment of effort in children: A systematic review. *Child Neuropsychology, 21,* 1–24.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64*, 135–168.

Evans, A. D., & Lee, K. (2011). Verbal deception from late childhood to middle adolescence and its relation to executive functioning skills. *Developmental Psychology, 47*, 1108–1116.

Evans, A. D., & Lee, K. (2013). Emergence of lying in very young children. *Developmental Psychology*, *49*, 1958–1963.

Feldman, R. S., Jenkins, L., & Popoola, O. (1979). Detection of deception in adults and children via facial expressions. *Child Development, 50*, 350–355.

Fu, G, Evans, A. D., Wang, L., & Lee, K. (2008). Lying in the name of the collective good: A developmental study. *Developmental Science, 11*(4), 495 – 503.

Fu, G., Evans, A. D., Xu, F., & Lee, K. (2012). Young children can tell strategic lies after committing a transgression. *Journal of Experimental Child Psychology, 113*, 147–158.

Huizinga, M., Dolan, C. V., & van der Molen, M. W. (2006). Age-related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia*, *44*, 2017–2036.

Garnham, W. A., & Perner, J. (2001). Actions really do speak louder than words—but only implicitly: Young children's understanding of false belief in action. *British Journal of Developmental Psychology, 19*, 413–432.

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin, 134*, 31–60.

Gershoff, E. T. (2002). Corporal punishment by parents and associated child behaviors and experiences: A meta-analytic and theoretical review. *Psychological Bulletin, 128*, 539–579.

Gombos, V. A. (2006). The cognition of deception: The role of executive processes in producing lies. *Genetic, Social, and General Psychology Monographs*, *132*, 197–214.

Greenglass, E. R. (1972). Effects of age and prior help on "altruistic lying." *Journal of Genetic Psychology, 121*, 303–313.

Hare, B., Call, J., & Tomasello, M. (2006). Chimpanzees deceive a human by hiding. *Cognition, 101*, 495–514.

Hartshorne, H., & May, M. A. (1928). *Studies in the nature of character: Vol. 1. Studies in deceit*. New York: Macmillan.

Kirkwood, M. W., Kirk, J. W., Blaha, R. Z., & Wilson, P. (2010). Noncredible effort during pediatric neuropsychological exam: A case series and literature review. *Child Neuropsychology, 16*, 604–618.

Lee, K. (2013). Little liars: Development of verbal deception in children. *Child Development Perspectives*, *7*, 91–96.
Leslie, A. M., & Thaiss, L. (1992). Domain specificity in conceptual development: Neuropsychological evidence from autism. *Cognition, 43*, 225–251.
Lewis, M. (1993). The development of deception. In M. Lewis & C. Saarni (Eds.), *Lying and deception in everyday life* (pp. 90–105). New York: Guilford Press.
Lewis, M., Stanger, C., & Sullivan, M. (1989). Deception in 3-year-olds. *Developmental Psychology*, *25*, 439–443.
Malloy, L. C., Johnson, J. L., & Goodman, G. S. (2013). Children's memory and event reports: The current state of knowledge and best practice. *Journal of Forensic Social Work, 3*, 106–132.
McCarthy, A., & Lee, K. (2009). Children's knowledge of deceptive gaze cues and its relation to their actual lying behavior. *Journal of Experimental Child Psychology*, *103*, 117–134.
Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science, 21*, 8–14.
Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive function and their contribution to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*, 49–100.
Munakata, Y., Herd, S. A., Chatham, C. H., Depue, B. E., Banich, M. T., & O'Reilly, R. C. (2011). A unified framework for inhibitory control. *Trends in Cognitive Science, 15*, 453–459.
Newton, P., Reddy, V., & Bull, R. (2000). Children's everyday deception and performance on false-belief tasks. *British Journal of Developmental Psychology, 18*, 297–317.
O'Reilly, R. C. (2011). The what and how of prefrontal cortical organization. *Trends in Neuroscience, 33*, 355–361.
Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science, 308*, 255–258.
Peskin, J. (1992). Ruse and representation: On children's ability to conceal information. *Developmental Psychology, 28*, 84–89.
Perner, J., & Wimmer, H. (1985). "John thinks that Mary thinks that . . . ": Attribution of second-order beliefs by 5- to 10-year-old children. *Journal of Experimental Child Psychology, 39*, 437–471.
Peterson, E., & Welsh, M. C. (2014). The development of hot and cool executive functions in childhood and adolescence: Are we getting warmer? In S. Goldstein & J. A. Naglieri (Eds.), *Handbook of executive functions* (pp. 45–68). New York: Springer.
Piaget, J. (1965). *The moral judgment of the child* (M. Gabain, Trans.). Glencoe IL: Free Press.
Polak, A., & Harris, P. L. (1999). Deception by young children following noncompliance. *Developmental Psychology, 35*, 561–568.
Poplinger, M., Talwar, V., & Crossman, A. (2011). Predictors of children's prosocial lie-telling: Motivation, socialization variables and moral understanding. *Journal of Experimental Child Psychology, 110*, 373–392.
Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*, 515–526.
Saykaly, C., Talwar, V., Lindsay, R. C., Bala, N. C., Lee, K., Bertrand, M., et al. (2013). Adults' ability to detect deception of stressful and non-stressful stories of children. *Psychology, Crime and Law*, *19*, 865–879.
Scholl, B. J., & Leslie, A. M. (1999) Modularity, development, and "theory of mind." *Mind and Language, 14*, 131–153.

Sears, R., Rau, L., & Alpert, R. (1965). *Identification and child rearing.* New York: Wiley.
Smith, R. M., & LaFreniere, P. J. (2011). Development of tactical deception from 4 to 8 years of age. *British Journal of Developmental Psychology*, *31*, 30–41.
Sodian, B., Taylor, C., Harris, P., & Perner, J. (1991). Early deception and the child's theory of mind: False trails and genuine markers. *Child Development, 62*, 468–483.
Stone, V. (2006). Theory of mind and the evolution of social intelligence. In J. T. Cacioppo, P. S. Visser, & C.L. Picket (Eds.), *Social neuroscience: People thinking about thinking people* (pp. 103–129). Cambridge, MA: MIT Press.
Stromwall, L. A., Granhag, P. A., & Landstram, S. (2007). Children's prepared and unprepared lies: Can adults see through their strategies? *Applied Cognitive Psychology*, *21*, 457–471.
Sullivan, K., Zaitchik, D., & Tager-Flusberg, H. (1994). Preschoolers can attribute second-order beliefs. *Developmental Psychology*, *30*, 395–402.
Talwar, V., & Crossman, A. M. (2012). Children's lies and their detection: Implications for child witness testimony. *Developmental Review*, *32*, 337–359.
Talwar, V., Crossman, A. M., Gulmi, J., Renaud, S., & Williams, S. (2009). Pants on fire?: Detecting children's lies. *Applied Developmental Science*, *13,* 119–129.
Talwar, V., Crossman, A., Williams, S., & Muir, S. (2011). Adult detection of children's selfish and polite lies: Experience matters. *Journal of Applied Social Psychology*, *41*, 2837–2857.
Talwar, V., Gordon, H. M., & Lee, K. (2007). Lying in the elementary school years: Verbal deception and its relation to second-order belief understanding. *Developmental Psychology*, *43*, 804–810.
Talwar, V., & Lee, K. (2002). Development of lying to conceal a transgression: Children's control of expressive behaviour during verbal deception. *International Journal of Behavioral Development*, *26*, 436–444.
Talwar, V., & Lee, K. (2008). Social and cognitive correlates of children's lying behavior. *Child Development*, *79*, 866–881.
Talwar, V., & Lee, K. (2011). A punitive environment fosters children's dishonesty: A natural experiment. *Child Development*, *82*, 1751–1758.
Talwar, V., Lee, K., Bala, N., & Lindsay, R. C. L. (2004). Children's lie-telling to conceal a parent's transgression: Legal implications. *Law and Human Behavior, 28*, 411–435.
Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities*. Chichester, UK: Wiley.
Wellman, H., Cross, D., & Watson, J. (2001). Meta-analysis of theory of mind development: The truth about false belief. *Child Development*, *72*, 655–684.
Welsh, M. C., & Pennington, B.F. (1988). Assessing frontal lobe functioning in children: Views from developmental psychology. *Developmental Neuropsychology*, *4*, 199–230.
Welsh, M. C., Pennington, B.F., & Grossier, D.B. (1991). A normative-developmental study of executive function: A window on prefrontal function in children. *Developmental Neuropsychology*, *7*, 131–149.
Welsh, M. C., & Peterson, E. (2014). Issues in the conceptualization and assessment of hot executive functions in childhood. *Journal of the International Neuropsychological Society*, *20*, 152–156.
Williams, S.M., Kirmayer, M., Simon, T., & Talwer, V. (2013). Children's antisocial and prosocial lies to familiar and unfamiliar adults. *Infant and Child Development, 22*, 430–438.
Wilson, A. E., Smith, M. D., & Ross, H. D. (2003). The nature and effects of young children's lies. *Social Development, 12*, 21–45.

Wimmer, H., & Hartl, M. (1991). Against the Cartesian view on mind: Young children's difficulty with own false beliefs. *British Journal of Developmental Psychology, 9*, 125–138.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13,* 103–128.

Zelazo, P. D., & Müller, U. (2002). Executive function in typical and atypical development. In U. Goswami (Ed.), *Blackwell handbook of childhood cognitive development* (pp. 445-469). Oxford, UK: Blackwell.

Zelazo, P. D., Qu, L., & Müller, U. (2005). Hot and cool aspects of executive function: Relations in early development. In W. Schneider, R. Schumann-Hengsteler, & B. Sodian (Eds.), *Young children's cognitive development: Interrelationships among executive functioning, working memory, verbal ability, and theory of mind* (pp. 71–93). Mahwah, NJ: Erlbaum.

4

Performance and Symptom Validity

A Perspective from the Adult Literature

GLENN J. LARRABEE

Psychologists and neuropsychologists have long recognized the need to evaluate both performance validity and symptom validity in the course of conducting psychological and neuropsychological evaluations. Our diagnostic tools require both accurate report of actual symptom experience on tests such as the Minnesota Multiphasic Personality Inventory–2—Restructured Form (MMPI-2-RF; Tellegen & Ben-Porath, 2008) and a full and complete effort on tests of ability such as the Wechsler Adult Intelligence Scale—Fourth Edition (WAIS-IV; Wechsler, 2008). Clearly, both symptom report and fully engaged performance on tests of ability can be controlled by the examinee. Consequently, one needs to evaluate both symptom validity with symptom validity tests (SVTs) to assess whether examinees are providing an accurate reporting of their actual symptom experience and performance validity with performance validity tests (PVTs) to assess whether examinees are providing an accurate measure of their actual abilities (Larrabee, 2012). In this chapter, I use the PVT and SVT distinction to refer to assessment procedures that previously have been characterized as measures of "symptom validity," "response bias," or "effort." The rationale for the PVT and SVT distinction is provided by Larrabee (2012). This chapter refers to the adult literature on PVTs, SVTs, and diagnosis of malingering.

THE HISTORY OF VALIDITY TESTING IN ADULT NEUROPSYCHOLOGY

Formal examination of performance validity in neuropsychological evaluation began with Rey's work with the Dot Counting Test (Rey, 1941) and the 15-Item Test (Rey, 1964) and with Spreen and Benton's work on simulation of mental deficiency on the Benton Visual Retention Test (BVRT; Benton & Spreen, 1961; Spreen & Benton, 1963). These early investigations demonstrate two features of PVTs: (1) PVTs can be stand-alone measures of performance validity that address only the issue of performance validity but do not evaluate other basic cognitive functions (Dot Counting Test, Rey 15-Item Test), or (2) PVTs can be derived from clinically atypical patterns of performance on standard neuropsychological tests such as the BVRT. Early stand-alone PVTs such as Dot Counting and the Rey 15-Item were designed to be sufficiently easy so that persons with bona fide neurological impairment could perform normally on these tasks. Benton and Spreen (1961; Spreen & Benton, 1963) also attempted to show that the patterns of performance of simulated mental deficiency on the BVRT were atypical compared with the performance of patients with actual clinical impairment, but variability in performance in the clinical patients precluded recommendation for routine clinical application of these patterns of simulated impairment.

In the 1970s, Pankratz, Fausti, and Peed (1975) demonstrated how a procedure based on application of the normal approximation to the binomial to two-alternative forced-choice testing could address the validity of an examinee's claimed sensory impairment, in this case, feigned deafness. This study influenced the development of many subsequent freestanding PVTs for evaluation of feigned memory impairment that used a two-alternative forced-choice stimulus presentation, including the Portland Digit Recognition Test (Binder, 1990), Test of Memory Malingering (TOMM; Tombaugh, 1996), Victoria Symptom Validity Test (VSVT; Slick, Hopp, Strauss, & Spellacy, 1996), and Word Memory Test (WMT; Green, Iverson & Allen, 1999).

The need to use a formal PVT assessment rather than relying upon clinical judgment to detect feigned impairment was demonstrated over 30 years ago by a study conducted by Heaton, Smith, Lehman, and Vogt (1978). These investigators found that the accuracy of a group of trained neuropsychologists ranged from 50% correct to 69% correct in discriminating feigned patterns of performance on the Halstead–Reitan Battery from patterns that were characteristic of severe traumatic brain injury (TBI). By contrast, a discriminant function equation correctly classified 100% of the cases on the basis of their Halstead–Reitan Battery and WAIS performance, and 94% on the basis of their MMPI profile scores. The participants feigning impairment did worse on the Speech-Sounds Perception Test, the Finger

Oscillation Test, finger agnosia, sensory suppressions, hand dynamometer, and WAIS Digit Span compared with the patients with severe TBI; these patterns have been replicated in subsequent investigations (Greiffenstein, Baker, & Gola, 1996; Mittenberg, Rotholc, Russell, & Heilbronner, 1996; Ross, Putnam, Millis, Adams, & Krukowski, 2006).

Other research on development of embedded/derived PVTs over the past 15 years has led to identification of invalid performance, with a particular focus on the recognition trials of verbal supraspan learning tests such as the California Verbal Learning Test—Second Edition (CVLT-II; Root, Robbins, Chang, & van Gorp, 2006) and the Rey Auditory Verbal Learning Test (Boone, Lu, & Wen, 2005; Barrash, Suhr, & Manzel, 2004; Davis, Millis, & Axelrod, 2012). Others have identified PVTs for visual memory tests, including atypical recognition errors on the Continuous Visual Memory Test (Larrabee, 2009) and atypical recognition errors and poor copy performance on the Rey Complex Figure Test (Reedy et al., 2013). Abnormally low performance on Finger Tapping has been identified as a PVT (Arnold et al., 2005; Larrabee, 2003), as has atypically poor visual discrimination on Benton Visual Form Discrimination (Larrabee, 2003) or on WAIS-III Picture Completion (Solomon et al., 2010). Atypical errors on problem-solving tests rarely seen in patients with severe TBI have also been identified for the Wisconsin Card Sorting Test (Larrabee, 2003) and for the Category Test (Tenhula & Sweet, 1996). Multivariable logistic regression equations have been identified for invalid performance patterns based on multiple scores from individual tests such as the CVLT-II (Wolfe et al., 2010) and Auditory Verbal Learning Test (AVLT; Davis et al., 2012) and based on multiple tests such as the five procedures comprising the Advanced Clinical Solutions (2009) effort tests (Miller et al., 2011). Sufficient numbers of embedded/derived PVT procedures exist to continuously sample performance validity throughout the neuropsychological evaluation for measures of perception, motor function, working memory, processing speed, verbal and visual learning and memory, and problem-solving skills (Boone, 2009).

Formal evaluation of the validity of symptom report also dates to the 1940s, with the original publication of the MMPI (Hathaway & McKinley, 1943). In its original format, there were three validity scales: "Cannot Say" (representing the unanswered items), L, and F. Subsequent research on the F scale demonstrated its particular relevance to detection of malingering of severe psychopathology. In this regard the F scale demonstrates one of the key features of detection of invalid symptom report: overendorsement of rare symptoms.

Over time, it became obvious that symptom exaggeration on omnibus personality tests such as the MMPI can manifest in other ways than through the overreporting of rarely endorsed symptoms characteristic of exaggeration of severe psychiatric disturbance. Lees-Haley, English, and Glenn (1991) developed the Fake Bad Scale (FBS) Symptom Validity Scale to detect exaggeration of nonpsychotic emotional and physical complaints; the Fs Scale was developed for the MMPI-2-RF to detect somatic symptom overreporting (Ben-Porath & Tellegen, 2008; Tellegen & Ben-Porath, 2008); and the Response Bias scale (RBS) of the MMPI-2-RF was devised to detect symptom overreporting in examinees who failed PVTs (Gervais, Ben-Porath, Wygant, & Green, 2007). Additionally, the F scale has been augmented by the Fp scale, developed to detect items rarely endorsed by an inpatient psychiatric sample (Arbisi & Ben-Porath, 1995). A cogent review of the MMPI/MMPI-2 and MMPI-2-RF validity scales is provided by Ben-Porath (2012).

THE IMPORTANCE OF OBJECTIVE VALIDITY TESTING

Failure to take into account performance validity can distort relationships with external criteria (Rohling et al., 2011). Grade point average did not show the expected relationship with IQ until those failing a PVT were excluded (Greiffenstein & Baker, 2003). Olfactory identification did not show the expected correlation with severity of TBI until those failing PVTs were excluded (Green, Rohling, Iverson, & Gervais, 2003). The median memory test correlation with hippocampal volume in mild cognitive impairment was .49 for those passing PVTs but –.11 in those failing PVTs (Rienstra et al., 2013). Memory performance did not differ between examinees with and without computerized tomography (CT) scan abnormalities until those failing PVTs were excluded (Green, 2007). Memory complaints only correlated with memory performance in those failing PVTs (Gervais, Ben-Porath, Wygant, & Green, 2008). A group of participants with TBI/neurological impairment did not differ in neuropsychological test performance from a group of participants with mild TBI, psychiatric disorder, or chronic pain until those participants who failed PVTs were excluded (Green, Rohling, Lees-Haley, & Allen, 2001). Neuropsychological test performance was associated with the presence or absence of brain injury only in those participants who passed PVTs (Fox, 2011).

Base Rates of Validity Test Failure

Failure rates of PVTs and SVTs in settings with external incentives are high. My review of 11 studies covering 1,363 cases, primarily with alleged mild TBI, showed a 40% failure rate (Larrabee, 2003). Ardolf, Denney, and Houston (2007) found that 54% of criminal litigants met diagnostic criteria for either probable or definite malingering. Chafetz (2008) found that 49% of Social Security claimants met criteria for either probable or definite malingering. Greve et al. (2006) found that 40% of persons with claims for toxic or environmental exposure met diagnostic criteria for probable or definite malingering. Rates of invalid performance and diagnoses of malingering were sufficiently consistent, within the range of 30–50%, for my colleagues and I to propose a new "magical number" of 40% ± 10 to represent the base rate of malingering in settings with external incentive (e.g., personal injury litigation, criminal prosecution, and disability claims; see Larrabee, Millis, & Meyers, 2009). At these base rates, the cost to society is very substantial. For example, Chafetz and Underhill (2013) provide data yielding estimated Social Security Disability costs of malingered mental disorders of $20.022 billion and of malingered musculoskeletal disorder (commonly associated with chronic pain) of $14.176 billion per year at a malingering base rate of 40%.

Interest and Consensus in the Field

There has been an explosion in research on PVTs and SVTs in the field of neuropsychology, starting in the early 1990s and accelerating since that time. Sweet and Guidotti Breting (2013) conducted a search using the key word *malingering* and found one publication in 1985, 18 in 1990, 66 in 2000, and 91 in 2010; in the first 6 months of 2011, the figure was 60. These authors also noted the appearance of 13 meta-analyses over the period of 1991 to 2011 and eight edited textbooks from 2001 to 2010. In 1999, Slick, Sherman, and Iverson published their landmark paper on diagnosis of malingered neurocognitive dysfunction (MND), and their criteria were adapted by Bianchini, Greve, and Glynn (2005) to the assessment of malingered pain-related disability (MPRD). In 2005 the National Academy of Neuropsychology published a position paper noting that symptom exaggeration or fabrication occurred in a sizeable minority of examinations, with a greater prevalence in forensic contexts, and advised that adequate assessment of performance and symptom validity was essential to maximizing confidence in results and that such assessment was medically necessary (Bush et al., 2005). In 2009, the American Academy of Clinical Neuropsychology published a consensus statement on the neuropsychological assessment of effort, response bias, and malingering based on the combined efforts of 30 recognized experts in this area (Heilbronner et al., 2009). The consensus statement focused on five topics: (1) definition and differential diagnosis, (2) ability issues, (3) somatic issues, (4) psychological issues, and (5) research evidence/scientific issues. This consensus statement concluded that assessing performance and symptom validity and determining the presence of malingering were important and necessary activities of clinicians, particularly in secondary gain contexts but also in routine clinical practice.

RESEARCH AND CLINICAL DIAGNOSTIC ISSUES

Research Design

Two basic research designs for creating PVTs and SVTs have been employed in the adult literature: the simulation design and the "known groups" or criterion groups design (Larrabee, 2007; Rogers, 1997). In the simulation design, two groups are compared on the PVT or SVT being developed: (1) a group of noninjured persons asked to believably feign deficits and (2) a group of participants who have identified neurological injury (e.g., moderate to severe TBI) who do not have any external incentive. In the "known groups" or criterion groups design, two groups are compared on the PVT/SVT being developed: (1) a litigating/compensation-seeking group with no known neurological deficits (e.g., uncomplicated mild TBI; normal CT scan; no or brief loss of consciousness; posttraumatic amnesia less than 24 hours) who have also failed two or more PVTs/SVTs independent of the PVT/SVT being investigated and (2) a group of participants with identified neurological injury (e.g., moderate to severe TBI) who do not have any external incentive.

False Positives

There are two keys to both simulation and criterion groups designs. The first is that the clinical comparison group must contain persons who have undeniably significant bona fide neuropsychological impairment. The second is that the cut score for the PVT or SVT being examined must be set so that a minimum of the bona fide clinical group, usually 10%, is misidentified; that is, the false-positive rate is 10% or less. Because the probability of any diagnosis, not just malingering, is based on (true positives) ÷ (true positives plus false positives), the smaller the false-positive rate, the greater the diagnostic probability (Baldessarini, Finkelstein, & Arana, 1983). Moreover, modern investigators are encouraged to report the characteristics of the cases misidentified as false positives to enhance future case-specific application of the PVT/SVT; that is, if the case being examined with the PVT does not show a clinical history associated with false-positive identification, it likely is not a false-positive result. Because the cut scores are set to keep false positives at a minimum, sensitivity is typically lower than specificity, as seen in the meta-analyses conducted by Vickery, Berry, Inman, Harris, and Orey (2001), who reported an average sensitivity of .56 and specificity of .95, and Sollman and Berry (2011), who reported a mean sensitivity of .69 with a mean specificity of .90.

Given the focus on minimizing PVT false-positive rates in patients with bona fide disorder (e.g., severe TBI) resulting in significant neurological impairment, most PVTs and SVTs are easy to pass for patients who have genuine central nervous system dysfunction. A review of the TOMM manual (Tombaugh, 1996) shows an average correct performance on Trial 2 of 98.7% for aphasic patients and 98.3% correct for patients with TBI who experienced anywhere from 1 day up to 3 months of coma. Indeed, one TBI patient had 38 days of coma, underwent right frontal lobectomy for the effects of a gunshot wound, and scored perfectly on TOMM Trial 2. In another investigation, Goodrich-Hunsaker and Hopkins (2009) found that three patients with radiologically confirmed bilateral hippocampal damage due to hypoxia all passed the PVT trials of the WMT (Green, 2003). Using an embedded/derived PVT, Silverberg and Barrash (2005) found that 94% of patients with intractable epilepsy, 57% with documented left temporal lobe seizure focus, and 43% with right temporal lobe seizure focus passed the PVT index for the AVLT, despite averaging a T-score of 39 on the long delay recall trial of the AVLT.

Performance on PVTs also does not appear to be affected by depression, anxiety, acute pain, or chronic pain. Cold-pressor stimulation does not affect performance on Reliable Digit Span or on the TOMM (Etherton, Bianchini, Ciota, & Greve, 2005; Etherton, Bianchini, Greve, & Ciota, 2005). Depression (Rees, Tombaugh & Boulay, 2001), depression and anxiety (Ashendorf, Constantinou, & McCaffrey, 2004), and depression with chronic pain (Iverson, Le Page, Koehler, Shojania, & Badii, 2007) do not affect performance on the TOMM. One hundred percent of fibromyalgia and rheumatoid arthritis patients not seeking compensation passed the Computerized Assessment of Response Bias, with 96% passing the WMT (Gervais et al., 2001).

Classification Statistics

Diagnostic statistics for PVTs and SVTs produce comparable if not superior positive likelihood ratios (LR+) to those found for other diagnostic contrasts. LR+ is defined as (sensitivity) ÷ (1 – specificity), or sensitivity/false alarm rate (Grimes & Schulz, 2005). LR+ gives the likelihood that the score came from the group with the condition of interest (COI), as opposed to the group without the COI. LR+ multiplied by the base rate odds gives the posttest odds, which can be converted to a diagnostic probability by the formula (odds) ÷ (odds + 1). Based on the meta-analysis of Vickery et al. (2001), the LR+ is .56/.05, or 11.2. Based on the Sollman and Berry (2011) meta-analysis, LR+ is .69/.10, or 6.9. These values compare quite favorably to the LR+ for detection of brain dysfunction on the Halstead–Reitan Average Impairment Rating, with .771/.146 yielding an LR+ of 5.28 (Heaton, Miller, Taylor, & Grant, 2004). The finding of superior LR+ for the discrimination of feigned from legitimate impairment, in comparison with the discrimination of abnormal from normal neuropsychological performance, is a testament to the successful development of PVTs with the goal of minimizing false-positive error rates.

As noted before, individual PVT cutoffs are typically set to keep the false-positive rate at 10% or less, resulting in a lower sensitivity relative to specificity. Diagnostic accuracy can be improved by use of multiple, independent PVTs. LR+ can be chained if the individual PVTs and SVTs are independent and either weakly correlated or uncorrelated (Grimes & Schulz, 2005). For example, if one has two independent PVTs, each with a sensitivity of .50 and specificity of .90, the LR+ for each will be .50/(1 – .90), or 5.0. If the base rate of malingering is .40, the base rate odds of malingering are .40/(1 – .40), or .67. If the LR+ for the first PVT is multiplied by the base rate odds of .67, this yields posttest odds of 3.35, which convert to a probability of 3.35/4.35, or .77. If both PVTs are used, the posttest odds of 3.35 are now multiplied by the new LR+ of 5.0 (for the second PVT), to yield new posttest odds of (3.35) (5.0), or 16.75, which converts to a probability of 16.75/17.75, or .94. I provide further discussion of this process in Larrabee (2008), in which I also argue that for most persons not engaging in invalid performance, the intercorrelations of PVTs are typically negligible owing to the fact that performance is at ceiling with restriction in range of scores, both factors which reduce PVT intercorrelation.

Use of Multiple Validity Tests

Some researchers have raised concerns over the possibility that administration of multiple PVTs and SVTs can cause an unwanted increase in the false-positive rate, over and beyond the per-test false-positive rate (Berthelson, Mulchan, Odland, Miller, & Mittenberg, 2013). Berthelson et al. attempted to directly test this possibility by using Monte Carlo simulated data, based on an average PVT intercorrelation of .31 that was determined from a meta-analysis they conducted. Their simulated data showed a significant increase in false alarm rate with increasingly larger sets of PVTs, each with either a .15 or .10 per-test false-positive rate. For example, using a per-test false-positive rate of 10% yielded a false-positive rate of 11.5% for a failure of two or more out of five PVTs, a value not much larger than the per-test rate of 10%. When the failure rate was two out of seven, however, the false-positive rate increased to 17.5%.

Because the Berthelson et al. (2013) analyses were based on simulated rather than actual data, Davis and Millis (2014) analyzed the association of the number of PVTs administered with number who failed in a sample of 158 outpatient physiatric referrals, not screened for medicolegal involvement, and found a small, insignificant correlation, $r_s = .13$, $p = .10$, between the number of PVTs failed and the number administered. In a subsample of 87 neurological cases not in litigation, false-positive rates were lower than those of Berthelson et al. (2013) for participants administered six, seven, or eight PVTs using a cutoff of failure of ≥ 2 PVTs. Davis and Millis

(2014) explained the difference between their actual patient data and the Berthelson et al. (2013) Monte Carlo simulated data as a result of likely violation of the Monte Carlo assumption of multivariate normality. Indeed the Monte Carlo simulation employed by Berthelson et al. (2013) generated simulated data with a mean of 0 and standard deviation of 1.0, producing scores that followed the standard normal curve. PVTs, however, typically are performed at ceiling, forming skewed distributions; for example, returning to the TOMM data mentioned earlier, 21 patients with aphasia averaged 98.7% correct on Trial 2, with over half, 16, achieving perfect scores; 22 patients with severe TBI averaged 98.2% correct on Trial 2, with over half, 14, achieving perfect scores.

The Berthelson et al. (2013) failure rates are also significantly higher than those reported by Victor, Boone, Serpa, Buehler, and Ziegler (2009) for failure of ≥ 2 of 4 PVTs at a per-test false-positive rate of .15, yielding a 6% false-positive rate versus 14.6% for the Berthelson et al. Monte Carlo estimate. In Larrabee (2014) I demonstrated lower false-positive rates for the Advanced Clinical Solutions (2009) five PVTs, for ≥ 2 of 5 failures, 6%, and for my own data set, ≥ 2 of 5 failures, 5.6%, compared with the Berthelson et al. (2013) value of 11.5%. For failure of ≥ 2 of 7, the Berthelson et al. (2013) failure rate of 17.5% was larger than the 11.1% failure rate for failure of ≥ 2 of 7 variables for my combined data set (Larrabee, 2003; Larrabee, 2009), as well as larger than the false-positive rates of 5% for nonpsychotic and 7% for psychotic patients reported by Schroeder and Marshall (2011) for failure ≥ 2 of 7 PVTs. Given these data, I concluded that the practice of using failure of 2 or more PVTs/SVTs as representing probable invalid clinical presentation was supported.

In light of the preceding discussion of diagnostic statistics and determination of performance and symptom validity, the failure of multiple PVTs and SVTs indicates a high probability of invalid data. In the context of multiple PVT and SVT failures, low scores in an evaluation are more likely the result of invalid performance, particularly in the absence of indicators for severe impairment, such as a history of prolonged coma or the need for 24-hour supervised care. Correspondingly, normal range scores themselves are likely underestimates of actual level of ability.

Diagnostic Considerations

PVT and SVT failure alone do not equal malingering. Rather, it is the context in which such failure occurs, such that failure of multiple PVTs and SVTs, in the context of a substantial external incentive, with no clear evidence of major neurological, psychiatric, or developmental contributions to test performance meets general criteria for probable malingering using either the MND criteria of Slick et al. (1999) or the MPRD criteria of Bianchini et al. (2005). Below-chance performance on two-alternative forced-choice testing, in the context of external incentive, meets the criteria for definite malingering (definite MND or MPRD), a pattern that Pankratz and Erickson (1990, p. 385) has characterized as the "smoking gun of intent."

My colleagues and I (Larrabee, Greiffenstein, Greve, & Bianchini, 2007) have described several common features of the Slick et al. (1999) MND criteria and the Bianchini et al. (2005) MPRD criteria. These include (1) the requirement of a substantial external incentive, (2) the requirement of multiple indicators of performance invalidity and/or symptom exaggeration, and (3) test performance and symptom report patterns that are atypical in nature and degree for bona fide neurological, psychiatric, or developmental disorders. As we concluded, it is the combined improbability of findings in the context of external incentive with the lack of any viable alternative neurological, psychiatric, or developmental explanation that establishes the intent of the examinee to malinger (Larrabee et al., 2007). Additionally, we have summarized various data from the adult literature to show the empirical support for the Slick et al. (1999) MND criteria and the Bianchini et al. (2005) MPRD criteria. In our review, we showed how the performance of litigants performing below chance was similar to that of noninjured simulators, supporting the idea that below-chance performance was intentional; in other words, the simulators were intentionally underperforming because they had been instructed to do so. Thus the similarity of litigating subjects performing below chance to the performance of the simulators supports that the performance of the below-chance litigants was also intentionally poor. We next demonstrated the similarity in performance between below-chance malingerers and criterion group probable malingerers (failing ≥ 2 PVTs), establishing the validity of the probable malingering criteria as supporting the presence of intentionally poor performance. Last, we discussed the reinforcing role of external incentives, demonstrated by the investigation of Bianchini, Curtis, and Greve (2006) showing how the frequency of malingering increased as a linear function of the amount of external incentive.

CONCLUSION

In summary, the adult literature on PVT and SVT development, the application of diagnostic statistics to detection of invalid test performance and symptom report, the use of these data to establish empirically based diagnostic criteria for malingering, and the demonstration of the power of aggregating multiple indicators to improve diagnostic accuracy by reducing false-negative and false-positive rates is substantial over the past 25 years. Adult clinicians now have an extensive armamentarium of freestanding and embedded/derived PVTs, as well as empirically derived measures of assessing accuracy of symptom report.

ACKNOWLEDGMENT

This chapter is based in part on a lecture presented to the Institute of Medicine Committee on Psychological Testing, Including Symptom Validity Testing for Social Security Disability Determination, June 25, 2014, Washington, DC. I receive royalties from Psychological Assessment Resources, Inc., for sales of the Continuous Visual Memory Test (1988).

REFERENCES

Advanced Clinical Solutions for the WAIS-IV and WMS-IV. (2009). San Antonio, TX: Pearson Education.

Arbisi, P. A., & Ben-Porath, Y. S. (1995). An MMPI-2 infrequent response scale for use with psychopathological populations: The Infrequency Psychopathology scale, *F(p)*. *Psychological Assessment*, *7*, 424–431.

Ardolf, B. R., Denney, R. L., & Houston, C. M. (2007). Base rates of negative response bias and malingered neurocognitive dysfunction among criminal defendants referred for neuropsychological evaluation. *The Clinical Neuropsychologist, 20*, 145–159.

Arnold, G., Boone, K. B., Lu, P., Dean, A., Wen, J., Nitch, S., et al. (2005). Sensitivity and specificity of Finger Tapping Test scores for the detection of suspect effort. *The Clinical Neuropsychologist, 19*, 105–120.

Ashendorf, L., Constantinou, M., & McCaffrey, R. J. (2004). The effect of depression and anxiety on the TOMM in community-dwelling older adults. *Archives of Clinical Neuropsychology, 19*, 125–130.

Baldessarini, R. J., Finklestein, S., & Arana, G. W. (1983). The predictive power of diagnostic tests and the prevalence of illness. *Archives of General Psychiatry, 40*, *569–573*.

Barrash, J., Suhr, J., & Manzel, K. (2004). Detecting poor effort and malingering with an expanded version of the Auditory Verbal Learning Test (AVLTX): Validation with clinical samples. *Journal of Clinical and Experimental Neuropsychology, 26*, 125–140.

Ben-Porath, Y. S. (2012). *Interpreting the MMPI-2-RF.* Minneapolis: University of Minnesota Press.

Ben-Porath, Y. S., & Tellegen, A. (2008). *MMPI-2-RF. Minnesota Multiphasic Personality Inventory—2 Restructured Form: Manual for administration, scoring, and interpretation*. Minneapolis: University of Minnesota Press.

Benton, A. L., & Spreen, O. (1961). Visual memory test: The simulation of mental incompetence. *Archives of General Psychiatry, 4,* 79–83.

Berthelson, L., Mulchan, S. S., Odland, A. P., Miller, L. J., & Mittenberg, W. (2013). False positive diagnosis of malingering due to the use of multiple effort tests. *Brain Injury, 27,* 909–916.

Bianchini, K. J., Curtis, K. L., & Greve, K. W. (2006). Compensation and malingering in traumatic brain injury: A dose–response relationship? *The Clinical Neuropsychologist, 20,* 831–847.

Bianchini, K. J., Greve, K. W., & Glynn, G. (2005). On the diagnosis of malingered pain-related disability: Lessons from cognitive malingering research. *Spine Journal*, *5,* 404–417.

Binder, L. M. (1990). Malingering following minor head trauma. *The Clinical Neuropsychologist, 4*, 25–36.

Boone, K. B. (2009). The need for continuous and comprehensive sampling of effort/

response bias during neuropsychological examinations. *The Clinical Neuropsychologist, 23,* 729–741.

Boone, K. B., Lu, P. H., & Wen, J. (2005). Comparisons of various RAVLT scores in the detection of noncredible memory performance. *Archives of Clinical Neuropsychology, 20*, 301–319.

Bush, S. S., Ruff, R. M., Troster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., et al. (2005). Symptom validity assessment: Practice issues and medical necessity: NAN Policy and Planning Committee. *Archives of Clinical Neuropsychology, 20,* 419–426.

Chafetz, M. (2008). Malingering on the Social Security Disability consultative exam: Predictors and base rates. *The Clinical Neuropsychologist, 22*, 529–546.

Chafetz, M., & Underhill, J. (2013). Estimated costs of malingered disability. *Archives of Clinical Neuropsychology, 28*, 633–639.

Davis, J. J., & Millis, S. R. (2014). Examination of performance validity test failure in relation to number of tests administered. *The Clinical Neuropsychologist, 28,* 199–214.

Davis, J. J., Millis, S. R., & Axelrod, B. N. (2012).Derivation of an embedded Rey Auditory Verbal Learning Test performance validity indicator. *The Clinical Neuropsychologist, 26,* 1397–1408.

Etherton, J. L., Bianchini, K. J., Ciota, M. A., & Greve, K. W. (2005). Reliable Digit Span is unaffected by laboratory-induced pain: Implications for clinical use. *Assessment, 12,* 101–106.

Etherton, J. L., Bianchini, K. J., Greve, K. W., & Ciota, M. A. (2005). Test of Memory Malingering performance is unaffected by laboratory-induced pain: Implications for clinical use. *Archives of Clinical Neuropsychology, 20*, 375–384.

Fox, D. D. (2011). Symptom validity test failure indicates invalidity of neuropsychological tests. *Clinical Neuropsychologist, 25,* 488–495.

Gervais, R. O., Ben-Porath, Y. S., Wygant, D. B., & Green, P. (2007). Development and validation of a Response Bias Scale (RBS) for the MMPI-2. *Assessment, 14,* 196–208.

Gervais, R. O., Ben-Porath, Y. S., Wygant, D. B., & Green, P. (2008). Differential sensitivity of the Response Bias Scale (RBS) and MMPI-2 validity scales to memory complaints. *The Clinical Neuropsychologist, 22,* 1061–1079.

Gervais, R. O., Russell, A. S., Green, P., Allen, L. M., Ferrari, R., & Pieschl, S. D. (2001). Effort testing in patients with fibromyalgia and disability incentives. *Journal of Rheumatology*, *28*, 1892–1899.

Goodrich-Hunsaker, N. J., & Hopkins, R. O. (2009). Word Memory Test performance in amnesic patients with hippocampal damage. *Neuropsychology, 23,* 529–534.

Green, P. (2003). *Word Memory Test*. Edmonton, Alberta, Canada: Green's Publishing.

Green, P. (2007). The pervasive influence of effort on neuropsychological tests. *Physical Medicine and Rehabilitation Clinics of North America, 18,* 43–68.

Green, P., Iverson, G. L., & Allen, L. M. (1999). Detecting malingering in head injury litigation with the Word Memory Test, *Brain Injury, 13*, 813–819.

Green, P., Rohling, M. L., Iverson, G. L., & Gervais, R. O. (2003). Relationships between olfactory discrimination and head injury severity. *Brain Injury, 17*, 479–496.

Green, P., Rohling, M. L., Lees-Haley, P. R., & Allen, L. M. (2001). Effort has a greater effect on test scores than severe brain injury in compensation claimants. *Brain Injury, 15*, 1045–1060.

Greiffenstein, M. F., & Baker, W. J. (2003). Premorbid clues? Preinjury scholastic performance and present neuropsychological functioning in late postconcussion syndrome. *The Clinical Neuropsychologist, 17*, 561–573.

Greiffenstein, M. F., Baker, W. J., & Gola, T. (1996). Motor dysfunction profiles in traumatic brain injury and post-concussion syndrome. *Journal of the International Neuropsychological Society*, *2*, 477–485.

Greve, K. W., Bianchini, K. J., Black, F. W., Heinly, M. T., Love, J. M., Swift, D. A., et al. (2006). The prevalence of cognitive malingering in persons reporting exposure to occupational and environmental substances. *NeuroToxicology, 27*, 940–950.

Grimes, D. A., & Schulz, K. F. (2005). Refining clinical diagnosis with likelihood ratios. *Lancet, 365,* 1500–1505.

Hathaway, S. R., & McKinley, J. C. (1943). *The Minnesota Multiphasic Personality Inventory*. Minneapolis: University of Minnesota Press.

Heaton, R. K., Miller, S. W., Taylor, M. J., & Grant, I. (2004). *Revised comprehensive norms for an expanded Halstead–Reitan Battery: Demographically adjusted norms for African American and Caucasian adults (HRB)*. Lutz, FL: Psychological Assessment Resources.

Heaton, R. K., Smith, H. H., Jr., Lehman, R. A., & Vogt, A. J. (1978). Prospects for faking believable deficits on neuropsychological testing. *Journal of Consulting and Clinical Psychology*, *46*, 892–900.

Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., Millis, S. R., & Conference Participants. (2009). American Academy of Clinical Neuropsychology Consensus Conference Statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist, 23,* 1093–1129.

Iverson, G. L., Le Page, J., Koehler, B. E., Shojania, K., & Badii, M. (2007). Test of Memory Malingering (TOMM) scores are not affected by chronic pain or depression in patients with fibromyalgia. *The Clinical Neuropsychologist, 21*, 532–546.

Larrabee, G. J. (2003). Detection of malingering using atypical performance patterns on standard neuropsychological tests. *The Clinical Neuropsychologist*, *17*, 410–425.

Larrabee, G. J. (2007). Introduction: Malingering, research designs and base rates. In G. J. Larrabee (Ed.), *Assessment of malingered neuropsychological deficits* (pp. 3–13). New York: Oxford University Press.

Larrabee, G. J. (2008). Aggregation across multiple indicators improves the detection of malingering: Relationship to likelihood ratios. *The Clinical Neuropsychologist, 22,* 666–679.

Larrabee, G. J. (2009). Malingering scales for the Continuous Recognition Memory Test and Continuous Visual Memory Test. *The Clinical Neuropsychologist, 23,* 167–180.

Larrabee, G. J. (2012). Performance validity and symptom validity in neuropsychological assessment. *Journal of the International Neuropsychological Society, 18*, 625–631.

Larrabee, G. J. (2014). False-positive rates associated with the use of multiple performance and symptom validity tests. *Archives of Clinical Neuropsychology, 29*, 364–373.

Larrabee, G. J., Greiffenstein, M. F., Greve, K. W., & Bianchini, K. J. (2007). Refining diagnostic criteria for malingering. In G. J. Larrabee (Ed.), *Assessment of malingered neuropsychological deficits* (pp. 334–371). New York: Oxford University Press.

Larrabee, G. J., Millis, S. R., & Meyers, J. E. (2009). 40 plus or minus 10, a new magical number: Reply to Russell. *The Clinical Neuropsychologist, 23,* 746–753.

Lees-Haley, P. R., English, L. T., & Glenn, W. J. (1991). A Fake Bad Scale for the MMPI-2 for personal injury claimants. *Psychological Reports*, *68*, 203–210.

Miller, J. B., Millis, S. R., Rapport, L. J., Bashem, J. R., Hanks, R. A., & Axelrod, B. N. (2011). Detection of insufficient effort using the Advanced Clinical Solutions for the Wechsler Memory Scale, Fourth Edition. *The Clinical Neuropsychologist, 25,* 160–172.

Mittenberg, W., Rotholc, A., Russell, E., & Heilbronner, R. (1996). Identification of malingered head injury on the Halstead–Reitan Battery. *Archives of Clinical Neuropsychology*, *11*, 271–281.

Pankratz, L., & Erickson, R. D. (1990). Two views of malingering. *The Clinical Neuropsychologist, 4,* 379–389.

Pankratz, L., Fausti, S. A., & Peed, S. A. (1975). A forced choice technique to evaluate deafness in a hysterical or malingering patient. *Journal of Consulting and Clinical Psychology, 43*, 421–422.

Reedy, S. D., Boone, K. B., Cottingham, M. E., Glaser, D. F., Lu, P.H., Victor, T. L., et al. (2013). Cross validation of the Lu and colleagues (2003) Rey–Osterrieth Complex Figure Test effort equation in a large known-group sample. *Archives of Clinical Neuropsychology, 28*, 30–37.

Rees, L. M., Tombaugh, T. N., & Boulay, L. (2001). Depression and the Test of Memory Malingering. *Archives of Clinical Neuropsychology, 16*, 501–506.

Rey, A. (1941). L'examen psychologique dans le cas d'encephalopathie traumatique [The psychological examination in the case of traumatic encephalopathy]. *Archives de Psychologie*, *28*, 286–340.

Rey, A. (1964). *L'examen clinique en psychologie* [The clinical examination in psychology]. Paris: Presses Universitaires de France.

Rienstra, A., Groot, P. F., Spaan, P. E., Majoie, C. B., Nederveen, A. J., Walstra, G. J., et al. (2013). Symptom validity testing in memory clinics: Hippocampal-memory associations and relevance for diagnosing mild cognitive impairment. *Journal of Clinical and Experimental Neuropsychology, 35*, 59–70.

Rogers, R. (1997). Researching dissimulation. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (2nd ed., pp. 398–426). New York: Guilford Press.

Rohling, M. L., Larrabee, G. J., Greiffenstein, M. F., Ben-Porath, Y. S., Lees-Haley, P. R., Green, P., et al. (2011). A misleading review of response bias: Comment on McGrath, Mitchell, Kim, & Hough, (2010). *Psychological Bulletin, 137,* 708–712.

Root, J. G., Robbins, R. N., Chang, L., & van Gorp, W. (2006). Detection of inadequate effort on the California Verbal Learning Test—Second Edition: Forced choice recognition and critical item analysis. *Journal of the International Neuropsychological Society, 12*, 688–696.

Ross, S. R., Putnam, S. H., Millis, S. R., Adams, K. M., & Krukowski, R. A. (2006). Detecting insufficient effort using the Seashore Rhythm and Speech Sounds Perception Tests in head injury. *The Clinical Neuropsychologist, 20,* 798–815.

Schroeder, R. W., & Marshall, P. S. (2011). Evaluation of the appropriateness of multiple symptom validity indices in psychotic and non-psychotic psychiatric populations. *The Clinical Neuropsychologist, 25,* 437–453.

Silverberg, N., & Barrash, J. (2005). Further validation of the expanded Auditory Verbal Learning Test for detecting poor effort and response bias: Data from temporal lobectomy candidates. *Journal of Clinical and Experimental Neuropsychology, 27*, 907–914.

Slick, D. J., Hopp, G., Strauss, E., & Spellacy, F. (1996). Victoria Symptom Validity Test: Efficiency for detecting feigned memory impairment and relationship to neuropsychological tests and MMPI-2 validity scales. *Journal of Clinical and Experimental Neuropsychology, 18,* 911–922.

Slick, D. J., Sherman, E. M. S., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist*, *13*, 545–561.

Sollman, M. J., & Berry, D. T. R. (2011). Detection of inadequate effort on neuropsychological testing: A meta-analytic update and extension. *Archives of Clinical Neuropsychology, 26,* 774–789.

Solomon, R. E., Boone, K. B., Miora, D., Skidmore, S., Cottingham, M., Victor, T.

et al. (2010). Use of the Picture Completion subtest as an embedded measure of response bias. *The Clinical Neuropsychologist, 24*, 1243–1256.

Spreen, O., & Benton, A. L. (1963). Simulation of mental deficiency on a visual memory test. *American Journal of Mental Deficiency, 67,* 909–913.

Sweet, J. J., & Guidotti Breting, L. M. (2013). Symptom validity test research: Status and clinical implications. *Journal of Experimental Psychopathology, 4*, 6–19.

Tellegen, A., & Ben-Porath, Y. S. (2008). *MMPI-2-RF (Minnesota Multiphasic Personality Inventory—2 Restructured Form) technical manual.* Minneapolis: University of Minnesota Press.

Tenhula, W. N., & Sweet, J. J. (1996). Double cross-validation of the Booklet Category Test in detecting malingered traumatic brain injury. *The Clinical Neuropsychologist, 10*, 104–116.

Tombaugh, T. N. (1996). *TOMM: Test of Memory Malingering.* New York: Multi-Health Systems.

Trahan, D. E., & Larrabee, G. J. (1988). *The Continuous Visual Memory Test*. Odessa, FL: Psychological Assessment Resources, Inc.

Vickery, C. D., Berry, D. T. R., Inman, T. H., Harris, M. J., & Orey, S. A. (2001). Detection of inadequate effort on neuropsychological testing: A meta-analytic review of selected procedures. *Archives of Clinical Neuropsychology*, *16*, 45–73.

Victor, T.L., Boone, K. B., Serpa, J. G., Buehler, J., Ziegler, E. A. (2009). Interpreting the meaning of multiple symptom validity test failure. *The Clinical Neuropsychologist, 23,* 297–313.

Wechsler, D. (2008). *WAIS-IV: Technical and interpretive manual*. San Antonio, TX: Pearson.

Wolfe, P. L., Millis, S. R., Hanks, R., Fichtenberg, N., Larrabee, G. J., & Sweet, J. J. (2010). Effort indicators within the California Verbal Learning Test—II (CVLT-II). *The Clinical Neuropsychologist, 24,* 153–168.

Part II

Detection Methods and Other Validity Test Usage Matters

5

Review of Pediatric Performance and Symptom Validity Tests

MICHAEL W. KIRKWOOD

Methodologies to identify noncredible effort and symptom feigning/exaggeration in adults are impressively well established. Hundreds of studies over the last several decades have documented the value of objective validity tests (Boone, 2007; Larrabee, 2007a; Sweet & Guidotti Breting, 2013). Multiple terms have been used to refer to these measures, including tests of effort, malingering, response bias, and symptom validity. Larrabee (2012) suggested that the term *performance validity test* (PVT) be used to refer to an objective measure evaluating validity during ability-based tests and that *symptom validity test* (SVT) be reserved for objective measures evaluating the validity of self-report data.

Larrabee's terminological distinction between PVTs and SVTs is used throughout this chapter, with PVTs being further divided into stand-alone measures and embedded indices (see Figure 5.1). Stand-alone PVTs are free-standing tests that are designed to appear to measure ability but that in actuality are simple enough that they most often measure effort instead. Embedded indicators are derived from atypical performance on tests (e.g., Digit Span) that examiners administer as part of their standard test battery. This chapter reviews the evidence for using PVTs and SVTs in child samples. The review is an update and adaptation of a previously published chapter (Kirkwood, 2012).

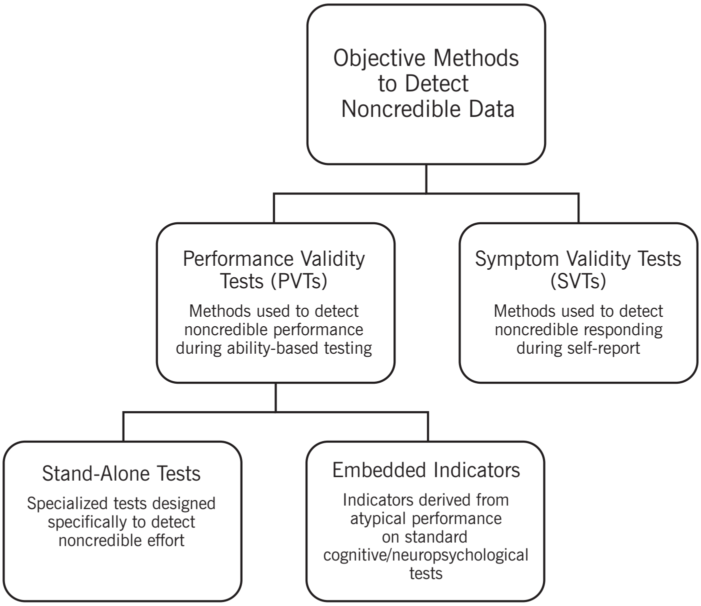

FIGURE 5.1. Terminological clarification of objective methods to detect noncredible performance and responding.

PERFORMANCE VALIDITY TESTS

Although the number of pediatric studies devoted to stand-alone PVTs pales in comparison to the number of adult-focused studies, a growing body of work has demonstrated that certain stand-alone PVTs can be used confidently in school-age children. Many fewer studies have focused on the appropriateness of embedded indices in child populations, but even in this area empirical investigation is increasing rapidly.

Stand-Alone PVTs

Table 5.1 provides a brief description of the most common stand-alone pediatric PVTs investigated in children. Table 5.2 provides estimates of the strength of empirical evidence for the tests. It should be noted that serious research into PVTs in children is only a decade old. As might be expected for a burgeoning but still initial area of investigation, many of the existing studies are methodologically imperfect (Rohling, 2004). Thus the current classification of the empirical support should be considered preliminary, weaker for most tests than the evidence supporting PVTs with adults.

TABLE 5.1. Description of Common Stand-Alone PVTs Investigated in Pediatric Populations

Test	Nature of task	Examinee response	Approximate administration time (minutes)	Description
Computerized Assessment of Response Bias (CARB; Allen, Conder, Green, & Cox, 1997)	Number recognition	Computerized	15	Examinee presented 5-digit numbers and then asked to choose number in forced-choice format
Dot Counting Test (DCT; Lezak, 1983; Rey, 1941)	Dot counting	Oral	10	Examinee counts the number of dots on six individually presented cards in both grouped and ungrouped formats
15-Item Test (FIT; Rey, 1964)	Letter, number, shape recall	Drawing	5	Examinee shown 15 items and asked to draw items from memory
Medical Symptom Validity Test (MSVT; Green, 2004)	Word recognition	Computerized	5	Examinee shown a list of semantically related words and asked to choose words in forced-choice format
Memory Validity Profile (MVP; Sherman & Brooks, in press)	Verbal and visual paradigms	Oral	10	Examinee provided verbal and visual items perceived to increase in difficulty and asked to respond in forced-choice format
Nonverbal Medical Symptom Validity Test (NV-MSVT; Green, 2008)	Object recognition	Computerized	5	Examinee shown pictures of common objects and asked to choose objects in forced-choice format
Test of Memory Malingering (TOMM; Tombaugh, 1996)	Object recognition	Oral	15	Examinee shown pictures of common objects and asked to choose objects in forced-choice format
Victoria Symptom Validity Test (VSVT; Slick et al., 1997)	Number recognition	Computerized	15	Examinee presented 5-digit numbers and then asked to choose number in forced-choice format
Word Memory Test (WMT; Green, 2003)	Word recognition	Computerized	15	Examinee shown a list of semantically related words and asked to choose words in forced-choice format

Note. Adapted from Kirkwood (2012). Copyright 2012 by Oxford University Press. Adapted by permission.

TABLE 5.2. Strength of Empirical Evidence Estimates for the Most Commonly Used Stand-Alone Performance Validity Tests in Pediatric Populations

	Strength of evidence for use in children			
	Community samples	Clinical samples	Secondary-gain samples	Simulation samples
Computerized Assessment of Response Bias (CARB)	–	+	–	–
Dot Counting Test (DCT)	–	–	–	–
15-Item Test (FIT)	+	+	–	–
Medical Symptom Validity Test (MSVT)	+	++	+	++
Memory Validity Profile (MVP)	++	+	–	+
Nonverbal Medical Symptom Validity Test (NV-MSVT)	–	+	–	–
Test of Memory Malingering (TOMM)	++	++	+	++
Victoria Symptom Validity Test (VSVT)	–	+	–	–
Word Memory Test (WMT)	+	++	–	+

Note. ++, adequate to strong evidence; +, modest evidence; –, no or conflicting evidence. Adapted from Kirkwood (2012). Copyright 2012 by Oxford University Press. Adapted by permission.

A few stand-alone PVTs have been investigated in only one or two published studies. The most promising of these are the Nonverbal Medical Symptom Validity Test (NV-MSVT; Green, 2008) and the Victoria Symptom Validity Test (VSVT; Slick, Hopp, Strauss, & Thompson, 1997), both of which have demonstrated less than 10% false-positive rates in clinically referred school-age children (Brooks, 2012; Green, Flaro, Brockhaus, & Montijo, 2012; Harrison, Flaro, & Armstrong, 2014). Further validation work with these tests is warranted to better characterize their utility across child conditions and settings.

The Computerized Assessment of Response Bias (CARB; Allen, Conder, Green, & Cox, 1997) has also been investigated in child samples, but with mixed results. Courtney, Dinkins, Allen, and Kuroski (2003) found that clinically referred children 11 years and older performed comparably to adults, although a more recent study by Harrison and colleagues (2014) found an unacceptably high false-positive rate (27%) in children with attention-deficit/hyperactivity disorder (ADHD). The Dot Counting Test (Rey, 1941) has also been investigated in children with less than favorable results. In a simulation design, Rambo, Callahan, Hogan, Hullmann, and Wrape (2015) found that it did not distinguish between children asked to try their best and those asked to intentionally perform badly, suggesting poor sensitivity.

Several additional stand-alone PVTs have received considerably more empirical attention in children, including the Test of Memory Malingering (TOMM; Tombaugh, 1996), the Word Memory Test (WMT; Green, 2003), the Medical Symptom Validity Test (MSVT; Green, 2004), and the Rey 15-Item Test (FIT; Rey, 1964). These tests are reviewed in detail next. The Memory Validity Profile (MVP; Sherman & Brooks, in press), to be published by Psychological Assessment Resources (PAR), is also described as a promising new pediatric PVT. For heuristic value, the discussion of the TOMM, WMT, MSVT, and FIT is organized by the type of sample primarily studied: (1) community-based children, (2) clinical patients, (3) examinees seen in a secondary gain context, and (4) simulators. In actuality, many of the studies relied on mixed samples of convenience without well-specified inclusion and exclusionary criteria, so overlap, known or unknown, in the types of patients included in each group may be apparent.

Test of Memory Malingering

DESCRIPTION

The Test of Memory Malingering (TOMM) is a 50-item forced-choice visual recognition test designed by Tombaugh (1996) to detect individuals who exaggerate or fake memory impairment. The test kit can be ordered from Multi-Health Systems (800-456-3003; *www.mhs.com*). The TOMM includes two learning trials, two recognition trials, and an optional retention trial. On each learning trial, the examinee is presented 50 line drawings one at a time for 3 seconds. Examinees are then asked to choose the correct drawing from a pair consisting of the target and a foil during the two recognition trials and after a 15-minute delay during the optional retention trial. Examiner feedback regarding the correctness of the response is provided after each item.

NORMATIVE DATA

Tombaugh (1996) first investigated the TOMM in primarily adult samples of community-dwelling individuals and clinical patients. The community samples included a small group of 16- and 17-year-olds, although the exact number of patients below 18 years was not reported. These studies suggested that individuals who provided adequate effort earned perfect or near perfect scores on Trial 2, whereas those instructed to feign impairment performed below the cutoff scores.

EMPIRICAL STUDIES

The TOMM has been investigated in child samples more than any other PVT. Table 5.3 presents the mean scores, standard deviations, and percentage of children performing above the recommended cutoff score available in published studies.

Two studies have focused on community samples. The first was conducted by Constantinou and McCaffrey (2003) and included children recruited from general school populations in upstate New York and in Cyprus. All children in upstate New York passed the TOMM using the recommended adult cutoffs, and all but two children from Cyprus passed. Performance was not influenced by age, educational level, culture, or gender. Rienstra, Spaan, and Schmand (2010) also investigated TOMM performance in a community sample of Dutch children. All children passed, with every child earning a perfect score on Trial 2.

Numerous studies have also investigated TOMM performance in clinical populations. Four of these were similar in design and focused on pass–fail rates in school-age children with mixed diagnoses referred for clinical evaluations (Donders, 2005; Kirk et al., 2011; Loughan & Perna, 2014; Ploetz, Mosiewicz, Kirkwood, Sherman, & Brooks, 2014). For these studies, TOMM performance was variably related to age and cognitive abilities; however, the vast majority of children passed using conventional cutoff scores. Donders (2005) found a pass rate of 97%, with two patients who failed considered true positives for noncredible effort. Kirk et al. (2011) obtained a pass rate of 96%, with all four patients who failed considered true positives. Ploetz et al. (2014) obtained a 94% pass rate, with 3% of these considered true positives; the false-positive cases were children mostly 8 years or younger with very low functioning. Loughan and Perna (2014) obtained a higher fail rate (10%), although it was not clear how many of these were true versus false positives.

In another clinical study, MacAllister and colleagues (2009) investigated TOMM performance in school-age children with epilepsy. Fifty-four of the children (90%) passed the TOMM, with 2 of the failures thought to be true positives and 4 to be false positives. Of the 11 patients with intellectual disability (ID), 7 passed the TOMM. The 4 patients who had IQ estimates below 50 passed. In the entire sample, TOMM performance was unrelated to age, though there was a significant correlation between Trial 2 scores and IQ. Gast and Hart (2010) investigated TOMM performance in adolescent males involved in the U.S. juvenile court system. Only one youth failed. All 13 adolescents whose IQs fell in the range of ID passed. Performance on Trials 1 and 2 did not covary with age, educational level, lifetime adjudications, or IQ.

TABLE 5.3. Summary of Pediatric Studies Focused on the Test of Memory Malingering

Source	Population	*N*	Age range (years)	Mean age (*SD*)	Trial 1 mean (*SD*)	Trial 2 mean (*SD*)	% passing[a]
Constantinou & McCaffrey (2003)	Cyprus community	61	5–12	8.4 (2.1)	46.8 (3.4)	49.5 (1.7)	97%
Constantinou & McCaffrey (2003)	U.S. community	67	5–12	7.9 (2.0)	45.9 (3.7)	49.9 (0.3)	100%
Rienstra et al. (2010)	Netherlands community	48	7–12	9.9 (1.6)	—	50.0 (0.0)	100%
Schneider et al. (2014)	U.S. community	30	4–7	5.6 (0.8)	43.3 (4.2)	47.1 (4.7)	85%[b]
Donders (2005)	U.S. clinical mixed	100	6–16	11.9 (3.4)	46.5 (4.2)	49.7 (0.72)	97%
MacAllister et al. (2009)	U.S. clinical epilepsy	60	6–17	~13.0 (~3.5)	43.5 (6.6)	47.5 (4.8)	90%
Kirk et al. (2011)	U.S. clinical mixed	101	5–16	10.6 (3.2)	46.7 (3.2)	49.6 (0.9)	96%
Loughan & Perna (2014)	U.S. clinical mixed	86	6–18	11.6 (3.2)	45.3 (5.6)	48.2 (4.0)	90%
Brooks et al. (2012)	U.S. clinical mixed	53	6–19	12.4 (4.1)	44.0 (5.6)	48.4 (5.0)	94%
Ploetz et al. (2014)	U.S. clinical mixed	266	5–18	13.0 (3.7)	46.9 (4.7)	46.9 6.3	94%
Schneider et al. (2014)	U.S. clinical ADHD	36	4–7	5.5 (1.0)	41.1 (6.3)	44.4 (9.2)	85%[b]
Gast & Hart (2010)	U.S. juvenile court	107	12–17	15.4 (1.4)	46.7 (3.4)	49.7 (0.9)	99%
Chafetz et al. (2007)	U.S. Social Security Disability applicants	96	6–16	10.6 (2.7)	38.2 (5.5)	40.6 (2.4)	40%
Nagle et al. (2006)	U.S. simulation controls	17	6–12	~8.6 (~2.9)	—	49.7 (0.8)	100%
Blaskewitz et al. (2008)	Germany simulation controls	51	6–11	8.9 (1.0)	—	49.8 (0.9)	100%
Gunn et al. (2010)	Australia simulation controls	50	6–11	~8.7 (~1.8)	46.6 (3.2)	49.2 (1.3)	98%
Rambo et al. (2015)	U.S. simulation controls	17	6–12	10.1 (1.8)	45.7 (4.4)	49.8 (0.75)	100%

Note. Adapted from Kirkwood (2012). Copyright 2012 by Oxford University Press. Adapted by permission.

[a]See text for available information about the number of children in each clinical sample judged to be true positive for noncredible effort.

[b]Reflects a combined passing rate among children with and without ADHD.

A clinical study by Schneider, Kirk, and Mahone (2014) is the first to include 4-year-olds in the sample. The study examined the TOMM in children ages 4–7 years with and without ADHD. There were no differences in total scores or passing rates between children with and without ADHD. However, significant age effects were found with regard to passing rates. The TOMM was passed by 93% of children ages 5–7 years; only 67% of the 4-year-olds passed, with those with ADHD showing the least success.

Two pediatric clinical studies have investigated performance on TOMM Trial 1 alone. Brooks, Sherman, and Krol (2012) examined 53 clinically referred 6- to 19-year-olds, and Perna and Loughan (2013) examined 75 clinically referred 6- to 18-year-olds. Both studies found that 100% of children who passed Trial 1 went on to pass Trial 2, suggesting that if a child passes Trial 1 the test can be discontinued without any loss of information. Whether or not a Trial 1 cutoff score can be used to indicate a fail will require further investigation.

Chafetz, Abrahams, and Kohlmaier (2007) is the only identified TOMM study focused on a sample with a clear external incentive to perform poorly (i.e., claimants for U.S. Social Security Disability benefits). In this secondary gain context, failure rates were considerably higher than in other pediatric studies, with 28% of the sample performing below the actuarial cutoff and another 31% scoring at chance levels or below. Data are presented to suggest that most cases failing the TOMM are likely true positives for noncredible effort (Chafetz, 2008; Chafetz et al., 2007).

Several studies have also investigated TOMM performance using simulation designs in which participants are assigned to an experimental condition in which they are asked to feign impairment or to a control condition in which they are asked to try hard to do well. Nagle, Everhart, Durham, McCammon, and Walker (2006) found that their optimal effort group performed comparably to adults on Trial 2. Counter to hypotheses, all but one of the children instructed to feign also performed comparably to adults and no differently from the control group. Blaskewitz, Merten, and Kathmann (2008) also found that all German children in the control condition passed. In the experimental condition, 32% of the children were not identified by TOMM performance. A recent simulation study by Rambo and colleagues (2015) found similar results, with all children in the control condition passing and 29% of the experimental group being missed. Gunn, Batchelor, and Jones (2010) found higher sensitivity in their Australian sample. One of 50 children in the control group failed the TOMM. Thirty-eight of the 40 children in the experimental malingering group were correctly identified.

SUMMARY COMMENTS

The TOMM has a number of strengths. The test stimuli consist of simple pictures of common objects and so can be used with young and language-impaired children. The stimulus books are also small and easily transportable, allowing them to be used readily in a variety of evaluative settings. More independent empirical work has focused on the TOMM than on any other PVT. Although performance may be affected slightly by age or cognitive ability, the vast majority of children 5 years and older exerting adequate effort can pass, regardless of setting, patient population, or culture. The MacAllister et al. (2009) and Ploetz et al. (2014) studies indicate that caution is needed when interpreting failing scores for those with significant cognitive impairment, particularly those 8 and younger. Notably, many individuals with IQs in the range of ID perform above adult cutoffs, so failure in a child with ID should not be assumed automatically to reflect ability-based deficits.

Drawbacks for the test include the length of administration time, as it can take 20–30 minutes to complete with children who are slow to respond. The recent research demonstrating that the test can be confidently discontinued if the child passes Trial 1 is welcomed from a time-efficiency perspective. Though failure is likely to be highly specific in all but the lowest functioning school-age children, questions about sensitivity are raised by the Nagle et al. (2006), Blaskewitz et al. (2008), and Rambo et al. (2015) simulation studies, not unlike findings from certain studies with adults (Green, 2007).

Word Memory Test

DESCRIPTION

Green's Word Memory Test (WMT) for Windows is a forced-choice verbal memory test that was designed to evaluate both effort and memory (Green, 2003). The test is available in approximately 10 languages and can be ordered from its developer and publisher, Paul Green (866-463-6968; *www.wordmemorytest.com*). The WMT involves the computerized presentation of 20 semantically related word pairs (e.g., pig–bacon) over two trials. Examinees are then asked to choose the correct word from pairs consisting of the target and a foil, during immediate recognition (IR) and delayed recognition (DR) conditions. These scores, along with the consistency of response during IR and DR, constitute the effort indices. After each IR and DR response, examinees receive auditory and visual feedback. Examinees are then asked to recognize or recall the words during several additional subtests intended to measure verbal memory rather than effort, including a multiple-choice task, a paired-associate subtest, a delayed free-recall subtest, and another free-recall subtest after a further delay. When a patient scores below the specified effort cutoffs, a profile analysis is recommended. Such analysis is predicated on the idea that the primary effort subtests are easier than the other subtests intended to tap memory ability. Decision rules have been created to help examiners determine when a profile may suggest severe cognitive impairment, rather than noncredible effort, as described in multiple publications (e.g., Green et al., 2012; Green, Flaro, & Courtney, 2009).

NORMATIVE DATA

As part of the development of the WMT, the test was administered by Lloyd Flaro to clinical pediatric patients from Canada who had a wide variety of medical, psychiatric, and developmental diagnoses. These data have been presented in several places, including the test manual, a computer software program, and select scientific publications (Green & Flaro, 2003; Green et al., 2012; Green et al., 2009). As reported in Green et al. (2012), as of December 2008, 380 children with a grade 3 reading level or higher had been tested on the WMT. Of these patients, 10% failed using the recommended actuarial cutoffs, and 5% failed, displaying a profile consistent with what is seen in those who display suboptimal effort. Of the 38 children who were administered the WMT and had less than a third-grade reading level 26% failed.

EMPIRICAL STUDIES

Seven identified case series or experimental studies have focused on the WMT: one involving typically developing children, five involving clinical populations, and one using a simulation design. Table 5.4 presents the mean WMT effort scores, standard deviations, and percentage of children performing above the recommended cutoff scores available in these published studies. Carone (2014) also presented a case of a single child with mild ID and significant brain pathology who was able to pass the WMT.

Rienstra and colleagues (2010) investigated WMT performance in their sample of typically developing Dutch children. When compared with the performance of healthy adults, the children's WMT scores were significantly lower, although all children passed using recommended adult cutoffs.

Courtney and colleagues (2003) administered the WMT to mixed clinical pediatric patients. No relation was found between IQ and performance on the effort subtests, whereas age was positively related to performance. Children 10 years old and younger were found to perform significantly worse on the WMT than older children. Once children reached age 11, performance was consistently at adult levels. Several other clinical WMT studies have utilized data from the case series of 7- to 18-year-olds referred to Flaro's private practice (these subsets are not described in Table 5.4 given that most of the patients overlap with those in Green et al., 2012). The first

TABLE 5.4. Summary of Pediatric Studies Focused on the Word Memory Test

Source	Population	*N*	Age range (years)	Mean age (*SD*)	IR % mean (*SD*)	DR % mean (*SD*)	CNS %	% passing[a]
Rienstra et al. (2010)	Netherlands community	48	7–12	9.9 (1.6)	—	—	—	100%
Green et al. (2012)	Canada clinical mixed ≥ 3rd grade reading level	380	—	13.4 (2.7)	95.9 (5.7)	95.9 (7.0)	93.8 (7.7)	90%
Courtney et al. (2003)	U.S. clinical mixed—younger group	55	6–9	8.5 (1.2)	Average effort scores 74.2 (18.8)			—
Courtney et al. (2003)	U.S. clinical mixed—older group	56	10–17	13.4 (2.0)	Average effort scores 93.4 (10.4)			—
Larochette & Harrison (2012)	U.S. clinical learning disability	63	11–14	12.2 (0.6)	—	—	—	91%
Gunn et al. (2010)	Australia simulation controls	50	6–11	~8.7 (~1.8)	90.6 (7.6)	95.3 (6.1)	—	98%

Note. IR, Immediate Recognition; DR, Delayed Recognition; CNS, Consistency.

Adapted from Kirkwood (2012). Copyright 2012 by Oxford University Press. Adapted by permission.

[a]See text for available information about the number of children in each clinical sample judged to be true positive for noncredible effort.

of these studies was by Green and Flaro (2003), who found that 14% of 135 referred children failed the WMT, with evidence presented that at least some failed because of fluctuations in effort. Performance was found to be unrelated to IQ. In contrast to the Courtney et al. (2003) study, Green and Flaro (2003) found WMT performance to be related to reading level rather than age. Those children at or below a grade 2 reading level scored worse than the other children, but once children could read at a grade 3 level, their performances were comparable to those of adults, regardless of age. Harrison, Flaro, and Armstrong (2014) examined WMT performance in 73 children with ADHD from Flaro's case series. Nine of the children performed below the actuarial cutoff. Three of the 9 were identified by profile analysis as potential true positives for noncredible effort. Green and Flaro (2015) also examined WMT performance in children with IQs 70 or below from the Flaro series. If the reading level was third grade or higher, failure rates were found not to be elevated.

In an independent clinical study, Larochette and Harrison (2012) administered the WMT to early adolescents referred for psychoeducational assessment. Six students (9.5%) performed below the suggested actuarial cutoffs. These six students had verbal intellectual and memory abilities within the average range; however, they showed significant reading impairment, with reading below the recommended third-grade level. The authors concluded that failing performance for these students likely represented poor reading ability rather than noncredible effort. Profile analysis would have correctly identified five of these six students as having authentic impairment sufficient enough to cause failure on the test.

The Gunn et al. (2010) simulation study described previously included an oral version of the WMT administered to schoolchildren from Australia. Age, IQ, and reading ability were not related to the DR trial. Oral vocabulary and reading did account for a small, significant amount of variance on the IR trial. One of the 50 children in the control group failed the WMT. Thirty-six of the 40 children in the experimental malingering group were correctly identified, resulting in 90% sensitivity and 98% specificity for this oral version of the WMT.

SUMMARY COMMENTS

The WMT has a number of strengths, including an extensive adult literature demonstrating its sensitivity and specificity. Administration, scoring, and data storage are automated and computerized and so are particularly quick and easy. The availability of the stimuli in multiple languages is a plus. Normative data and empirical studies suggest that the vast majority of clinical patients who can read at a third-grade level should be able to pass using adult cutoffs.

Because the WMT requires reading, it can be expected to have less utility in younger or very impaired children. It also takes more time than certain other PVTs (e.g., MSVT), so it is apt to be less useful when testing time is limited. Though adult-based studies suggest that it is likely to be quite sensitive to suboptimal effort, further independent work examining sensitivity in child samples will be necessary. Additional independent work is also necessary to examine specificity in pediatric patients who have significant cognitive impairment. The proposed profile analysis available for the WMT and Green's other PVTs (i.e., the MSVT and the NV-MSVT) is unique and could be a clear added benefit in helping examiners differentiate failure resulting from true impairment from that resulting from noncredible effort; initial work with adults and the work of Harrison and colleagues with children is quite promising, but further independent study will be required to establish the classification accuracy of such analysis in pediatric samples.

Medical Symptom Validity Test

DESCRIPTION

Green's Medical Symptom Validity Test (MSVT) for Windows is a forced-choice verbal memory test that was designed to evaluate effort and memory in both adults and children (Green, 2004). The test is available in approximately 10 languages and can be ordered from its developer and publisher, Paul Green (866-463-6968; *www.wordmemorytest.com*). It is similar in design to the WMT but shorter and easier, taking only about 5 minutes to administer. It involves the computerized presentation of 10 semantically related word pairs (e.g., school–book) over two trials. Examinees are then asked to choose the correct word from pairs consisting of the target and a foil during IR and DR conditions. After each response, examinees receive auditory and visual feedback. Examinees are then asked to recall the words during paired-associate (PA) and free-recall (FR) conditions. The primary effort indices are IR, DR, and consistency of response (CNS) during these subtests. Like the WMT, when a patient scores below the specified cutoffs on the MSVT, a profile analysis is recommended, with decision rules available that were designed to help identify severe ability-based impairment rather than noncredible effort (e.g., Green et al., 2009; Green et al., 2012).

NORMATIVE DATA

As part of the development of the MSVT, the test was administered by colleagues of the publisher who work with children. These data have been presented in several places, including the test manual, a computer software program, and select scientific publications (Green et al., 2012; Green et al., 2009). The two primary datasets are from Lloyd Flaro and John Courtney. Flaro administered the MSVT to 55 healthy Canadian children without psychiatric or neurological illness ages 8–11 years (and one 7-year-old). Out of the 55 children, 53 passed the MSVT validity subtests, with no age/grade effect found. Courtney's data are from 82 healthy Brazilian children asked to do their best and 27 healthy Brazilian children asked to simulate memory impairment. In the children asked to try their best, 80 out of 82 passed the effort subtests. All 27 simulators failed the effort subtests. Flaro has also administered the MSVT to clinical pediatric patients with a wide variety of medical, psychiatric, and developmental diagnoses. As reported in Green et al. (2012), as of December 2008, 265 children with a grade 3 reading level or higher had been administered the MSVT. Of these patients, 5% failed using the recommended actuarial cutoffs, and 3% failed with a poor effort profile. Of the 46 children who were administered the MSVT and had less than a third-grade reading level, 20% failed the MSVT.

EMPIRICAL STUDIES

Eight identified case series or experimental studies have focused on the MSVT in children: six with clinical populations, one in a secondary gain context, and one using a simulation design. Table 5.5 presents the mean MSVT effort scores, standard deviations, and percentage of children performing above the recommended cutoff scores available in published studies. Carone (2014) also presented a case of a single child with mild ID who was able to pass the MSVT.

A study by Carone (2008) compared the performance of children with moderate to severe TBI or other significant neurological or developmental problems with that of adults who had sustained a mild TBI. Whereas only 5% of the children failed the MSVT, 21% of the adults did. The two children who failed were deemed to be accurately identified as noncredible responders. A study by Kirkwood and Kirk (2010) investigated MSVT performance in a clinical sample of school-age children following mild traumatic brain injury (mild TBI). Seventeen percent of the sample failed the MSVT. Only three patients who failed were considered possible false positives, with the rest true positives for noncredible effort.

Like the WMT, two other studies have utilized data from Flaro's case series (these subsets are not described in Table 5.5 given that they are mostly captured in Green et al., 2012). Harrison et al. (2014) examined the MSVT in a subset of children with ADHD. Two of the children performed below the actuarial cutoff, with one of the two identified by profile analysis as a potential true positive for noncredible effort. Green and Flaro (2015) also examined MSVT performance in the subset of children with IQs of 70 or below. Similar to the findings for the WMT, failure rates on the MSVT were not elevated, if reading level was third grade or higher.

Two other clinical studies have focused on understanding the implications of MSVT failure in children. In adults presenting for neuropsychological evaluation, large amounts of variance in neuropsychological ability-based test performance is accounted for by whether examinees exert adequate effort as measured by PVTs (Constantinou, Bauer, Ashendorf, Fisher, & McCaffrey, 2005; Green, Rohling, Lees-Haley, & Allen, 2001; Lange, Iverson, Brooks, & Rennison, 2010; Meyers, Volbrecht, Axelrod, & Reinsch-Boothby, 2011). Using data from the Children's Hospital Colorado Concussion Program, we found similar effects in 276 8- to 16-year-olds; the MSVT correlated significantly with performance on all ability-based tests and explained 38% of the total ability-based test variance, far more than any demographic or injury-related variable (Kirkwood, Yeates, Randolph, & Kirk, 2012). In a nonoverlapping group of 191 youth ages 8–17 years referred clinically in the same case series, we also demonstrated that participants failing the MSVT endorsed significantly more postconcussive

TABLE 5.5. Summary of Pediatric Studies Focused on the Medical Symptom Validity Test

Source	Population	*N*	Age range (years)	Mean age (*SD*)	IR % mean (*SD*)	DR % mean (*SD*)	CNS % mean (*SD*)	% passing[a]
Green et al. (2009)	Canada community	56	7–11	9.2 (1.7)	98.6 (3.8)	98.6 (3.0)	97.6 (5.4)	96%
Green et al. (2009)	Brazil community—young	36	6–10	8.7 (1.4)	95 (5)	99 (3)	94 (8)	98%
Green et al. (2009)	Brazil community—old	34	11–15	12.4 (1.3)	96 (4)	100 (2)	96 (4)	98%
Green et al. (2012)	Canada clinical mixed ≥ 3rd-grade reading level	265	—	13.6 (2.9)	98.8 (3.7)	98.0 (4.3)	97.3 (5.8)	95%
Carone (2008)	U.S. clinical mixed	38	—	11.8 (3.1)	98.6 (3.7)	97.6 (6.3)	96.7 (9.0)	95%
Kirkwood & Kirk (2010)	U.S. clinical mild TBI	193	8–17	14.5 (2.4)	95.5 (5.3)	93.6 (5.4)	93.9 (4.8)	83%
Chafetz et al. (2007)	U.S. Social Security Disability applicants	27	6–16	11.36 (2.6)	86.4 (8.0)	84.2 (9.9)	87.8 (9.1)	37%
Blaskewitz et al. (2008)	Germany simulation controls	51	6–11	8.9 (1.0)	98.6 (2.5)	99.6 (1.2)	98.2 (3.6)	98%

Note. IR, Immediate Recognition; DR, Delayed Recognition; CNS, Consistency.

Adapted from Kirkwood (2012). Copyright 2012 by Oxford University Press. Adapted by permission.

[a]See text for available information about the number of children in each clinical sample judged to be true positive for noncredible effort.

symptoms than those passing (Kirkwood, Peterson, Connery, Baker, & Grubenhoff, 2014).

The Chafetz et al. (2007) study that focused on individuals seeking Social Security Disability compensation is the only identified study with the MSVT that has included a sample with a clear incentive to perform poorly. The authors administered the MSVT to 27 children under a nonstandardized administration procedure (i.e., the examiner read the directions and the stimuli to the claimants as they were presented on the computer screen, instead of allowing the examinees to read on their own). Based on actuarial cutoffs, 37% of the children failed the MSVT, and another 26% scored at chance levels or below.

The Blaskewitz et al. (2008) study with German children has been the only simulation study to include the MSVT. All but one child in the control group instructed to give good effort passed the MSVT. The one child who failed was in second grade and scored right at the recommended cutoff. In the experimental suboptimal condition, 90% of the children were correctly identified by the MSVT, better than the 68% identified by the TOMM.

SUMMARY COMMENTS

One of the MSVT's clear strengths is its brief administration time, which makes it an ideal screening test for noncredible effort. Like the WMT, administration, scoring, and data storage are automated and computerized and so are particularly quick and easy. The availability of the stimuli in multiple languages is a plus as well. Normative data and a growing body of independent work suggest that the vast majority of children who can read at a third-grade level or higher can pass using adult cutoffs. The Blaskewitz et al. (2008) study also suggests that the test may be more sensitive than the TOMM in the detection of noncredible effort.

Because the MSVT requires reading, it will generally be inappropriate for children in the earliest of school years or who have lower functioning. Further independent research is necessary to examine sensitivity compared with other well-validated measures and to examine specificity in pediatric patients with significant learning, attentional, and neurological impairment. As is the case for the WMT, the proposed profile analysis could be a clear added benefit in helping examiners differentiate failure resulting from true impairment from that resulting from suboptimal effort. Initial work is promising, but further independent work will be required to establish the classification utility of such analysis in children.

Rey 15-Item Test

DESCRIPTION

The 15-Item Test (FIT) was developed originally by Rey (1964) to detect memory feigning and has since been adapted by a number of authors (e.g., Boone, Salazar, Lu, Warner-Chacon, & Razani, 2002; Lezak, 1983). The test is not copyrighted and can be created simply from available descriptions (e.g., Strauss, Sherman, & Spreen, 2006). It consists of 15 items that are arranged in 3 columns by 5 rows, which examinees are shown for 10 seconds before being asked to draw the items from memory. Because of item redundancy, the examinee needs to recall only a few ideas rather than 15 independent items.

NORMATIVE DATA

No child normative data were provided originally.

EMPIRICAL STUDIES

Three published pediatric studies have included administration of the FIT: one with typically developing children, one with a clinical population, and one using a simulation design.

In the Constantinou and McCaffrey (2003) study described previously, the authors administered the FIT along with the TOMM to community child samples from upstate New York and Cyprus. Performance on the FIT at both sites correlated significantly with age and educational level. Above about age 10 years, performance was nearly errorless, though the sample sizes in the older age range were quite small.

Using the Children's Hospital Colorado Concussion Program case series, we recently examined the FIT and the recognition trial developed by Boone et al. (2002) in 319 children, ages 8 to 17, who were referred clinically following mild TBI (Green, Kirk, Connery, Baker, & Kirkwood, 2014). Failure on the MSVT was used as the criterion for noncredible effort. The traditional cutoff score of below 9 on the recall trial yielded excellent specificity (98%) but very poor sensitivity (12%). Adding the recognition trial resulted in a combined cutoff score with reasonable sensitivity (55%) and good specificity (91%) in this relatively high-functioning population.

The FIT has also been used in one study with simulators. In the Blaskewitz et al. (2008) simulation study, the FIT was administered to 70 typically developing German children. No child in the entire control group scored below the established cutoffs for adults. In the experimental condition, only 10% of the children failed the FIT, suggesting very low sensitivity.

SUMMARY COMMENTS

The FIT has been historically one of the most frequently used PVTs by neuropsychologists, presumably because it is freely available and quick to administer. Nonetheless, in adults, it appears to be sensitive to genuine cognitive dysfunction and fairly insensitive to malingering (Strauss et al., 2006). No identified studies have used the FIT with lower functioning pediatric samples, so the test needs to be used very cautiously, if at all, in patients with significant cognitive problems. Extant pediatric studies also suggest that it needs to be used cautiously with younger children as well, as performance appears correlated significantly with ages younger than about 11 years. Although it may be passed by older, higher functioning children, the Blaskewitz et al. (2008) and Green et al. (2014) data suggest that traditional cutoff scores are likely to be quite insensitive to feigning. The Green et al. (2014) study supports adding the recognition trial developed by Boone et al. (2002), as it increased sensitivity considerably, without substantially altering specificity, at least among higher functioning 8- to 17-year-olds.

Memory Validity Profile

DESCRIPTION

The Memory Validity Profile (MVP) is a stand-alone PVT to be published by Psychological Assessment Resources that consists of both verbal and visual paradigms (Sherman & Brooks, in press). Each paradigm consists of three components that are perceived to increase progressively in difficulty. Responses are obtained through forced choice, which allows for computation of binomial probabilities and classification of performances as valid, questionable, or invalid.

NORMATIVE DATA

The MVP is normed on over 1,200 children, adolescents, and young adults between the ages of 5 and 21 years. As part of the test development process, the MVP was administered to more than 200 youth with clinical diagnoses, as well as to 45 children in a simulation study in which children were asked to either provide optimal or noncredible effort.

EMPIRICAL STUDIES

Empirical studies are not available.

SUMMARY COMMENTS

As of 2015, the MVP has not yet been independently studied. As such, more information will be needed to make definitive statements about its utility as a pediatric PVT. With that said, the development of the test is exciting because it represents the first PVT designed for youth specifically. It is also the first pediatric PVT to undergo a rigorous national norming and validation process and the only one to be conormed with a memory battery (the Child and Adolescent Memory Profile, described in the following section). Thus, although caution about the test is warranted until it can be properly reviewed and used with children in real-world settings, the test can be considered at this time to have much promise.

Embedded Performance Validity Tests

The value of embedded indices to detect noncredible performance is well established in adult populations, as they are time efficient, resistant to coaching, and allow more continual monitoring of effort than stand-alone PVTs. Less research has focused on child samples. Other than Digit Span, the indicators described next have thus far been investigated in only one pediatric study or by one investigative group, so they are probably not yet ready for widespread adoption. One exception to this could be the embedded indices in the Child and Adolescent Memory Profile (ChAMP; Sherman & Brooks, 2015), a memory battery published by Psychological Assessment Resources in 2015.

Digit Span

Digit Span performance is one of the most thoroughly investigated of all embedded indicators; it includes examination of raw scores, age-corrected scale scores (ACSS) and a variety of derived scores. The most popular of the derived scores is the Reliable Digit Span (RDS), introduced by Greiffenstein, Baker, and Gola (1994). RDS is calculated by summing the longest string of digits repeated without error over two trials under both forward and backward conditions.

Six pediatric studies have focused on Digit Span as an embedded indicator: five with clinical populations and one using a simulation design. Using data from the Children's Hospital Colorado Concussion Program, we examined the sensitivity and specificity of both RDS and ACSS in the detection of noncredible effort (Kirkwood, Hargrave, & Kirk, 2011). Fourteen percent of the sample failed both the MSVT and TOMM, which was used as the criterion for noncredible effort. An RDS ≤ 7 cutoff, often used with adults, had an unacceptably high false-positive rate (32%). However, an RDS cutoff of ≤ 5 resulted in 51% sensitivity and 91% specificity. An ACSS cutoff of ≤ 5 also resulted in sensitivity of 51%, with even better specificity at 95%. A study by Welsh, Bender, Whitman, Vasserman, and MacAllister (2012) examined RDS in 54 children with epilepsy. Only 65% of children passed using an RDS cut score of 6. RDS scores were also significantly correlated with IQ and age. Using an RDS cutoff of ≤ 4 achieved specificity of 90%, but with only a 20% sensitivity rate. A study by Loughan, Perna, and Hertza (2012) used the TOMM to split 51 dually diagnosed children into valid (86%) and invalid (14%) responders. An ACSS cutoff of ≤ 4 resulted in sensitivity of 43% with a specificity score of 91%. Perna, Loughan, Hertza, and Segraves (2014) found very similar results in 75 clinically referred children, with an ACSS cut score of ≤ 4 yielding a sensitivity of 44% and specificity of 94%. Harrison and Armstrong (2014) examined RDS and ACSS in 86 adolescents undergoing psychoeducational evaluations. The students were screened to include only credible responders. When using an ACSS cut score of 5 and Canadian Wechsler Intelligence Scale for Children (WISC) norms, 26% of the sample was identified as investing noncredible effort. Even an ACSS cut score of 4 resulted in an 11% false-positive rate when using Canadian norms. In contrast, scores for RDS showed more promise, as only 1% of the participants earned an RDS score of ≤ 6.

The only simulation work to use Digit Span was the previously discussed study by Blaskewitz et al. (2008). In the German children, 90% of the feigners were identified using adult RDS cutoffs; however, 59% of the matched controls performed below this cutoff as well, supporting the sensible idea that RDS cutoffs for adults are unlikely to be appropriate for young children. Unfortunately, the authors did not publish the classification statistics for lower RDS cutoff scores or for other Digit Span scores.

Collectively, the findings suggest that Digit Span ACSS and RDS scores may have reasonable sensitivity and specificity as embedded PVTs in older school-age children and adolescents who have relatively high functioning. The Harrison and Armstrong (2014) study indicates that RDS may also have value in adolescents with learning problems. In contrast, Digit Span scores are unlikely to have much utility as a PVT in young children or those faced with significant neurological problems owing to low sensitivity when specificity is set above 90%.

Child and Adolescent Memory Profile

In a positive development in the field of pediatric validity testing, the Child and Adolescent Memory Profile (ChAMP; Sherman & Brooks, 2015) was designed a priori to include embedded validity indicators. It is a brief memory test with two verbal and two visual subtests. All subtests have immediate and delayed trials, and the verbal subtests also include a delayed recognition trial. Each subtest contains at least one embedded validity indicator that allows ongoing monitoring of task engagement. The embedded validity indicators were derived using binomial probability on three-item forced-choice responding. Using the embedded validity indicators, performance on the subtests can be classified as valid, questionable, or invalid. The ChAMP was standardized and validated on the same children as the MVP: 1,200 youth ages 5–21 years, 200 youth with clinical diagnoses, and 45 children in a simulation design study.

Matrix Reasoning

McKinzey, Prieler, and Raven (2003) examined the value of Raven's Standard Progressive Matrices in detecting feigned impairment. They administered the test under standard instructions to 44 typically developing

Austrian children ages 7–17 years. The participants were then asked to take the test again and were instructed the second time to do as badly on the test as possible without being detected. The authors used item difficulty analyses to create a three-item scale using a floor effect strategy, with the modified formula resulting in 95% sensitivity and 95% specificity. Interestingly, two other studies have also provided indirect support to indicate that matrix reasoning tasks may be worth further investigation as embedded PVTs. In the Children's Hospital Colorado Concussion Program case series investigating the relationship between the MSVT and neuropsychological tests, Matrix Reasoning from the Wechsler Abbreviated Scale of Intelligence was significantly different between those passing and failing the MSVT and resulted in one of the largest effect sizes (Kirkwood et al., 2012). Similarly, in the Rambo et al. (2015) simulation study, performance on Matrix Reasoning was the only non-PVT that differed significantly between children asked to try their best and those asked to provide noncredible effort.

Symptom Validity Scale

Some sophisticated embedded PVT work with children comes from Chafetz and colleagues (Chafetz, 2008; Chafetz et al., 2007). In the context of conducting psychological consultative examinations with adult and child claimants for Social Security Disability benefits, Chafetz developed a rating scale that relied on data collected routinely as part of the exam. The rating scale is now referred to as the Symptom Validity Scale (SVS) for Low Functioning Individuals. The scale has been validated against the TOMM and MSVT (Chafetz, 2008; Chafetz et al., 2007), with reasonable classification statistics found for different cut scores across a variety of effort levels (e.g., below chance and below actuarial criteria for the respective PVT).

Automatized Sequences Test

The Children's Hospital Colorado Concussion Program case series was used to investigate the Automatized Sequences Test (AST) as an embedded indicator in 439 clinically referred school-age patients (Kirkwood, Connery, Kirk, & Baker, 2014). On the AST, children are asked to say the alphabet, the days of the week, and the months of the year and to count to 20 as fast as they can. Total time on the AST resulted in sensitivity of 55% and specificity of 90% when validated against multiple other PVTs.

California Verbal Learning Test—Children's Version

The Children's Hospital Colorado case series was also used to investigate the California Verbal Learning Test—Children's Version (CVLT-C) as an embedded indicator (Baker, Connery, Kirk, & Kirkwood, 2014). Although most CVLT-C variables differed between credible and noncredible responders, Recognition Discriminability was by far and away the most robust predictor of noncredible effort and again produced solid sensitivity (55%) and specificity (91%) for an embedded indicator.

CNS Vital Signs

Brooks, Sherman, and Iverson (2014) examined the embedded indicators from the CNS Vital Signs in 275 children with varying neurological conditions. When cross-validated against the TOMM and VSVT, the seven subtests of the CNS demonstrated high specificity (0.94 to 0.99) but very low sensitivity (0.04 to 0.29), suggesting that the CNS indicators would not identify the majority of children classified by the TOMM and VSVT as providing noncredible effort.

SYMPTOM VALIDITY TESTS

The validity of self-report data is measured by SVTs designed to detect “faking bad.” In adult neuropsychological populations, a number of self-report instruments have demonstrated good value in detecting symptom feigning or exaggeration. The Minnesota Multiphasic Personality Inventory–2 (MMPI-2) has garnered the most investigative attention and has impressively strong support in identifying individuals who not only exaggerate psychiatric symptoms but who also feign cognitive, health, and injury-related concerns (Larrabee, 2007b).

Numerous pediatric self-report instruments include validity scales designed to detect symptom exaggeration. Commonly used measures in pediatric neuropsychological evaluations that contain a “fake bad” scale include general personality instruments such as the MMPI-Adolescent (Infrequency scale), Personality Inventory for Youth (PIY; Dissimulation scale), Personality Assessment Inventory—Adolescent (PAI-A; Negative Impression scale), and Behavior Assessment System for Children—Second Edition (BASC-2; F Index), as well as domain- and disorder-specific scales such as the Behavior Rating Inventory of Executive Function—Self-Report (Negativity scale) and Trauma Symptom Checklist for Children (Hyperresponse scale). Each of these scales has solid normative data and at least adequate psychometric properties. However, to date, remarkably little published research has focused on the utility of the validity indices in particular.

A few studies have provided initial support for the MMPI-A in identifying feigned psychopathology (Baer, Kroll, Rinaldo, & Ballenger, 1999; Lucio, Duran, Graham, & Ben-Porath, 2002; Rogers, Hinds, & Sewell, 1996; Stein, Graham, & Williams, 1995). One study also provided support for the PIY Dissimulation Scale in identifying feigned emotional distress and psychosis (Wrobel, Lachar, Wrobel, Morgan, & Gruber, 1999), and a study found support for the PAI-A validity scales in identifying feigned ADHD in a group of adolescent university students (Rios & Morey, 2013). However, all of these studies have focused on simulators, so the value of the scales to detect symptom exaggeration in real-world child examinees remains largely unknown. Moreover, none of the studies was conducted with individuals presenting for neuropsychological or psychoeducational evaluation, so their applicability to children who may be more likely to present with exaggerated physical or cognitive complaints than psychiatric problems is uncertain.

In a recent study using the Children's Hospital Colorado Concussion Program dataset, we examined the relationship between the BASC-2 Self-Report of Personality validity scales and the MSVT in a sample of 274 children, ages 8 to 17 years, after mild TBI (Kirk, Hutaff-Lee, Connery, Baker, & Kirkwood, 2014). Only 7 of the 274 patients (2.5%) fell in the "Caution" or "Extreme Caution" range on the BASC-2 F Index. Of the 50 patients who failed the MSVT, only 3 were identified by the F Index. These data suggest that sole reliance on the BASC-2 self-report validity indices is likely to substantially underestimate the number of children who display noncredible presentations following TBI. Very little overlap in this sample was found between patients identified by the self-report validity indices and patients failing a performance-based PVT, which contrasts dramatically with findings from a number of adult validity scales (e.g., MMPI-2 Fake Bad Scale; Larrabee, 2007b).

CONCLUSION AND FUTURE DIRECTIONS

At this point, the number of PVTs that have demonstrated utility in children pales in comparison to those available to the adult practitioner. Even so, sufficient data now exist to strongly support the use of several PVTs with school-age children, including the TOMM, WMT, and MSVT. The FIT and Digit Span have also demonstrated value in higher functioning older school-age children. Although the MVP and ChAMP from Psychological Assessment Resources will need to be independently studied, they are the first PVTs designed, normed, and validated specifically with children, so they can be considered quite promising.

Of course, regardless of the specific PVTs that are being used, examiners need to remember that PVT performance depends in part on the particular demands of the task and can vary for a multitude of reasons, including true cognitive impairment and temporary fluctuations in arousal, attention, emotional state, and effort. Determining whether a child is responding more broadly in a noncredible fashion requires not only careful examination of PVT performance but also a solid understanding of the natural history of the presenting condition; scrutiny of the child's developmental, medical, educational, and environmental background; and thorough consideration of the consistency and plausibility of the behavioral, self-report, and test data.

Future work is needed to demonstrate the base rate of noncredible effort and PVT classification statistics across child samples ranging more widely in ability level, presenting condition, and evaluative context. Further work will allow cut scores to be established for specific conditions and settings, rather than relying on a single cut score for all children alike, regardless of level of functioning or varying needs for increased sensitivity or specificity. The introduction of new developmentally grounded PVTs is also in order, to establish a more complete armamentarium of child-specific measures similar to that which is available to those who examine adults. In this regard, the introduction of the MVP is a welcomed development in the field of pediatric validity testing.

Given that "effort" is a continuous variable that can fluctuate throughout an evaluation, test publishers also need to design pediatric tests that integrate embedded indices into individual tests across ability-based tasks. The development of the ChAMP is an encouraging sign in this regard. As has been done for many adult tests, pediatric investigators also need to continue to work to independently validate embedded indices from existing tests to establish classification statistics that could be used confidently in different settings and for different conditions.

Finally, as the focus of research to date has been on performance-based validity almost exclusively, a separate line of study needs to focus on validating pediatric SVTs to detect exaggerated or feigned self-report data in examinees presenting for psychoeducational or neuropsychological evaluation. Tools to evaluate the credibility of self-report data will be especially valuable when assessing patients or students with conditions in which subjective symptoms play a defining role in the diagnostic process, such as ADHD, posttraumatic stress, mild TBI, and pain conditions.

AUTHOR DISCLOSURE

I was paid as an expert consultant reviewer by Psychological Assessment Resources during the development of the Memory Validity Profile (MVP; Sherman & Brooks, in press) and the Children and Adolescent Memory Profile (ChAMP; Sherman & Brooks, 2015). I receive no financial compensation from the sales of these tests.

REFERENCES

Allen, L. M., Conder, R. L., Green, P., & Cox, D. R. (1997). *Manual for the Computerized Assessment of Response Bias*. Durham, NC: CogniSyst.

Baer, R. A., Kroll, L. S., Rinaldo, J., & Ballenger, J. (1999). Detecting and discriminating between random responding and overreporting on the MMPI-A. *Journal of Personality Assessment, 72,* 308–320.

Baker, D., Connery, A.K., Kirk, J. W., & Kirkwood, M. W. (2014). Embedded performance validity indicators within the California Verbal Learning Test, Children's Version. *The Clinical Neuropsychologist, 28*, 116–127.

Blaskewitz, N., Merten, T., & Kathmann, N. (2008). Performance of children on symptom validity tests: TOMM, MSVT, and FIT. *Archives of Clinical Neuropsychology, 23*, 379–391.

Boone, K. B. (2007). *Assessment of feigned cognitive impairment: A neuropsychological perspective*. New York: Guilford Press.

Boone, K. B., Salazar, X., Lu, P., Warner-Chacon, K., & Razani, J. (2002). The Rey 15-Item Recognition Trial: A technique to enhance sensitivity of the Rey 15-Item Memorization Test. *Journal of Clinical and Experimental Neuropsychology, 24,* 561–573.

Brooks, B. L. (2012). Victoria Symptom Validity Test performance in children and adolescents with neurological disorders. *Archives of Clinical Neuropsychology, 27,* 858–868.

Brooks, B. L., Sherman, E. M., & Iverson, G. L. (2014). Embedded validity indicators on CNS vital signs in youth with neurological diagnoses. *Archives of Clinical Neuropsychology, 29*, 422–431.

Brooks, B. L., Sherman, E. M., & Krol, A. L. (2012). Utility of TOMM Trial 1 as an indicator of effort in children and adolescents. *Archives of Clinical Neuropsychology, 27*, 23–29.

Carone, D. A. (2008). Children with moderate/severe brain damage/dysfunction outperform adults with mild-to-no brain damage on the Medical Symptom Validity Test. *Brain Injury, 22*, 960–971.

Carone, D. A. (2014). Young child with severe brain volume loss easily passes the Word Memory Test and Medical Symptom Validity Test: Implications for mild TBI. *The Clinical Neuropsychologist, 28*, 146–162.

Chafetz, M. D. (2008). Malingering on the Social Security Disability consultative exam: Predictors and base rates. *The Clinical Neuropsychologist, 22,* 529–546.

Chafetz, M. D., Abrahams, J. P., & Kohlmaier, J. (2007). Malingering on the Social Security Disability consultative exam: A new rating scale. *Archives of Clinical Neuropsychology, 22*, 1–14.

Constantinou, M., Bauer, L., Ashendorf, L., Fisher, J. M., & McCaffrey, R. J. (2005). Is poor performance on recognition memory effort measures indicative of generalized poor performance on neuropsychological tests? *Archives of Clinical Neuropsychology, 20,* 191–198.

Constantinou, M., & McCaffrey, R. J. (2003). Using the TOMM for evaluating children's effort to perform optimally on neuropsychological measures. *Child Neuropsychology, 9*, 81–90.

Courtney, J. C., Dinkins, J. P., Allen, L. M., III, & Kuroski, K. (2003). Age-related effects in children taking the Computerized Assessment of Response Bias and Word Memory Test. *Child Neuropsychology, 9*, 109–116.

Donders, J. (2005). Performance on the Test of Memory Malingering in a mixed pediatric sample. *Child Neuropsychology, 11*, 221–227.

Gast, J., & Hart, K. J. (2010). The performance of juvenile offenders on the Test of Memory Malingering. *Journal of Forensic Psychology Practice, 10,* 53–68.

Green, C. M., Kirk, J. W., Connery, A. K., Baker, D. A., & Kirkwood, M. W. (2014). The use of the Rey 15-Item Test and recognition trial to evaluate noncredible effort after pediatric mild TBI. *Journal of Clinical and Experimental Neuropsychology, 36*, 261–267.

Green, P. (2003). *Manual for the Word Memory Test*. Edmonton, Alberta, Canada: Green's Publishing.

Green, P. (2004). *Manual for the Medical Symptom Validity Test*. Edmonton, Alberta, Canada: Green's Publishing.

Green, P. (2007). Spoiled for choice: Making comparisons between forced-choice effort tests. In K. B. Boone (Ed.), *Assessment of feigned cognitive impairment* (pp. 50–77). New York: Guilford Press.

Green, P. (2008). *Manual for the Nonverbal Medical Symptom Validity Test*. Edmonton, Alberta, Canada: Green's Publishing.

Green, P., & Flaro, L. (2003). Word Memory Test performance in children. *Child Neuropsychology, 9*, 189–207.

Green, P., & Flaro, L. (2015). Results from three performance validity tests in children with intellectual disability. *Applied Neuropsychology*. Available at *www.tandfonline.com/doi/abs/10.1080/23279095.2014.925903#preview*.

Green, P., Flaro, L., Brockhaus, R., & Montijo, J. (2012). Performance on the WMT, MSVT, & NV-MSVT in children with developmental disabilities and in adults with mild traumatic brain injury. In C. R. Reynolds & A. Horton (Eds.), *Detection of malingering during head injury litigation* (2nd ed., pp. 201–219). New York: Plenum Press.

Green, P., Flaro, L., & Courtney, J. (2009). Examining false positives on the Word Memory Test in adults with mild traumatic brain injury. *Brain Injury, 23,* 741–750.

Green, P., Rohling, M., Lees-Haley, P., & Allen, L. M. (2001). Effort has a greater effect on test scores than severe brain injury in compensation claimants. *Brain Injury, 15,* 1045–1060.

Greiffenstein, M. F., Baker, W. J., & Gola, T. (1994). Validation of malingered amnesia measures with a large clinical sample. *Psychological Assessment, 6,* 218–224.

Gunn, D., Batchelor, J., & Jones, M. (2010). Detection of simulated memory impairment in 6- to 11-year-old children. *Child Neuropsychology, 16,* 105–118.

Harrison, A. G., & Armstrong, I. (2014). WISC-IV unusual digit span performance in a sample of adolescents with learning disabilities. *Applied Neuropsychology: Child, 3*, 152–160.

Harrison, A. G., Flaro, L., & Armstrong, I. (2014). Rates of effort test failure in children with ADHD: An exploratory study. *Applied Neuropsychology: Child*. Available at *www.tandfonline.com/doi/abs/10.1080/21622965.2013.850581#.VOT9wWA5Ccw*.

Kirk, J. W., Harris, B., Hutaff-Lee, C. F., Koelemay, S. W., Dinkins, J. P., & Kirkwood, M. W. (2011). Performance on the Test of Memory Malingering (TOMM) among a large clinic- referred pediatric sample. *Child Neuropsychology, 17,* 242–254.

Kirk, J. W., Hutaff-Lee, C., Connery, A., Baker, D. A., & Kirkwood, M. W. (2014). The relationship between the self-report BASC-2 validity indicators and performance validity test failure after pediatric mild traumatic brain injury. *Assessment, 21,* 562–569.

Kirkwood, M. W. (2012). Overview of tests and techniques to detect negative response bias in children. In E. M. S. Sherman & B. L. Brooks (Eds.), *Pediatric forensic neuropsychology* (pp. 136–161). New York: Oxford University Press.

Kirkwood, M. W., Connery, A. K., Kirk, J. W., & Baker, D. A. (2014). Detecting performance invalidity in children: Not quite as easy as A, B, C, 1, 2, 3 but automatized sequences appears promising. *Child Neuropsychology, 20*, 245–252.

Kirkwood, M. W., Hargrave, D., & Kirk, J. W. (2011). The value of the WISC-IV Digit Span subtest in detecting noncredible performance during pediatric neuropsychological exam. *Archives of Clinical Neuropsychology, 26*, 377–384.

Kirkwood, M. W., & Kirk, J. W. (2010). The base rate of suboptimal effort in a pediatric mild TBI sample: Performance on the Medical Symptom Validity Test. *The Clinical Neuropsychologist, 24,* 860–872.

Kirkwood, M. W., Peterson, R. L., Connery, A. K., Baker, D. A., & Grubenhoff, J. A.

(2014). Postconcussive symptom exaggeration after pediatric mild traumatic brain injury. *Pediatrics, 133*, 643–650.

Kirkwood, M. W., Yeates, K. O., Randolph, C., & Kirk, J. W. (2012). The implications of symptom validity test failure for ability-based test performance in a pediatric sample. *Psychological Assessment, 24*, 36–45.

Lange, R. T., Iverson, G. L., Brooks, B. L., & Rennison, L. A. (2010). Influence of poor effort on self-reported symptoms and neurocognitive test performance following mild traumatic brain injury. *Journal of Clinical and Experimental Neuropsychology, 32*, 961–972.

Larochette, A. C., & Harrison, A. G. (2012). Word Memory Test performance in Canadian adolescents with learning disabilities: A preliminary study. *Applied Neuropsychology: Child, 1*, 38–47.

Larrabee, G. J. (2007a). *Assessment of malingered neuropsychological deficits*. New York: Oxford University Press.

Larrabee, G. J. (2007b). Evaluation of exaggerated health and injury symptomatology. In G. J. Larrabee (Ed.), *Assessment of malingered neuropsychological deficits* (pp. 264–286). New York: Oxford University Press.

Larrabee, G. J. (2012). Performance validity and symptom validity in neuropsychological assessment. *Journal of the International Neuropsychological Society, 18,* 625–631.

Lezak, M. D. (1983). *Neuropsychological assessment*. New York: Oxford University Press.

Loughan, A. R., & Perna, R. (2014). Performance and specificity rates in the Test of Memory Malingering: An investigation into pediatric clinical populations. *Applied Neuropsychology: Child, 3*, 26–30.

Loughan, A. R., Perna, R., & Hertza, J. (2012). The value of the Wechsler Intelligence Scale for Children—Fourth Edition digit span as an embedded measure of effort: An investigation into children with dual diagnoses. *Archives of Clinical Neuropsychology, 27*, 716–724.

Lucio, E., Duran, C., Graham, J. R., & Ben-Porath, Y. S. (2002). Identifying faking bad on the Minnesota Multiphasic Personality Inventory—Adolescent with Mexican adolescents. *Assessment, 9,* 62–69.

MacAllister, W. S., Nakhutina, L., Bender, H. A., Karantzoulis, S., & Carlson, C. (2009). Assessing effort during neuropsychological evaluation with the TOMM in children and adolescents with epilepsy. *Child Neuropsychology, 15*, 521–531.

McKinzey, R. K., Prieler, J., & Raven, J. (2003). Detection of children's malingering on Raven's Standard Progressive Matrices. *British Journal of Clinical Psychology, 42,* 95–99.

Meyers, J. E., Volbrecht, M., Axelrod, B. N., & Reinsch-Boothby, L. (2011). Embedded symptom validity tests and overall neuropsychological test performance. *Archives of Clinical Neuropsychology, 26*, 8–15.

Nagle, A. M., Everhart, D. E., Durham, T. W., McCammon, S. L., & Walker, M. (2006). Deception strategies in children: Examination of forced choice recognition and verbal learning and memory techniques. *Archives of Clinical Neuropsychology, 21,* 777–785.

Perna, R., Loughan, A. R., Hertza, J., & Segraves, K. (2014). The value of embedded measures in detecting suboptimal effort in children: An investigation into the WISC-IV Digit Span and CMS Verbal Memory subtests. *Applied Neuropsychology: Child, 3*, 45–51.

Perna, R. B., & Loughan, A. R. (2013). Children and the Test of Memory Malingering: Is one trial enough? *Child Neuropsychology, 19*, 438–447.

Ploetz, D. M., Mosiewicz, A. M., Kirkwood, M. W., Sherman, E. M. S., & Brooks, B. L. (2014). Performance on the Test of Memory Malingering in children with neurological conditions. *Child Neuropsychology*. Available at *www.tandfonline.com/doi/abs/10.1080/09297049.2014.986446#preview.*

Rambo, P. L., Callahan, J. L., Hogan, L. R., Hullmann, S., & Wrape, E. (2015). Effort testing in children: Can cognitive and symptom validity measures differentiate malingered performances? *Applied Neuropsychology: Child, 4,* 1–8.

Rey, A. (1941). L'examen psychologique dans les cas d'encéphalopathie traumatique [The psychological examination in cases of traumatic encepholopathy]. *Archives de Psychologie, 28*, 215–285.

Rey, A. (1964). *L'examen clinique en psychologie* [The clinical examination in psychology]. Paris: Presses Universitaires de France.

Rienstra, A., Spaan, P. E., & Schmand, B. (2010). Validation of symptom validity tests using a "child-model" of adult cognitive impairments. *Archives of Clinical Neuropsychology, 25*, 371–382.

Rios, J., & Morey, L. C. (2013). Detecting feigned ADHD in later adolescence: An examination of three PAI-A negative distortion indicators. *Journal of Personality Assessment, 95*, 594–599.

Rogers, R., Hinds, J. D., & Sewell, K. W. (1996). Feigning psychopathology among adolescent offenders: Validation of the SIRS, MMPI-A, and SIMS. *Journal of Personality Assessment, 67,* 244–257.

Rohling, M. L. (2004). Who do they think they're kidding: Review of the use of symptom validity tests with children. *Division of Clinical Neuropsychology Newsletter, 22,* 1, 21–26.

Schneider, H. E., Kirk, J. W., & Mahone, E. M. (2014). Utility of the Test of Memory Malingering (TOMM) in children ages 4–7 years with and without ADHD. *The Clinical Neuropsychologist, 28*, 1133–1145.

Sherman, E. M. S., & Brooks, B. L. (in press). *Memory Validity Profile.* Lutz, FL: Psychological Assessment Resources.

Sherman, E. M. S., & Brooks, B. L. (2015). *Children and Adolescent Memory Profile.* Lutz, FL: Psychological Assessment Resources.

Slick, D. J., Hopp, G., Strauss, E., & Thompson, G. B. (1997). *Victoria Symptom Validity Test.* Odessa, FL: Psychological Assessment Resources.

Stein, L. A. R., Graham, J. R., & Williams, C. L. (1995). Detecting fake-bad MMPI-A profiles. *Journal of Personality Assessment, 65,* 415–427.

Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary* (3rd ed.). New York: Oxford University Press.

Sweet, J. J., & Guidotti Breting, L. M. (2013). Symptom validity test research: Status and clinical implication. *Journal of Experimental Psychopathology, 4,* 6–19.

Tombaugh, T. N. (1996). *Test of Memory Malingering (TOMM).* North Tonawanda, NY: Multi-Health Systems.

Welsh, A. J., Bender, H. A., Whitman, L. A., Vasserman, M., & MacAllister, W. S. (2012). Clinical utility of reliable digit span in assessing effort in children and adolescents with epilepsy. *Archives of Clinical Neuropsychology, 27*, 735–741.

Wrobel, T. A., Lachar, D., Wrobel, N. H., Morgan, S. T., & Gruber, C. P. (1999). Performance of the Personality Inventory for Youth validity scales. *Assessment, 6,* 367–380.

6

Clinical Strategies to Assess the Credibility of Presentations in Children

DOMINIC A. CARONE

Objective methods to assess the validity of a patient's reported symptoms and performance on neuropsychological testing are well established in the scientific literature and recommended for both adults and children (Bush et al., 2005; Heilbronner, et al., 2009). These measures are commonly referred to as *effort tests, symptom validity tests*, or *performance validity tests* (PVTs). The term *PVT* is generally used throughout this chapter. The use of PVTs has been noted to be more widespread with adults compared with children (Constantinou & McCaffrey, 2003; Donders, 2005; Greve et al., 2007; Kirk et al., 2011; McCann, 1998).

Although there has been a marked increase in the use of PVTs with children in recent years (Blaskewitz, Merten, & Kathmann, 2008; Carone, 2008; Chafetz & Prentkowski, 2011; Kirk et al., 2011; Kirkwood, Connery, Kirk, & Baker, 2013; Kirkwood & Kirk, 2010; Perna & Loughan, 2013), research in this area remains in its relative infancy. Because there are fewer available PVTs for children and because of factors that can reduce or eliminate their applicability to certain children (e.g., very young age, low reading level), clinicians sometimes need to rely on other methods to assess the credibility of a child's clinical presentation. Whereas other chapters discuss the use of PVTs in detail, the focus of this chapter is to aid the clinician in utilizing information obtained from behavioral observations, clinical interviews, comprehensive records review, and evidence-based clinical reasoning to evaluate the credibility of a child's presentation. Unidentified case examples are utilized to illustrate various points.

Of note, clinicians will generally be on safer ground in drawing conclusions about the reliability and validity of a patient's presentation when objective data support the examiner's judgment. Relying on subjective impressions alone is fraught with potential limitations and inaccuracies (Guilmette, 2013). However, differences exist between relying exclusively on clinical instinct and relying systematically on extra-test data to form conclusions about the credibility of a patient's presentation. As is shown in this chapter, the use of extra-test data can be quantified at times and can also be combined with objective data from PVTs to *supplement* the examiner's conclusions. Such an approach is consistent with formal guidelines of malingering assessment (Slick & Sherman, 2013; Slick, Sherman, & Iverson, 1999). Specifically, those guidelines note that other evidence beyond PVT results can be used to make a determination of malingering (and by extension, conclusions about noncredible effort and exaggeration). This evidence includes:

- Evidence of a substantial external incentive(s) toward negative response bias.
- Discrepancy between test data and observed behavior, reliable collateral reports, and/or documented background history.
- Discrepancy between test data and known patterns of brain functioning.
- Discrepancy between self-reported history and documented history, known patterns of brain functioning, behavioral observations, and/or reliable collateral reports.

The Slick criteria require neuropsychologists to utilize and integrate multiple sets of data to determine whether they cohere together in a credible way. This process has also been referred to as *coherence analysis*, and it can be remembered by the mnemonic of "the seven C's" (Stewart-Patterson, 2010). That is, the examiner should evaluate for: (1) Continuity (clinical progression in the expected manner given what is scientifically known about the patient's condition), (2) Consistency of presentation over time, (3) Congruence (whether various aspects of the patient's clinical presentation are compatible), (4) Compliance (with assessments and treatment), (5) Causality (is the alleged/known condition the cause of the patient's presentation or is an alternative/complementary cause indicated?), (6) Comorbidity (are comorbidities present that can explain the patient's presentation?), and (7) Cultural factors that may affect the patient's presentation. Although there is some overlap to the seven C's, they are worth remembering when evaluating children, as well as adults, and can be utilized in conjunction with the methods described in this chapter when assessing the credibility of a child's clinical presentation.

UNDERSTANDING THE CONTEXT

When exploring reasons for noncredible clinical presentations in children, examiners need to keep the context of the evaluation in mind. To begin with, many children do not fully understand why they were referred for a neuropsychological evaluation when they arrive to the office. This is more common as the age of the child decreases. Typically, parents may have told the child that he or she is going to see a doctor who will do some tests and ask some questions that are designed to figure out how to help him or her do better in school. Most pediatric neuropsychology cases include an academic component, because there are often complaints of the child experiencing difficulties in school or questions about educational programming or school attendance. Children being seen for evaluation not uncommonly say they "hate school" and their schoolwork because of their difficulties. In addition, many of these children prefer brief tutoring to an entire day of school because of less structure, easier access to foods and leisure activities (e.g., video games), avoidance of social conflict with peers, and not needing to awaken early in the morning. Some children, such as those with conduct disorder, have a history of conflict with authority figures (e.g., parents and teachers) that can transfer to the examiner and evaluation setting (Chafetz & Prentkowski, 2011). Some children will arrive to the office in an upset and/or oppositional mood because they have associated what they have heard about the evaluation (e.g., that there is "testing") with a place (school) and activities (taking tests, schoolwork) that they strongly dislike.

In addition to understanding the aforementioned contextual variables, clinicians should recognize that some children may be aware of a compensation-seeking model within the family (e.g., multiple family members receiving disability). This situation can influence behavior (e.g., attention seeking) and lead to noncredible behavior during a neuropsychological evaluation, particularly if the reason for the evaluation is to determine eligibility for Social Security Disability insurance benefits, if there is ongoing or possible future personal injury litigation, and/or if the parents coached the child to perform poorly (Chafetz, Abrahams, & Kohlmaier, 2007; Chafetz & Prentkowski, 2011; Lu & Boone, 2002). Evaluators also need to be very careful not to solely rely on parental statements about premorbid functioning because such statements may be unintentionally or intentionally inaccurate (Lu & Boone, 2002). Last, the evaluator should be aware of psychosocial factors (e.g., parental divorce, abuse history, frequent moving, bullying in school) and medical comorbidities that may be associated with emotional pathology, behavioral disturbances, and poor compliance with academics, authority figures, and cognitive assessments. With the context of the evaluation and the child understood, the examiner can make better use of the observational techniques and clinical strategies discussed herein.

BEHAVIORAL OBSERVATIONS DURING INTERVIEW AND TESTING

Observations of the child's emotional state and behaviors can provide an important initial clue regarding possible upcoming problems with test compliance. These observations can take place in the initial meeting with the child, during the clinical interview, and during testing. Common examples are discussed in more detail in this section, as are some more general behavioral signs of poor engagement with testing that can be observed by examiners.

Signs of Passive Negativity/Aggressiveness

Signs of passive negativity/aggressive behavior are typically manifested by poor eye contact, failure to verbally acknowledge the examiner upon meeting, sitting with arms crossed, responding with a weak handshake, eye rolling, sarcastic-looking smirks, repeated and/or loud sighing, poor social reciprocity, and quick or dismissive answers during interview. Before interpreting such signs as indicators of passive negativity/aggressiveness, the examiner must take care to ensure that such factors are not primarily caused by depression, shyness (particularly in very young children), or a neurodevelopmental disorder (e.g., autism spectrum disorder). Signs of family tension during the meeting with the child are important to note. These may include an argument in the waiting area or on the way to the appointment (which may be disclosed during interview), the patient and family members not sitting close together in the waiting area or office, minimal or hostile interactions between the patient and family, and tense facial expressions.

A good example of a passive–aggressive presentation is that of an adolescent female who sat in the waiting room with her arms crossed, not speaking with her parents. She and her parents appeared tense and quiet. In the initial meeting with the examiner, the adolescent minimally responded and reciprocated with a sarcastic-looking smirk and nonchalant handshake. When the interview began, she was asked what she remembered about the etiology of a concussion that allegedly caused her persisting symptoms. She dismissively responded, "I don't know, the first I am hearing about it is from you." However, medical records showed that she was well aware of the injury cause. When confronted about her response, she stood up and walked out of the office. The behavior was very consistent with a long history of aggression noted in the medical records. She returned about 30 minutes later, after one parent gave chase to her, sat in the chair with her arms folded, had an angry-looking facial expression, and refused to respond verbally. The examiner decided that the patient was not ready for a neuropsychological evaluation at the time because she clearly was not in a mind-set to provide credible and cooperative responses. The parents were provided information for mental health intervention because she had yet to receive such care.

Signs of Active Negativity/Aggressiveness

Active negativity and aggressiveness often present in the form of negative comments about the examination (e.g., “This is stupid,” “I hate having to be here,” “Why do I have to do these dumb tests?”). One child was noted to begin crying at the beginning of testing because having to come to the examination reminded her of the car accident that caused a moderate traumatic brain injury and resulted in the death of her mother. The child repeatedly exclaimed “I don’t want to be here!” while crying, even when the father tried to calm her down. A decision was made to terminate the evaluation for the time being so as not to further traumatize the child. The child would likely have not provided valid responses while experiencing such emotional turmoil. In rare cases, children may make negative comments towards the examiner (e.g., “I don’t like you,” “You are getting on my nerves”) or cross personal space boundaries and become confrontational, which should also obviously alert the examiner to the potential for invalid responding.

Repeated Signs of a Desire to Leave

A child’s *repeated* showing of signs of a desire to leave the examination is an indicator of possible compliance problems with the examination. Such children may also be negative and aggressive when expressing a desire to leave, but this is not always the case. For example, some may repeatedly whine and make statements such as “I don’t want to be here”, “I want to go home”, “Are we almost done?” or “How many more tests do we have to do?”. When children learn that there is still much left to do, they may demonstrate further whining, sighing, and sometimes crying.

One of the most common physical behaviors in children demonstrating a desire to leave the evaluation is frequent clock checking. Some children may get out of their seats and try to open the door to leave, whereas others may be more passive in their attempts to stall the evaluation by requesting frequent breaks or trips to the restroom. An example is the case of a 7-year-old child with cerebral palsy who was repeatedly opening the office door and saying that he wanted to leave and that he could not stand being bored. He was asked to wait outside on a couch observable from the examination room, and later he slipped a note under the door stating, “I’m running away now. Goodbye forever.” This child later went on to provide random responding and fail a freestanding PVT (the Medical Symptom Validity Test); he is one of the pediatric PVT failure cases described by Carone (2008).

Signs of Separation Anxiety

Signs of separation anxiety are typically obvious when the child leaves the presence of the adult who brought him or her to the evaluation. This anxiety is frequently characterized by crying, looking out the window or door to find the adult in the waiting area, expressing a desire to be reunited with the adult, or leaving the room to be reunited. When separation anxiety is apparent before testing begins, it can be addressed by discussing the process openly with the child, requesting that the parent give the child an incentive to separate amicably, and reassuring the child that the parent will be nearby. Separation anxiety is also one of the rare instances in which a third-party observer (i.e., parent) can be permitted into the evaluation setting to increase compliance with testing (Axelrod et al., 2000). In some instances, children may not express their separation anxieties for an hour or more and then finally erupt in a crying episode. In such situations, the examiner must take into consideration that test results could have been negatively affected during that time period due to the child's pent-up emotional distress.

General Signs of Poor Engagement with Testing

Behavioral observations can provide strong direct evidence of compliance problems with testing. In some cases, this may not be purposeful and can reflect a poor night's sleep; for instance, closing one's eyes repeatedly, putting one's head down on the table, and/or falling asleep. In other instances, further inquiry may reveal that sleep was normal the night before and that the sleepy behaviors likely reflect a motivational problem when engaging in activities that are not viewed as pleasurable. Other behavioral signs provide clearer evidence of poor engagement with testing, such as not complying with test instructions. An example of this behavior would be a child not looking at test materials during part or all of a memory stimuli exposure task. To use this as evidence of poor engagement with testing, the examiner must be sure that this problem is not due to an underlying neuropsychological deficit.

Another example of poor engagement with testing can be the occurrence of frequent and quick "I don't know" answers. This behavior can occur during memory tests, particularly word list recall tasks. In some cases of poor engagement with testing, after the initial grouping of words is said and perhaps a few additional ones after that, a quick "I don't know" reply follows without any noticeable attempt to think of additional words. Sometimes, the quick "I don't know" reply may occur as soon as the child is asked to recall any words. Quick "I don't know" answers can also occur on verbal fluency tasks well before the time limit has expired. They are also common on tests that require additional querying, such as on the Similarities and Vocabulary subtests of the Wechsler intelligence scales.

Instances of quick "I don't know" answers can provide the examiner with evidence of effort problems if the examiner is able to successfully prompt the child for more information without coaching. For example, on a list-learning task, a child may quickly say 3 of the words from a 16-word list that was read and then say, "I don't know" or "That's it." If the examiner senses that the child provided this response too quickly, it is acceptable to say something such as "Well, how about you just think for a little more and see if any more words pop into your head. If not, we'll move on." In this way, the examiner does not provide any cues as to what the words are but merely provides extra time when sensing the child is rushing on an untimed test. Many times, this technique reveals that the child can actually provide one to several more words. This observation can be discussed during feedback and in the report as evidence that the child sometimes gives up too easily and can do more with additional effort. Such information helps teachers and parents realize the need to sometimes encourage the child to provide more information and not automatically accept quick "I don't know" replies for an answer.

The visual-motor corollary to a quick "I don't know" response in children is typically observed on constructional tasks such as design copying. In such cases, the child responds by rapidly producing the figure in less than 5 seconds in a sloppy and haphazard manner, clearly putting minimal effort into reproducing the items correctly. This often seems to be motivated by a desire to end the examination quickly.

Rarely, children will be so disengaged with testing that they display evidence of response sets that indicate they are not paying attention to the test content. Such response sets are usually accompanied by quick responses and immature behavior (e.g., laughing while responding). Examples of such responding include alternating "yes" and "no" responses on forced-choice tests (e.g., recognition memory) or providing the same response (e.g., all "yes" or all "no" responses) throughout a test.

Invalid response sets can also occur on questionnaires. A number of questionnaires have standardized validity indicators to help identify invalid responding. On scales that do not, clinicians can look for invalid response patterns themselves, at least to some extent. For example, I have observed young children provide quick response sets on the Beck Youth Inventories (BYI; Beck, Beck, Jolly, & Steer, 2005), a self-report scale that contains subscales for self-concept, depression, anxiety, anger, and disruptive behavior. Each subscale has 20 items and requires the child to respond to sentences by saying whether they are never, sometimes, often, or always true. Given the diversity of the item content on each scale, it is highly unlikely for a child to validly endorse the same response for all 20 items. For example, a 9-year-old child once responded on the BYI self-concept scale that he *always* felt smart but also responded that he *always* felt stupid. This is an inconsistency that he could not explain when asked about it. When the responses on these scales are invalid, one must rely on behavioral observations, interview data, and records review to draw conclusions about emotional and behavioral functioning.

REVIEW OF SCHOOL RECORDS AND MEDICAL RECORDS

A review of the child's school and medical records can play a very important supplementary role in assessing the validity of a child's presentation, because this information helps put the obtained test performance, interview data, and behavioral observations in historical context.

Review of School Records

A review of report cards (especially the comments section), progress notes, individualized educational plans, and notes from teachers, school psychologists, and school therapists (e.g., speech therapists, occupational therapists, physical therapists) may contain useful information about effort in the classroom or other school settings. This information is almost always subjective, observational, and descriptive, but it is useful nonetheless. A common example would be a note from a teacher stating that "Johnny needs to put more effort into his work. He is clearly not working up to his capabilities." Other examples include observations that the child is rushing through work in order to socialize, not handing in homework, and not studying for tests. The latter may be apparent during a study hall and/or may also be supported by the parents (who may also agree that the child does not put forth adequate effort during homework).

Important information can also be documented during classroom observations about the degree to which the child's behaviors appear off-task related to peers. These off-task behaviors can be related to genuine neuropsychological problems (e.g., attention impairment, impulsivity) and/or poor effort. It may not be possible to distinguish between the causes of these off-task behaviors based on a review of school records alone, although observations that these behaviors are accompanied by an "I don't care" attitude would give more credence to a poor effort etiology. In most cases, knowledge of the clinical history, behavioral observations during the examination, cognitive test results, and PVT results can aid in making informed judgments about the likely cause(s) of observed off-task behaviors in school.

Not only can school records provide helpful qualitative information, but they can also be used to gather quantitative information about behavioral patterns and trends. In certain cases, this information can be used to evaluate the credibility of specific causal attributions (e.g., traumatic brain injury) on functioning (e.g., academic performance) and can provide useful data regarding effort. For example, a 14-year-old boy was referred for neuropsychological assessment after suffering two concussions separated by a 2-month interval. He was at the end of the seventh grade at the time of the first injury. The second injury occurred over the summer. Both traumas met liberal criteria for a concussion (Mild Traumatic Brain Injury Committee of the Head Injury Interdisciplinary Special Interest Group of the American Congress of Rehabilitation Medicine, 1993) in that he did report a brief mental status change afterward (i.e., amnesia around the time of these events). However, in neither event did his parents notice any type of altered mental status minutes after the injuries occurred; no acute medical care was pursued after either injury; and symptoms from his orthopedic injuries were what he and his parents noticed the most afterward, at which point he followed up with his medical doctors. An electroencephalogram (EEG) and two brain magnetic resonance imaging studies were negative.

When reviewing his postinjury school grades, health care providers interpreted his low grades and decline in the first marking period of the eighth grade as due to postconcussion effects (see Figure 6.1). However, examining the preinjury report cards places the postinjury records into an entirely different context. As can be seen in Figure 6.2, his grade point average (GPA) in academic classes after his injuries was actually higher than in six of the seven prior marking periods and very close to the seventh. After his second injury, his GPA for the first marking period in the eighth grade was higher than in two prior marking periods. His fourth-grade report cards (not depicted) showed that 57% of his grades for academic classes were in the C to D range. In the fifth grade, 33% of his grades were in the C to D range (not depicted). Thus, when viewed in complete context, his postinjury report card grades do not support an academic decline due to concussion. In fact, the greatest decline did not take place until he began home instruction in the eighth grade, which was initiated after his health care providers decided to pull him out of school completely. This decline is consistent with poor effort and decreased motivation in the home instruction environment, likely due to decreased structure. Cases such as this one point to the importance of considering other information besides PVT results when assessing the impact of motivational and/or other factors on academic performance and cognitive test results.

When reviewing school psychology test reports, neuropsychologists must be cautious when interpreting statements about the validity of the test scores. Such reports often contain a statement that reads somewhat like the following: "Jane demonstrated a good attention span for a child her age and appeared to put her best effort forth on all items presented to her. These test results are felt to be an accurate estimate of her current functioning." However, these observations are never (in my experience), or almost never, accompanied by the administration of PVTs, likely because effort testing has primarily been the focus of research by neuropsychologists as

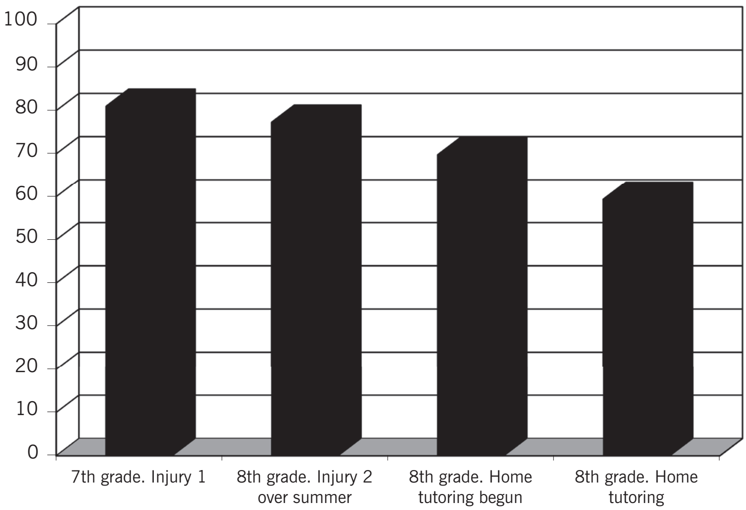

FIGURE 6.1. Report card grades from the time of the first and second head injuries/concussions.

opposed to school psychologists. Due to the omission of PVTs by school psychologists, obtained test results from such evaluations need to be interpreted cautiously when the subjective impression is that effort applied to testing was good.

A recent case highlights an example of this problem. A teenager with chronic epilepsy presented for a neuropsychological evaluation. Numerous school records documented concerns about effort in the classroom. She was tested by a school psychologist over multiple days 6 months prior to the neuropsychological evaluation. The school psychology report stated that the patient "put forth good effort and appeared to be doing her best" and that "The results reported here can be viewed as valid estimates of her current levels of functioning." The school psychology report revealed a 22-point difference between her average score on the Wechsler Intelligence Scale for Children (WISC-IV) Verbal Comprehension index (99) and her borderline score on the Perceptual Reasoning index (77); no Full-Scale IQ was calculated "due to the significant discrepancy among her index scores."

Several observations led me to have significant concerns about the validity of the WISC-IV administration in school, including the facts that (1) no objective validity tests were used, (2) an extremely low performance (scaled score of 1) on the Letter–Number Sequencing (LNS) and Symbol Search subtests was observed (which is highly atypical on the WISC-IV and yielded a Working Memory index score of 56), and (3) significant differences were apparent between her WISC-IV performance in school versus

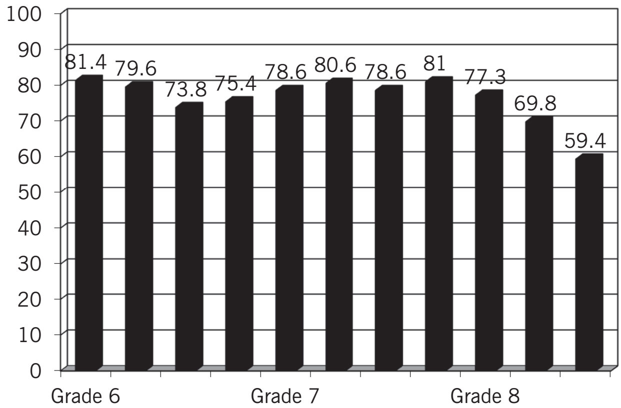

FIGURE 6.2. Report card grades pre- and postinjury.

the neuropsychological evaluation that could not be explained by practice effects. The most glaring examples of the latter included her improving from a scaled score of 1 on the LNS to a scaled score of 9 on the neuropsychological evaluation and from a scaled score of 1 to a scaled score of 8 on Symbol Search. She was administered the Wechsler Abbreviated Scale of Intelligence-II (WASI-II) during the neuropsychological evaluation, and her Full-Scale IQ was average (104), with no significant difference between her Verbal Comprehension index score (105) and her Perceptual Reasoning index score (101), very much unlike the 22-point difference previously noted on the WISC-IV.

Importantly, all of these improvements occurred in the context of her seizures becoming more intractable. Thus there was no plausible neurological explanation for the improvement. Overall, it appears that something happened during the course of the IQ testing in school that led to invalid performance and a gross underestimate of her intellectual potential, despite subjective statements by the examiner at that time to the contrary. During the neuropsychological evaluation, the patient passed multiple PVTs, including a perfect performance on the MSVT and a perfect to near-perfect performance on the Word Memory Test (Green, 2003). This example highlights how integrating knowledge about formal PVTs, reevaluation results, records review, and knowledge about biologically plausible findings can add important clarification regarding a patient's actual cognitive abilities.

Although observations by school officials can provide important information regarding effort level, it is also important to put this information into clinical context before using it as supplemental evidence of performance invalidity. Here is an example. A teenager survived a rare form of cancer that metastasized to the lungs when she was 6 years old. The cancer did not involve the brain. The cancer went into remission after seven cycles of chemotherapy and a stem cell transplant. About 1 year later she developed seizures, possibly as a result of high-dose experimental chemotherapy. The seizures were associated with profoundly abnormal EEG results, characterized by electrical status epilepticus of sleep. The seizures were intractable, and she was referred for a neuropsychological evaluation. At the time of the evaluation, brain surgery was being considered, and she was being treated with five different anticonvulsants. The child was overweight, drooling from medications, and visibly fatigued. She was anxious, depressed, tearful, and angry and irritable about her medical status.

Six months prior to the neuropsychological evaluation, the child was tested in school by a school psychologist and a speech-language pathologist. Despite borderline to extremely low working memory scores on both evaluations, word retrieval problems noted by the speech pathologist, and reports of cognitive problems from various teachers, she was denied special education services. No explicit reason for the denial was provided, but it appeared that one possible reason was the oft-mentioned observation by teachers that her work was inconsistent, that she needed to put forth more effort, and that she was not working up to her capabilities.

Although this may have been the case, it was pointed out in the neuropsychological report that (1) she passed multiple measures of test-taking effort during the examination and still demonstrated significant cognitive impairments, (2) her most recent math teacher noted that she was working to her ability yet was still having difficulty on tests, and (3) this child had been dealing with very serious medical issues throughout most of her life, which now involved the possibility of brain surgery, and she was understandably emotionally overwhelmed by this prospect (verified by interview and her responses to emotional self-report scales). This case provides an important example of how passing effort tests, comprehensive objective testing, a thorough interview, and a thorough review of the case file (including school records) can help establish the credibility of a patient's neurological presentation despite poor effort in the classroom setting.

REVIEW OF MEDICAL RECORDS

A review of medical records can also play an important role in forming conclusions about the validity of a patient's presentation. In some cases, this review may also involve a review of prior cognitive test results performed in a medical setting as opposed to a school setting. As in the school setting, the data can be graphically analyzed and evaluated for trends regarding data reliability and validity. An example follows. A 16-year-old child was referred for a neuropsychological evaluation for a concussion he had suffered in physical education class 18 months prior. In between that time, he had been administered ImPACT (Lovell, Maroon, & Collins, 1990) on three different occasions at 5-month intervals. As can be seen in Table 6.1, this youngster scored below the fifth percentile on multiple occasions in multiple domains, a performance that is grossly inconsistent with the degree of injury sustained. In fact, decades of methodologically rigorous mild TBI research, including prospective controlled and meta-analytic studies with both adults and children, generally show no significant measurable cognitive impairment on neuropsychological testing after 3 months (Babikian & Asarnow, 2009; Babikian et al., 2011; Belanger & Vanderploeg, 2005; Binder, 1997; Dikmen, Machamer, & Temkin, 2001; Dikmen, Machamer, Winn, & Temkin, 1995; Frencham, Fox, & Maybery, 2005; Ponsford et al., 2000; Rohling et al., 2011; Schretlen & Shapiro, 2003). Very low and/or inconsistent neuropsychological test scores after concussion are a red flag for invalid data and/or other factors unrelated to the injury, such as significant psychopathology, preexisting psychopathology, prior neurodevelopmental problems, or litigation stress.

As a result of these concerns, the child was asked whether or not he put forth his best effort on these tests during the interview for the neuropsychological evaluation. Initially, he stated that he tried his best and was emphatic about this point. However, he was then gently confronted about the fact that his scores did not make sense neurologically and was asked again about the degree of effort he put forth. At this point, he smiled, looked down, and acknowledged not trying on the tests because he saw them as a waste of time. Likewise, review of his physical therapy and occupational therapy notes stated that he exhibited low motivation and poor frustration tolerance when he found a task challenging. This was reportedly occurring in school as well.

Results on the Personality Assessment Inventory—Adolescent (PAI-A; Morey, 2007) for this child showed significant depression, anxiety, and somatization and no negative impression management. Thus this became

TABLE 6.1. Inconsistent and Noncredible ImPACT Test Results via File Review

ImPACT composite scores	Time 1 score	Time 1 percentile	Time 2 score	Time 2 percentile	Time 3 score	Time 3 percentile
Memory (verbal)	59	1%	74	3%	82	20%
Memory (visual)	45	1%	50	2%	50	2%
Visual motor speed	**25**	**5%**	**32.23**	**21%**	**15.25**	**1%**
Reaction time	0.80	<1%	0.70	4%	0.60	24%
Impulse control	60	—	5	—	8	—
Total symptom score	30	—	13	—	4	—

Note. Each time point was 5 months apart. Inconsistent visual motor speed scores are highlighted in **bold**.

a psychiatric case and not a neurological case by the time he was referred for a neuropsychological evaluation. The medical file review and clinical interview provided a clear picture of poor effort without the formal use of PVTs. This case also highlights the need for having an open discussion with children about their effort level and that gentle confrontation may elicit a more truthful response in some cases due to the human tendency to initially deny wrongdoing when asked.

Physical Signs of Noncredible Presentations

In working with adults, there are numerous measures and indicators of exaggeration of physical problems. These include qualitative indicators such as inconsistencies across tasks. An example here might include extremely low grip strength with the dominant hand, yet high average speeded peg placement with the same hand. This is a nonphysiological pattern because it is opposite of established patterns of motor dysfunction in patients with genuine upper motor neuron disease (Greiffenstein, Baker, & Gola, 1996). A neuromedical indicator of response bias is gait discrepancies relative to the direction of requested ambulation because gait dysfunction due to genuine hemiparesis should present similarly in all directions. This indicator is important to consider in children who claim to be unable to return to school because of chronic complaints of dizziness and imbalance after mild injuries or illnesses. One such child in late adolescence claimed to have these and many other symptoms for 2.5 years after a concussion and presented to clinic appointments with extremely low tolerance for walking even minimal distances. However, her mother, who became suspicious of these claims, brought in a picture from one of her social media accounts of her being held upside down by two adolescent males while she drank from a keg. Obviously, such behavior is inconsistent with someone who genuinely had a gait disturbance.

Another good technique to assess physical response bias is to use a distraction test, in which the examiner distracts the patient from focusing on certain symptoms and their impact on functioning during direct examination. In this way, the examiner can check for discrepancies between self-reported physical functioning and observed behavior. An example is the object drop test, in which the examiner appears to inadvertently drop an item on the floor to see if a patient who claims to be unable to bend down will bend over to pick up the dropped object. For an extensive review of this and other physical signs of response bias, see Carone (2013). To the best of my knowledge, these physical indicators of response bias have not been validated with children. Although there is no readily apparent reason as to why these indicators would not also be applicable to children, future research is needed to determine whether there are any caveats regarding their use and interpretation with younger populations.

Although neuropsychologists can use self-report scales that include validity checks of physical symptom reporting, such as the PAI-A and the Minnesota Multiphasic Personality Inventory—Adolescent (Butcher et al., 1992), use of such techniques is generally limited to the teenage years and is not the focus of this chapter. In cases in which physical exaggeration is suspected in children, neuropsychologists are generally left with utilizing medical records review, behavioral observations, and other extra-test data to form conclusions about this issue. Examples include reviewing the consistency of physical presentations over time (e.g., Romberg sign, tandem walk, finger-to-nose testing), examining whether physicians have documented nonorganic signs (e.g., sensory loss in a nondermatomal pattern, astasia–abasia), results of diagnostic studies (most commonly results from EEGs showing psychogenic nonepileptic events), and discrepancies between physical signs during a medical office visit and collateral information. As an example of the latter, a child may claim quick fatigue when asked to walk around a track as a way to avoid going to school, but the parents report that the child attended the prom and danced throughout the night without any noticeable difficulties.

In chronic concussion cases, I have observed that frequent sunglass use is often associated with physical exaggeration, because it is a way to make what is sometimes described as "an invisible injury" more visible to others. This is referred to as the "sunglass sign" and has been shown to predict nonorganic vision loss in adults and children as young as age 8 (Bengtzen, Woodward, Lynn, Newman, & Biousse, 2008). In adults, the use of props in people with mild to no verifiable neurological injuries can extend to canes, walkers, walking sticks, wheelchairs, seeing-eye dogs, and helmets (when not playing sports), although I have yet to see these props utilized by children suspected of physical exaggeration.

SUMMARY AND FUTURE DIRECTIONS

Neuropsychologists can and should utilize additional information beyond the results of formal PVTs to evaluate the credibility of children's clinical presentations. Although some of these data will be qualitative in nature (e.g., teacher comments on report cards), other information can be utilized as an additional objective data source (e.g., report card grades, prior cognitive test results). Information can also be tracked over time to determine whether observed patterns are biologically plausible given what is known about the patient's neurological condition and whether information about effort and motivation can be ascertained. This sometimes requires discussing prior data discrepancies with the patient and family and gently confronting the patient if needed and if clinically appropriate.

The main obstacle to utilizing the additional data gathering approaches discussed in this chapter is time. For example, it takes time to acquire the necessary school and medical records, review the records (which can sometimes be more numerous than in adult cases), and calculate grade point averages, as they are not always provided on the report cards. Although forensic neuropsychologists can be reimbursed for the extra time that it takes to gather and use such information, this is not always the case in clinical settings.

Future directions in this area include the need for research focused on the application of physical measures of response bias, including the development of new measures that are specific to children. Research is also needed on which positive qualitative measures of response bias and physical measures of response bias are most common in children, particularly in those with no objective biomarkers of neuropathology. Studying empirical clinical correlates (e.g., abuse history, domestic violence exposure, parental divorce, psychopathology) of these response bias indicators would also be helpful in pediatric case conceptualization and management.

REFERENCES

Axelrod, B., Barth, J., Faust, D., Fisher, J., Heilbronner, R., Larrabee, G., et al. (2000). Presence of third party observers during neuropsychological testing: Official statement of the National Academy of Neuropsychology. *Archives of Clinical Neuropsychology, 15*, 379–380.

Babikian, T., & Asarnow, R. (2009). Neurocognitive outcomes and recovery after pediatric TBI: Meta-analytic review of the literature. *Neuropsychology, 23*, 283–296.

Babikian, T., Satz, P., Zaucha, K., Light, R., Lewis, R. S., & Asarnow, R. F. (2011). The UCLA longitudinal study of neurocognitive outcomes following mild pediatric traumatic brain injury. *Journal of the International Neuropsychological Society, 17*, 886–895.

Beck, J. S., Beck, A. T., Jolly, J. B., & Steer, R. A. (2005). *Beck Youth Inventories for Children and Adolescents—Second Edition: Manual*. San Antonio, TX: Harcourt Assessment.

Belanger, H. G., & Vanderploeg, R. D. (2005). The neuropsychological impact of sports-related concussion: A meta-analysis. *Journal of the International Neuropsychological Society, 11*, 345–357.

Bengtzen, R., Woodward, M., Lynn, M. J., Newman, N. J., & Biousse, V. (2008). The "sunglasses sign" predicts nonorganic visual loss in neuro-ophthalmologic practice. *Neurology, 70*, 218–221.

Binder, L. M. (1997). A review of mild head trauma: Part II. Clinical implications. *Journal of Clinical and Experimental Neuropsychology, 19*, 432–457.

Blaskewitz, N., Merten, T., & Kathmann, N. (2008). Performance of children on symptom validity tests: TOMM, MSVT, and FIT. *Archives of Clinical Neuropsychology, 23*, 379–391.

Bush, S., Ruff, R., Troster, A., Barth, J., Koffler, S., Pliskin, N., et al. (2005). Symptom validity assessment: Practice issues and medical necessity: NAN Policy and Planning Committee. *Archives of Clinical Neuropsychology, 20*, 419–426.

Butcher, J. N., Williams, C. L., Graham, J. R., Archer, R. P., Tellegen, A., Ben-Porath,

Y. S., et al. (1992). *MMPI-A: Manual for administration, scoring, and interpretation.* Minneapolis: University of Minnesota Press.

Carone, D. A. (2008). Children with moderate/severe brain damage/dysfunction outperform adults with mild-to-no brain damage on the Medical Symptom Validity Test. *Brain Injury, 22*, 960–971.

Carone, D. A. (2013). Strategies for non-neuropsychology clinicians to detect noncredible presentations after mild traumatic brain injury. In D. A. Carone & S. B. Bush (Eds.), *Mild traumatic brain injury: Symptom validity assessment and malingering* (pp. 183–202). New York: Springer.

Chafetz, M. D., Abrahams, J. P., & Kohlmaier, J. (2007). Malingering on the Social Security Disability consultative exam: A new rating scale. *Archives of Clinical Neuropsychology, 22*, 1–14.

Chafetz, M. D., & Prentkowski, E. (2011). A case of malingering by proxy in a Social Security Disability psychological consultative examination. *Applied Neuropsychology, 18*, 143–149.

Constantinou, M., & McCaffrey, R. J. (2003). Using the TOMM for evaluating children's effort to perform optimally on neuropsychological measures. *Child Neuropsychology, 9*, 81–90.

Dikmen, S., Machamer, J., & Temkin, N. (2001). Mild head injury: Facts and artifacts. *Journal of Clinical and Experimental Neuropsychology, 23*, 729–738.

Dikmen, S., Machamer, J., Winn, R., & Temkin, N. R. (1995). Neuropsychological outcome at one-year post head injury. *Neuropsychology, 9*, 80–90.

Donders, J. (2005). Performance on the Test of Memory Malingering in a mixed pediatric sample. *Child Neuropsychology, 11*, 221–227.

Frencham, K. A., Fox, A. M., & Maybery, M. T. (2005). Neuropsychological studies of mild traumatic brain injury: A meta-analytic review of research since 1995. *Journal of Clinical and Experimental Neuropsychology, 27*, 334–351.

Green, P. (2003). *Green's Word Memory Test for Windows: User's manual*. Edmonton, Alberta, Canada: Green's Publishing.

Greiffenstein, M. F., Baker, W. J., & Gola, T. (1996). Motor dysfunction profiles in traumatic brain injury and postconcussion syndrome. *Journal of the International Neuropsychological Society, 2*, 477–485.

Greve, K. W., Bianchini, K. J., Stickle, T. R., Love, J. M., Doane, B. M., & Thompson, M. D. (2007). Effects of a community toxic release on the psychological status of children. *Child Psychiatry and Human Development, 37*, 307–323.

Guilmette, T. J. (2013). The role of clinical judgment in symptom validity assessment. In D. A. Carone & S. Bush (Eds.), *Mild traumatic brain injury: Symptom validity assessment and malingering* (pp. 31–43). New York: Springer.

Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., Millis, S. R., & Conference Participants. (2009). American Academy of Clinical Neuropsychology Consensus Conference Statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist, 23*, 1093–1129.

Kirk, J. W., Harris, B., Hutaff-Lee, C. F., Koelemay, S. W., Dinkins, J. P., & Kirkwood, M. W. (2011). Performance on the Test of Memory Malingering (TOMM) among a large clinic-referred pediatric sample. *Child Neuropsychology, 17*, 242–254.

Kirkwood, M. W., Connery, A. K., Kirk, J. W., & Baker, D. A. (2013). Detecting performance invalidity in children: Not quite as easy as A, B, C, 1, 2, 3 but automatized sequences appears promising. *Child Neuropsychology, 20*, 245–252.

Kirkwood, M. W., & Kirk, J. W. (2010). The base rate of suboptimal effort in a pediatric mild TBI sample: Performance on the Medical Symptom Validity Test. *The Clinical Neuropsychologist, 24*, 860–872.

Lovell, M., Maroon, J., & Collins, M. W. (1990). ImPACT. Retrieved April 17, 2011, from *www.impacttest.com*.

Lu, P. H., & Boone, K. B. (2002). Suspect cognitive symptoms in a 9-year-old child: Malingering by proxy? *The Clinical Neuropsychologist, 16*, 90–96.

McCann, J. T. (1998). *Malingering and deception in adolescents: Assessing credibility in criiminal and forensic settings*. Washington, DC: American Psychological Association.

Mild Traumatic Brain Injury Committee of the Head Injury Interdisciplinary Special Interest Group of the American Congress of Rehabilitation Medicine. (1993). Definition of mild traumatic brain injury. *Journal of Head Trauma Rehabilitation, 8*, 86–87.

Morey, L. C. (2007). *Personality Assessment Inventory—Adolescent: Professional manual*. Lutz, FL: Psychological Assessment Resources.

Perna, R. B., & Loughan, A. R. (2013). Children and the test of memory malingering: is one trial enough? *Child Neuropsychology, 19*, 438–447.

Ponsford, J., Willmott, C., Rothwell, A., Cameron, P., Kelly, A. M., Nelms, R., et al. (2000). Factors influencing outcome following mild traumatic brain injury in adults. *Journal of the International Neuropsychological Society, 6*, 568–579.

Rohling, M. L., Binder, L. M., Demakis, G. J., Larrabee, G. J., Ploetz, D. M., & Langhinrichsen-Rohling, J. (2011). A meta-analysis of neuropsychological outcome after mild traumatic brain injury: Re-analyses and reconsiderations of Binder et al. (1997), Frencham et al. (2005), and Pertab et al. (2009). *The Clinical Neuropsychologist*, *25*,1–16.

Schretlen, D. J., & Shapiro, A. M. (2003). A quantitative review of the effects of traumatic brain injury on cognitive functioning. *International Review of Psychiatry, 15*, 341–349.

Slick, D., & Sherman, E. M. (2013). Differential diagnosis of malingering. In D. A. Carone & S. S. Bush (Eds.), *Mild traumatic brain injury: Symptom validity assessment and malingering* (pp. 57–72). New York: Springer.

Slick, D., Sherman, E. M., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist, 13*, 545–561.

Stewart-Patterson, C. (2010). Detection of potential malingering indicators through document review. *IAIABC Journal, 47*, 19–43.

7

Motivations Behind Noncredible Presentations

Why Children Feign and How to Make This Determination

DAVID A. BAKER
MICHAEL W. KIRKWOOD

Over the last decade, an increasing amount of attention has been devoted to noncredible cognitive performance in children. Nonetheless, the potential reasons or motives driving noncredible presentations have been largely unexplored, despite their importance to accurate case conceptualization, diagnostic decision making, and clinical management. Based on the small amount of existing literature on the topic, it appears that noncredible presentations in children likely vary depending on the context in which the evaluation is being conducted (Flaro & Boone, 2009; Henry, 2005; Kirkwood, Kirk, Blaha, & Wilson, 2010; Lu & Boone, 2002; McCaffrey & Lynch, 2009; Chafetz, Abrahams, & Kohlmaier, 2007; Kirkwood & Kirk, 2010). In the medicolegal or compensation-seeking context (e.g., litigation, disability claims), the reason for the noncredible performance is most often concluded to be related to the compensation seeking. In clinical settings, the reasons that children provide invalid performance or data can be more complex. This chapter aims to assist the pediatric clinician in navigating the murky waters often confronted in the presence of noncredible presentations by exploring some common potential motivators, presenting applicable case examples, and outlining methods for discerning motivations in children and teens.

RELEVANT TERMINOLOGY

The attempt to understanding the motivations that underlie noncredible presentations usually begins with exploring what type of incentives or gains might be present (see Sherman, Chapter 2, this volume).

One of the most distinct motivators behind noncredible performance is the presence of external incentives or secondary gains, which Slick and Sherman (2012) define as any possible advantages or benefits that an individual could attain by behaving in a particular manner. These secondary gains are generally believed to drive deliberate or conscious behavior and can be delineated as material–legal or psychosocial. Material–legal gains refer to substantial tangible rewards, such as a financial settlement/compensation or disability benefit, as well as evasion of formal duties or responsibilities, such as criminal charges or institutional placement. Psychosocial secondary gains are those incentives that are less tangible and could include any interpersonal, social, or emotional benefit that could be negatively or positively reinforcing the behavior (e.g., more attention from parents or friends, avoiding aversive social situations).

In contrast to secondary gains, primary gains or internal incentives are thought to be those rewards that are less tangible and more internally driven by underlying psychological processes (see Table 7.1). Rather than being deliberately driven by conscious motives, as is often the case with secondary gains, primary gains are often believed to be driven unconsciously. These primary gains may come in the form of positively reinforcing internal benefits, such as gaining attention or affection from others, as well as negatively reinforcing gains, such as avoiding aversive stimuli or unwanted feelings (e.g., stress, anxiety). Although primary and secondary gains are often described as isolated motivators, they are not mutually exclusive and frequently co-occur (Slick & Sherman, 2012; Delis & Wetter, 2007; Boone, 2007).

POTENTIAL MOTIVATORS

Material–Legal

Performance validity tests (PVTs) and symptom validity tests (SVTs) have been propelled by the clear need for valid cognitive data within the context of civil litigation and other compensation-seeking settings. Plaintiffs or claimants have an obvious motivation to feign or exaggerate a condition in order to gain more reward (e.g., money). Other claims outside the context of civil litigation also have clear external incentives, including avoiding duties or responsibilities (e.g., criminal charges). Among various compensation-seeking adult populations, base rates of malingering have been found

TABLE 7.1. Differences between Primary and Secondary Gains Driving Noncredible Presentations

	Primary gain	Secondary gain
Incentive	Internal	External
Benefit	Immediate relief from unwanted feelings such as guilt, anxiety, tension, internal conflict, etc.	External advantages or benefits that reward or reinforce behaviors
Motivation	Unconscious processes	Conscious processes
Examples	Gaining attention due to illness or injury; alleviating pressure to excel with illness or injury; decreasing stress	Gaining monetary rewards such as disability incentives; receiving extra supports at school; avoiding certain responsibilities such as school or work

to vary depending on the evaluation context and the type of presenting problem or impairment. Mittenberg, Patton, Canyock, and Condit (2002) conducted a large national survey of adult neuropsychologists and found that base rates of malingering and symptom exaggeration varied depending on the reason for referral (29% for personal injury cases, 30% for disability cases, 19% for criminal cases, and 8% for medical cases). Among adult litigants or claimants with a suspected mild traumatic brain injury, noncredible presentations are found remarkably often, in approximately 40% of cases (Larrabee, 2003; Mittenberg et al., 2002). Individuals seeking incentives due to chronic pain or fibromyalgia also have been found to have high rates of malingering ranging from 20 to 50% (Greve, Etherton, Ord, Bianchini, & Curtis, 2009; Gervais, Green, Allen, & Iverson, 2001).

The frequency of noncredible presentations in compensation-seeking pediatric populations is much less defined. Several single-case reports have documented clearly that children and adolescents are capable of feigning cognitive impairment within the context of civil litigation. Lu and Boone (2002) present the case of a 9-year-old boy whose family was involved in litigation after he was struck by a car and sustained a moderate TBI. His performance during the independent neuropsychological evaluation was notable for failure of multiple PVTs, as well as a noncredible pattern of performance on standard neurocognitive measures. Lu and Boone (2002) determined that his parents appeared to be guiding much of his deceitful behavior, so a diagnosis of malingering by proxy was assigned. Several other case reports have highlighted children and teens presenting in a noncredible manner on testing in order to gain material benefit (e.g., financial compensation; McCaffrey & Lynch, 2009), to evade criminal responsibility (Flaro & Boone, 2009), and to secure less restrictive psychiatric placement (Flaro, Green, & Blaskewitz, 2007). In the only identified case series with children or families who had a clear external incentive to feign impairment, Chafetz (2008) found that up to 60% of a pediatric sample being evaluated for Social Security Disability benefits failed at least one PVT.

School

The potential role of school-related factors in driving noncredible presentations in pediatric patients should routinely be considered. School refusal is well covered in the pediatric psychology and educational literature and has been conceptualized as a behavior motivated potentially by several desires: to avoid school-based stimuli that provoke negative affect (e.g., anxiety, depression); to avoid uncomfortable social or evaluative situations (e.g., tests or being called on in class); to get attention from significant others (e.g., friends or parents); and/or to receive tangible reinforcers outside of school (e.g., playing video games all day; Kearney & Albano, 2004; Kearney, 2008). School refusal behavior has been found to be associated with weaker social skills and increased social isolation, high levels of family conflict and parental alcoholism, and heightened levels of comorbid psychiatric diagnoses such as depression and anxiety (McCune & Hynes, 2005). Not surprisingly, several negative outcomes have been found to be associated with school refusal, including poor academic achievement and eventual school dropout (Alexander, Entwisle, & Kabbani, 2001).

Academic struggles, adjustment difficulties, and subsequent school avoidance can be amplified at certain "stress" points such as transition from lower elementary to upper elementary, elementary to middle school, and middle school to high school (Anderson, Jacobs, Schramm, & Splittgerber, 2000). Thus, in the presence of a noncredible presentation in a pediatric patient, clinicians should consider the possibility of school refusal behavior, especially if the patient recently entered one of these academic stress points and might be struggling with adjustment to a new academic setting or the increased demands of that grade. Developmentally based conditions (e.g., attention-deficit/hyperactivity disorder [ADHD], learning disability [LD]) also add academic stress for some youth. High-achieving students might also be prone to experience heightened levels of academic stress or pressure to maintain their elevated performance. Certain patients who present in a noncredible manner may in fact be avoiding school because they are struggling to adjust to the temporary symptoms of an illness or injury and the consequent stress of missed classes and makeup work. In these cases of heightened academic stress and/or temporary illness-related conditions, informal or formal school accommodations (e.g., Section 504 Plan) can be helpful in supporting the student's transition back to school.

Case Example: FT

FT was a 14-year-old male referred for neuropsychological consultation by his neurologist due to increasing concerns with regard to headaches, forgetfulness, memory complaints, and declining academic performance since entering high school. This apparent surge in headaches and change in cognition and learning over the preceding year was concerning enough to the parents that they sought neurological consultation. Neurological examination, physical examination, blood work, and brain magnetic resonance imaging (MRI) were all unremarkable. A review of academic records revealed grade decline and standardized reading score decline beginning in sixth grade; no special education services or other formal supports were provided. FT's history was further notable for early delays in acquiring preacademic skills such as reciting the alphabet and counting and informal reading supports in second grade. Teachers had long-standing concerns about his ability to pay attention.

During the evaluation, FT failed numerous PVTs, scoring 75%, 70%, and 70% on the Medical Symptom Validity Test (MSVT; Green, 2004) Immediate Recall (IR), Delayed Recall (DR), and Consistency (CNS) scales, respectively. He also scored a raw score of 32 on Trial 1 and 33 on Trial 2 of the Test of Memory Malingering (TOMM; Tombaugh, 1996). He often stated during testing, "My memory is so bad!" Several embedded validity indicators were also suggestive of noncredible effort, including a Wechsler Intelligence Scale for Children—Fourth Edition (WISC-IV; Wechsler, 2003), Digit Span Scaled Score of 5 (Kirkwood, Hargrave, and Kirk, 2011), and a California Verbal Learning Test—Children's Version (CVLT-C; Delis, Kramer, Kaplan, & Ober, 1994) Discriminability z-score of –2.0 (Baker, Connery, Kirk & Kirkwood, 2014). Despite these results, there were some indications on testing that FT likely had at least average intellect, receiving a Wechsler Abbreviated Scale of Intelligence—Second Edition (WASI-2; Wechsler, 2011) estimated IQ of 99. Overall, based largely on the clinical history, underlying attention and reading problems were judged to be a likely contributing factor to his noncredible performance and memory and somatic complaints over the previous year. FT was thought to be trying to portray himself as more impaired than he actually was to further avoid school and/or receive more supports at school, which it appeared he probably needed. His underlying difficulties were thought to be amplified because of the stress associated with recently entering high school. Feedback to the family included reassurance that there was nothing to suggest a neurological condition. Recommendations included additional diagnostic assessment (which would need to include PVTs) to allow further exploration of FT's cognitive and academic profiles. Consistent school attendance was also strongly encouraged, as was additional support for FT at school in the form of a Section 504 Plan or an individualized education program (IEP) depending on the results of a recommended psychoeducational evaluation (which was to include objective validity testing). Another recommendation was that he participate in cognitive-behavioral therapy to improve his emotional coping skills and decrease somatic reactions.

Social

Given the pervasive social demands and influences present throughout childhood and adolescence, clinicians need to consider social factors as potential contributors to noncredible presentations. Very broadly, social factors could include issues related to psychosocial secondary gains, such as increased positive attention from others. Social stressors could also drive attempts to avoid aversive feelings or situations. A common social stressor that affects a significant number of children and teens is bullying or peer victimization (Phillips, 2007). National survey data indicate that in 2001 approximately 5.7 million American children in grades 6–10 experienced or witnessed bullying at school and that one in four students reported being bullied every month (National Institute of Child Health and Human Development, 2001).

Case Example: RD

RD was a 9-year-old male with a 3-year history of vague somatic complaints (e.g., stomachaches, headaches) with no identified etiology despite multiple medical diagnostic workups. Due to ongoing school absenteeism in the context of no known medical diagnoses, RD was referred for a neuropsychological evaluation by his primary care physician, who wanted to rule out a possible learning disorder. RD's medical and developmental histories were unremarkable. Academically, he had always done well and currently was maintaining above-average grades in the fourth grade despite missing many days of school. Socially, his parents reported that RD had a history of difficulties making and maintaining friends, which they attributed to the fact that he was the youngest and smallest in his class. He also had complained for several years about being bullied in school, but his parents assumed the school was properly handling this problem. During the evaluation, RD failed multiple PVTs, with scores of 85%, 85%, and 80% on the MSVT IR, DR, and CNS, respectively. He produced scores of 46 and 45 on Trials 1 and 2 of the TOMM but 35 on the Retention trial. He also provided a scaled score of 4 on WISC-IV Digit Span. Despite these results, RD displayed average performance in several areas, including intellect, verbal learning and memory, and reading and math skills. Additionally, the fact that school records revealed no areas of academic concern suggested that he was unlikely to have a specific learning disorder. Overall, RD's noncredible performance was judged to be motivated by his desire to avoid the social stress of school and, in particular, the potential for further bullying or peer victimization. He discussed the stressors associated with daily bullying at length during a clinical interview. Thus school personnel were encouraged to be more proactive in addressing the bullying and supporting RD socially, and RD was encouraged to work with the school psychologist. Parents were strongly encouraged to have RD attend school regularly. RD had been more socially isolated in the past few years, so a recommendation was also made that he engage in more structured social activities outside of school.

Sports

In the United States, between 40 and 50 million youth are estimated to participate in organized sports every year (National Federation of Youth Sports Associations, 2010). Athletes are often assumed to be motivated to return to their sport as quickly as possible after an injury or illness, and this is likely true for most athletes. However, in certain cases, athletes do not actually want to return to their sport. Clinically, we not infrequently see young athletes present noncredibly because they are hoping that a doctor restricts them from play rather than openly stating that they do not want to return to their sport after an injury. Motivations in these cases seem to be driven most often by fear of sustaining additional injury, not being able to play at preinjury level, or external pressure from a parent, coach, or peer to play a sport that the athlete has lost (or never had) interest in.

Case Example: FW

FW was a 15-year-old male with an unremarkable history who sustained his second concussion during a freshman football practice, with the first having occurred a year earlier. The most recent injury occurred when FW was reportedly pushed backward on a blocking drill and struck the back of his helmet on the ground. FW was then administered a computerized cognitive test and reportedly "failed it" compared with his baseline performance. He was restricted from playing football. FW was then followed for 2 weeks by his pediatrician, who noted worsening symptoms and referred FW for further consultation. An MRI was ordered by his primary care physician and was negative. Neuropsychological consultation was requested. At the time of clinical contact, FW reported multiple ongoing postconcussive symptoms. He presented to testing walking very slowly and needing parental assistance to get in and out of his chair. On testing, he failed multiple PVTs and also endorsed several noncredible symptoms (e.g., forgetting family members' names). FW's MSVT performance scores were 65%, 70%, and 60% on three primary effort indices, and his paired associates (PA) and free recall (FR) scores were both 40%. FW also provided raw scores of 32 and 35 on Trials 1 and 2 of the TOMM, respectively. He also took 14 seconds to recite the alphabet and 10 seconds to say the months of the year (Kirkwood, Connery, Kirk, & Baker, 2013). Embedded validity indicators were also notable, including a WISC-IV Digit Span scaled score of 4 and a CVLT-C Discriminability z-score of –1.5. In further discussion of these results, FW and his mother reported that there were several significant family stressors over the past year, including a recent parental separation and considerable pressure from FW's family for him to play football. FW ultimately admitted that he did not want to play football because he was worried about the risk of being seriously injured by having a third concussion. Overall, FW's noncredible performance was determined to be motivated by his desire to avoid returning to football, combined with other complicating family stressors. FW and his family were encouraged to participate in family therapy to work through many of the present issues. FW was provided education about the known risks of sport-related concussion (Kirkwood, Randolph, & Yeates, 2012; Randolph & Kirkwood, 2009). At the same time, a strong recommendation was provided to FW's parents that he be allowed to participate in activities of his choosing and be supported in his decisions.

Primary Psychological (Internalizing Disorders)

Preexisting or co-occurring psychopathology should also be routinely considered in the presence of noncredible presentations. Somatic complaints occur frequently in healthy children but become more problematic and pathological when they begin interfering with daily functioning and become a preoccupation (somatic symptom disorder). By the time a diagnosis of somatic symptom disorder (formerly somatization disorder) is made, a pattern of somatization is often evident throughout a child's life (e.g., frequent complaints of stomachaches and headaches; Gilleland, Suveg, Jacob, & Thomassin, 2009). Several risk factors predispose children to a somatization diagnosis, including female gender, exposure to trauma, lower socioeconomic status, and comorbid internalizing disorder (Kugler, Bloom, Kaercher, Truax, & Storch, 2012). Comorbid internalizing affective disorders such as anxiety and depression have been found to be associated with higher rates of somatization and worse functional disruption (Rhee, Holditch-Davis, & Miles, 2005). More specifically, internalizing disorders such as anxiety and depression also are commonly associated with several chronic health conditions, including chronic abdominal pain (Campo et al., 2004), fibromyalgia (Thieme, Turk, & Flor, 2006), chronic pain syndromes (Edwards, Augustson, & Fillingim, 2003), migraine headaches (Merikangas & Rasmussen, 2000), tension-type headaches (Janke, Holroyd, & Romanek, 2004), and lingering postconcussive symptoms (Ponsford et al., 2012). Very few studies have examined the relationship between noncredible test performance and somatization or other internalizing disorders.

However, in a recent study of ours examining a pediatric population with mild TBI, noncredible test performance was found to be associated with significantly higher rates of postconcussive symptom reporting (Kirkwood, Peterson, Connery, Baker, & Grubenhoff, 2014).

Case Example: AH

AH was a 13-year-old female with a history of long-standing anxiety who sustained a concussion while playing in a soccer game. She was seen by her pediatrician the day after her injury and was encouraged to remain out of school for at least 2 weeks, avoid any screens (e.g., computer, television), and limit socialization. AH reportedly was slow to improve over those 2 weeks at home and was reportedly "very stressed" and preoccupied about missing class and assignments. Her sleep also became problematic. When she returned to school, AH reportedly experienced an increase in her symptoms and was worse than in the days following her injury. AH reportedly had a long history of worrying and heightened anxiety, although psychological treatment had never been sought. There were select periods over the previous year (prior to her injury) in which AH would reportedly become so stressed about schoolwork that she would have "meltdowns." A family history of anxiety and depression also was ascertained. The family eventually consulted with a physiatrist, who documented a nonfocal neurological examination. Physical therapy was recommended, as was amitriptyline for headaches. A brain MRI was normal. Nonetheless, minimal improvement was seen, and AH was eventually referred for neuropsychological consultation. AH presented as highly immature and often turned to her parents for answers during information gathering. She appeared hesitant and reserved. On testing, AH produced valid scores on the MSVT of 90%, 95%, and 90% for IR, DR, and CNS, respectively. However, her performance showed questionable decline as the evaluation progressed, including a WISC-IV Digit Span Scaled Score of 1, a CVLT-C Discriminability z-score of –4.5, and TOMM raw scores of 33 and 35 for Trials 1 and 2, respectively. She also endorsed several noncredible symptoms, including seeing the world in black and white. Overall, it was concluded that AH's noncredible performance was likely driven by several factors. AH's long-standing predisposition toward anxiety, somatization, and poor stress coping likely played a prominent role in her presentation. AH's anxieties about performing poorly in school were quite likely amplified during her time away from school and her mounting missed assignments, likely leading to maladaptive coping such as somatization and avoidance. Additionally, her pediatrician's instructions to remain out of school for 2 weeks, as well as to refrain from electronics and socialization, were thought to have an iatrogenic effect and to contribute greatly to her social isolation and internalizing emotional difficulties. AH was referred for individual psychological treatment to address her anxiety and poor coping. In consultation with her physiatrist, it was agreed that AH should resume most normal activities, including conditioning for soccer, and returning to school full time with accommodations.

Primary Psychological (Conversion Disorder)

Albeit rare, conversion disorders (CD) occasionally account for noncredible presentations in children (see Sherman, Chapter 2, this volume). Conversion disorder (also referred to in DSM-5 as functional neurological symptoms disorder) is diagnosed in the presence of one or more altered motor or sensory functions (e.g., hemiparesis, visual changes) with obvious incompatibility in such symptom(s) and possible medical or neurological conditions (American Psychiatric Association, 2013). Although the presence of psychological distress is no longer required under DSM-5 diagnostic criteria, CD is often conceptualized as the manifestation of psychological distress that is "converted" into a physical symptom (Krasnik, Meaney, & Grant, 2013). In other words, underlying and complex psychological processes are believed to play a substantial role in the majority of CD cases. The incidence of pediatric CD in the United States is largely unknown, but an Australian surveillance study reported the incidence of pediatric CD to be 2.3 to 4.2 per 100,000 (Kozlowska et al., 2007). In this same Australian study, the three most common presentations of CD in children and teens were found to be weakness, ataxia, and psychogenic nonepileptic seizures (PNES). A large proportion of their population with CD (56%) also reported experiencing chronic pain.

Case Example: BR

BR was a 15-year-old female with an unremarkable medical history. Eight months prior to clinical contact, BR fell ill with mononucleosis. She remained out of school and activity for almost 3 months. Toward the end of her recovery period, BR developed abnormal symptoms, including brief convulsive episodes described by her mother as "seizures"; she also began stuttering. BR was referred to neurology by her primary care physician, and an extensive epilepsy workup revealed PNES. BR was then referred for a neuropsychological evaluation. During the clinical interview with both BR and her mother, it was learned that BR's father had been murdered 2 years previously. BR reportedly adjusted well to this tragic event and never showed signs of needing treatment. However, a previously close relationship with her mother intensified after her father's death to the point of enmeshment. At the time of neuropsychological contact, BR presented as a quiet and subdued young lady. She needed assistance from her mother to walk to the testing office. She reported several ongoing symptoms at the time of assessment, including daily "seizure" attacks, stuttering, headaches, and balance difficulties. BR also reportedly could not remember biographical information or names of familiar acquaintances. On testing, BR failed multiple PVTs, achieving scores of 65%, 60%, and 55% on MSVT IR, DR, and CNS, respectively. On the TOMM, she scored a 28 on Trial 1 and a 32 on Trial 2. She took 18 seconds to recite the alphabet. She earned a scaled score of 3 on the WISC-IV Digit Span, and on the CVLT-C Discriminability, she attained a z-score of –3.5. Her stuttering also was noted to be inconsistent throughout the evaluation. Overall, BR's noncredible performance was determined to be consistent with her conversion disorder, likely associated with the unresolved trauma surrounding her father's death. Her seizure-like episodes, memory difficulties, and gait and speech abnormalities were incompatible with any known medical condition. BR was referred for intensive outpatient therapy and weekly physical therapy to improve functioning. Considerable psychoeducation about CD was provided to the family. Unfortunately, 1 month later, BR was admitted to the hospital due to inability to move all extremities (i.e., quadriplegia), and she was eventually transferred to the inpatient psychiatric unit for more intensive treatment, highlighting the challenging nature of treating cases of true conversion disorder with well-established behavioral patterns.

Additional Factors Contributing to Noncredible Performance

The factors discussed here do not represent an exhaustive list of potential motivators. In actuality, numerous other factors can help to explain a youth's noncredible presentation, as described in the following list. It is important to note that most of the factors accounting for noncredible presentations are not mutually exclusive.

Family Stressors

Family stressors should be routinely considered as a possible contributing factor in pediatric noncredible presentations. Stressors such as marital discord, divorce, domestic violence, child abuse, parental medical or mental illness, and financial strain all have been found to contribute to chronic physical conditions and comorbid depression (Gonzalez et al., 2012). Furthermore, a child's illness or injury may serve to strengthen family cohesion or lead to increased attention for the child. In these cases, children may exaggerate or feign symptoms in order to maintain the desired family reaction. Examiners should also consider whether any family members have chronic health conditions or are seeking disability services, as this could potentially contribute to a learned pattern of behavior for the young patient and/or represent a factitious disorder by proxy (Schulte & Petermann, 2011).

Iatrogenic Factors

Iatrogenesis is preventable harm inflicted on a patient following health care examination or treatment. In some cases, certain interventions or treatments have the potential to cause patients to believe that they are more ill than they are and/or to attribute benign symptoms to an incorrect medical etiology. These perceived expectations of illness could perpetuate symptoms or reinforce symptom exaggeration. In the context of a neuropsychological examination, the belief that one is ill or impaired could contribute to an individual's believing that he or she *should* perform poorly on a neuropsychological exam. In the case of mild TBI, iatrogenic effects appear to be more frequent due to the nonspecific symptoms often associated with concussion and the potential for one to expect certain symptoms (i.e., "expectancy as etiology"; Mittenberg, DiGiulio, Perrin, & Bass, 1992). As described earlier in the case of AH, iatrogenesis can occur in the prescribed treatments for concussion (e.g., prolonged or absolute rest), especially in a child or adolescent who might be predisposed to somatization, depression, or anxiety. Additionally, patients who are told that their symptoms might be due to a specific medical diagnosis could begin to internalize such a diagnosis and behave in a manner consistent with such a diagnosis (i.e., diagnosis threat), which could affect effort and credible performance. In fact, Suhr and Gunstad (2002) examined diagnosis threat in a college-age population with mild TBI and found that the participants who were told that they were being evaluated because of their past head injuries did significantly worse on neuropsychological measures than those who were given neutral instructions with no mention of their head injuries. The diagnosis threat group additionally reported diminished effort on testing compared with the neutral group.

"Cry for Help"

When children are distressed, they may be apprehensive or unsure how to ask for help and can do so passively at times. Additionally, a child might be experiencing a certain degree of symptoms but worry that his or her symptoms might be dismissed or minimized. Thus one possible explanation for a noncredible performance could be that the youth is seeking more validation, assistance, and/or support by exaggerating an ailment or impairment. For example, consider the aforementioned case of FT, in which lifelong learning challenges went relatively undetected until, as difficulties surfaced at a time of increased demands and stress (e.g., transition to higher grades and/or an acquired ailment), more challenges with coping emerged. Thus it might not be uncommon that school accommodations and supports are needed to address long-standing weakness but the child or adolescent is unsure how to effectively ask for such assistance; thus he or she may do so in a maladaptive manner (by exaggerating symptoms or presenting in a noncredible manner).

Noncompliant Behavior

Youth may also present noncredibly as a manifestation of oppositionality or noncompliance (see Sherman, Chapter 2, this volume). In fact, Donders (2005) examined the performance of 100 general pediatric patients on the TOMM and found that two children with a history of noncompliant behavior displayed noncredible performances. Carone (2008) also found that one of the children in his case series who failed the MSVT was thought to be driven by noncompliance.

DISCERNING THE MOTIVES

Uncovering and understanding a child's motives for presenting in a noncredible manner is not an easy task. Practitioners are often faced with several viable possibilities and frequently experience initial uncertainty as to why a patient is presenting noncredibly. Despite the value of PVTs and SVTs in determining whether the data are valid or not, these tests are not designed to provide insight into the motives underlying the invalid data. Moreover, the literature provides minimal guidance on how to best discern motives behind noncredible presentations in clinical cases. Drawing on our own experiences, we have found several core clinical practices to be effective in identifying the motives behind children's noncredible presentations.

Establish Rapport

Establishing good rapport with the patient and family is a critical initial step in any clinical encounter but is highlighted in this context due to the delicate navigation often required in exploring, understanding, and then eventually explaining noncredible performance to the patient and family. Rapport is important because the exploration of a noncredible presentation is often something unexpected by the patient and family and frequently entails examination of personal and sensitive data. Establishing positive rapport can potentially alleviate a child's anxieties or apprehensions and lead to more complete reporting (Siegman & Reynolds, 1984). Furthermore, by establishing a positive rapport with the pediatric patient, it is possible that more detailed and/or accurate reporting can be elicited to further clarify motives (Roberts, Lamb, & Sternberg, 2004). At the risk of stating the obvious, rapport building can be enhanced by maintaining a warm but nonjudgmental and unbiased stance that allows clinicians to validate the presenting concerns. Patients and caregivers are more apt to disclose sensitive inormation (often critical to understanding the underlying motives) if they believe that the clinician is invested in their well-being and trustworthy from the outset of the evaluation. In the aforementioned case of FW, positive rapport was established at the onset of the evaluation by gaining an understanding of FW's personal interests and hobbies. Enthusiasm and interest was conveyed nonverbally and verbally with regard to his extracurricular interests (which had nothing to do with football). Without this positive rapport, it is possible that the underlying motives (i.e., avoiding football) could have gone undetected.

Take a Thorough History

Taking a thorough history from the patient and caregiver(s) is often a crucial step in illuminating the possible reasons behind a noncredible presentation. The history should include examination of behavioral patterns throughout the child's life, including somatic, anxious, and depressive tendencies, as well as examples of maladaptive coping in the face of current or previous stressors (see Table 7.2). Exploring how the child has dealt with adversity, pain, or illness in the past often can elucidate long-standing personality characteristics that could help to explain noncredible performance. Keeping this sort of questioning open-ended initially can be useful in allowing for spontaneous elaboration (e.g., "Have there been any significant stressful events over the past few years?"), although specific probing may be necessary as well (e.g., "Have there been any problems in the home recently?").

Conduct a Thorough Clinical Interview

Clinical interviews with the patient can also provide considerable insight into underlying motivations. Making the clinical interview conversational, open-ended, and unstructured can be helpful in eliciting more genuine and accurate responses (Turner, Hersen, & Heiser, 2003). Similar to the history taking, asking specific questions about recent or past stressors from the patient's perspective can provide data that allow clinicians to understand motivations. Exploring patients' likes and dislikes at school, their overall enjoyment of school, and any notable academic problems they might have is important. Furthermore, it is often useful to spend time discussing and exploring the child or teen's social life, including any problems with bullying or teasing. It is sometimes helpful to make a child or teen more at ease and feel less ashamed by saying something to the effect of "A lot of

TABLE 7.2. Example Interview Queries That May Assist in Understanding the Motives in Cases of Noncredible Presentations

Secondary gains/external incentives

- "Have you consulted or do you have plans to consult with an attorney about your child's injury?"
- "Is anyone in the family, including your child, receiving or seeking disability benefits?"
- "Has your child ever been involved with the police or legal/court system for any reason?"

Primary gains/internal incentives

- "Have there been any large family stressors over the last several years that have possibly affected your child?"
- "Have there been any significant social or school stressors over the past few years?"
- "Is there a history of your child having excessive reactions to pain or illness?"
- "Describe how your child has dealt with stress or adversity in the past."
- "Is there a history of chronic pain, including stomachaches or headaches that are not medically explained?"
- "How motivated is your child to succeed in school?"
- "Have you ever had difficulties getting your child to school in the past?"
- "In general, would you describe your child as perfectionistic?"

kids your age complain of *X* [e.g., being bullied at school]; do you have a problem with that?"

Use Objective Rating Forms

In addition to a thorough clinical interview, standardized objective questionnaires can often add important clinical information regarding emotional, social, and behavioral functioning in order to further discern potential motives driving a noncredible presentation. We frequently ask patients, caregivers, and occasionally teachers to complete a Behavior Assessment System for Children—Second Edition (BASC-2; Reynolds & Kamphaus, 2004) to assess current and/or premorbid level of functioning. Standardized questionnaires such as the BASC-2 allow us to gather important information regarding functioning that can further assist in conceptualization and discernment of motives behind noncredible performance. Of note, although the BASC-2 can be helpful in identifying emotional adjustment problems that may drive invalid responding, the validity indices may not actually be useful in detecting symptom exaggeration, at least in children presenting for neuropsychological evaluation. In a recent study with a large pediatric sample with mild TBI, we found that the validity scales infrequently fell in the invalid range and were not associated with PVT scores (Kirk, Hutaff-Lee, Connery, Baker, & Kirkwood, 2014). If additional mood assessment appears warranted, then specific diagnostic screening ratings might also be used (e.g., Children's Depression Inventory—Second Edition [CDI-2]; Kovacs, 2010).

Case Example: LS

LS was a 16-year-old male with an unremarkable developmental and medical history until age 14, when he began to complain of severe chronic headaches eventually diagnosed by his neurologist as migraines. He also had missed over 30 days of school over the past 2 school years. He had historically been an A–B student, but his grades now were mostly C's. LS also frequently complained of difficulties focusing and memorizing information. He also was displaying more irritability, mood swings, and isolative tendencies over the preceding year. LS was referred by his neurologist for a neuropsychological evaluation. On testing, LS's scores revealed subtle validity concerns. His performance on the MSVT revealed scores of 90%, 90%, and 85% on the three validity indices. TOMM scores were 48 and 47 on Trials 1 and 2, respectively. However, his Retention trial score was a 42. He was slow to recite the alphabet at 10 seconds. Additionally, LS denied any signs or symptoms of depression based on a self-report questionnaire (CDI-2); however, the clinical interview revealed many concerns regarding depressive symptomatology, including periods of very low motivation, having low energy, and being easily irritated by his family and peers, as well as some passive thoughts of suicide without any specified plan. Furthermore, during the clinical interview it was learned that LS had lost interest in playing his guitar, as well as in seeing friends, doing well in school, and playing sports. He also shared with the examiner that he was unsure of his sexual orientation and had been teased at school for this issue. He described "hating" school now and having few friends.

Overall, based on the clinical interview with LS and his mother, he was thought to be clinically depressed, despite the fact that he responded on performance-based tests in a noncredible fashion. Additionally, LS was thought to be struggling with his sexual orientation and was having difficulties with bullying and peer victimization at school. Hence, his noncredible performance on testing was concluded to be driven by a combination of internal distress (i.e., depression, difficulties surrounding sexual orientation) and avoidance of an aversive environment (i.e., school). This conceptualization allowed recommendations to be tailored to LS's clinical needs, including individual psychotherapy and consideration of an antidepressant, as well as the need to address the bullying with school personnel. The fact that LS failed PVTs was crucial in knowing how to interpret the test data but did not inform why he was performing noncredibly. The underlying motivations in his case were identified only by establishing good rapport with him and his mother and then systematically collecting sensitive data through comprehensive history taking and interviews.

CONCLUSIONS AND FUTURE DIRECTIONS

When a child or adolescent presents in a noncredible manner during a clinical assessment, practitioners must attempt to uncover the reasons for such a presentation given the significant implications the motivations have for case conceptualization and clinical management. In contrast to what is often seen in adult neuropsychological assessments, noncredible presentations in children are less likely to relate to obvious external incentives (e.g., monetary reward), and thus the motivations driving noncredible presentations in children can be difficult to discern. As a larger proportion of pediatric neuropsychologists include objective validity testing in their test batteries, an improved understanding of the pediatric populations at highest risk for noncredible performance should become clearer. Future research will also need to focus on the specific factors that increase the risk for noncredible presentations, including personality, psychosocial, and learning characteristics of the individual child, family variables (e.g., parental marital status), and other environmental conditions (e.g., school match).

REFERENCES

Alexander, K. L., Entwisle, D. R., & Kabbani, N. S. (2001). The dropout process in life course perspective: Early risk factors at home and school. *Teachers College Record, 103,* 760–822.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.) Arlington, VA: Author.

Anderson, L. W., Jacobs, J., Schramm, S., & Splittgerber, F. (2000). School transitions: Beginning of the end or a new beginning? *International Journal of Educational Research, 33,* 325–339.

Baker, D. A., Connery, A. K., Kirk, J. W., & Kirkwood, M. W. (2014). Embedded performance validity indicators within the California Verbal Learning Test, Children's Version. *The Clinical Neuropsychologist*, *28,* 116–127.

Boone, K. B. (2007). Commentary on "Cogniform disorder and cogniform condition: Proposed diagnoses for excessive cognitive symptoms" by Dean C. Delis and Spencer R. Wetter. *Archives of Clinical Neuropsychology, 22,* 675–679.

Campo, J. V., Bridge, J., Ehmann, M., Altman, S., Lucas, A., Birmaher, B., et al. (2004). Recurrent abdominal pain, anxiety, and depression in primary care. *Pediatrics*, *113*, 817–824.

Carone, D. A. (2008). Children with moderate/severe brain damage/dysfunction outperform adults with mild-to-no brain damage on the Medical Symptom Validity Test. *Brain Injury*, *22,* 960–971.

Chafetz, M. D. (2008). Malingering on the Social Security Disability consultative exam: Predictors and base rates. *The Clinical Neuropsychologist, 22,* 529–546.

Chafetz, M. D., Abrahams, J. P., & Kohlmaier, J. (2007). Malingering on the Social Security disability consultative exam: A new rating scale. *Archives of Clinical Neuropsychology, 22*, 1–14.

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (1994). *California Verbal Learning Test—Children's version*. San Antonio, TX: Psychological Corporation.

Delis, D. C., & Wetter, S. R. (2007). Cogniform disorder and cogniform condition: Proposed diagnoses for excessive cognitive symptoms. *Archives of Clinical Neuropsychology*, *22*, 589–604.

Donders, J. (2005). Performance on the Test of Memory Malingering in a mixed pediatric sample. *Child Neuropsychology, 11*, 221–227.

Edwards, R. R., Augustson, E., & Fillingim, R. B. (2003). Differential relationships between anxiety and treatment-associated pain reduction among male and female chronic pain patients. *Clinical Journal of Pain, 19*, 208–216.

Flaro, L., & Boone, K. (2009). Using objective effort measures to detect noncredible cognitive test performance in children and adolescents. In J. E. Morgan & J. J. Sweet (Eds.), *Neuropsychology of malingering casebook* (pp. 369–376). New York: Psychology Press.

Flaro, L., Green, P., & Blaskewitz, N. (2007). Die bedeutung der beschwerdenvalidierung im kindesalter [The importance of symptom validity testing in children: WMT & MSVT]. *Praxis der Rechtpsychologie, 17,* 125–139.

Gervais, R. O., Green, P., Allen, L. M., & Iverson, G. L. (2001). Effects of coaching on symptom validity testing in chronic pain patients presenting for disability assessments. *Journal of Forensic Neuropsychology, 2,* 1–20.

Gilleland, J., Suveg, C., Jacob, M., & Thomassin, K. (2009). Understanding the medically unexplained: Emotional and familial influences on children's somatic functioning. *Child, 35,* 383–390.

Gonzalez, A., Boyle, M. H., Kyu, H. H., Georgiades, K., Duncan, L., & MacMillan, H. L. (2012). Childhood and family influences on depression, chronic physical conditions, and their comorbidity: Findings from the Ontario Child Health Study. *Journal of Psychiatric Research, 46*, 1475–1482.

Green, P. (2004). *Manual for the Medical Symptom Validity Test*. Edmonton, Alberta, Canada: Green's Publishing.

Greve, W., Etherton, J., Ord, J., Bianchini, K., & Curtis, K. (2009). Detecting malingered pain-related disability: Classification accuracy of the Test of Memory Malingering. *The Clinical Neuropsychologist, 23*, 1250–1271.

Henry, G. K. (2005). Childhood malingering: Faking neuropsychological impairment in an 8-year-old. In R. L. Heilbronnner (Ed.), *Forensic neuropsychology casebook* (pp. 205–217). New York: Guilford Press.

Janke, E. A., Holroyd, K. A., & Romanek, K. (2004). Depression increases onset of tension-type headache following laboratory stress. *Pain, 111,* 230–238.

Kearney, C. A. (2008). School absenteeism and school refusal behavior in youth: A contemporary review. *Clinical Psychology Review, 28,* 451–471.

Kearney, C. A., & Albano, A. M. (2004). The functional profiles of school refusal behavior: Diagnostic aspects. *Behavior Modification, 28,* 147–161.

Kirk, J. W., Hutaff-Lee, C., Connery, A. K., Baker, D. A., & Kirkwood, M. W. (2014). The relationship between the self-report BASC-2 validity indicators and performance validity test failure after pediatric mild traumatic brain injury. *Assessment, 21, 562–569.*

Kirkwood, M. W., Connery, A. K., Kirk, J. W., & Baker, D. A. (2013). Detecting performance invalidity in children: Not quite as easy as A, B, C, 1, 2, 3 but automatized sequences appears promising. *Child Neuropsychology*, *20*, 245–252.

Kirkwood, M. W., Hargrave, D., & Kirk, J. W. (2011). The value of the WISC-IV Digit

Span subtest in detecting noncredible performance during pediatric neuropsychological exam. *Archives of Clinical Neuropsychology*, *26*, 377–384.

Kirkwood, M. W., & Kirk, J. W. (2010). The base rate of suboptimal effort in a pediatric mild TBI sample: Performance on the Medical Symptom Validity Test. *The Clinical Neuropsychologist, 24,* 860–872.

Kirkwood, M. W., Kirk, J. W., Blaha, R. Z., & Wilson, P. E. (2010). Noncredible effort during pediatric neuropsychological exam: A case series and literature review. *Child Neuropsychology, 16,* 604–618.

Kirkwood, M. W., Peterson, R. L., Connery, A. K., Baker, D. A., & Grubenhoff, J. A. (2014). Postconcussive symptom exaggeration after pediatric mild traumatic brain injury. *Pediatrics, 133,* 643–650.

Kirkwood, M. W., Randolph, C., & Yeates, K. O. (2012). Sport-related concussion: A call for evidence and perspective amidst the alarms. *Clinical Journal of Sports Medicine, 22*, 383–384.

Kovacs, M. (2010). *Manual for the Children's Depression Inventory* (2nd ed.). North Tonawanda, NY: Multi-Health Systems.

Kozlowska, K., Nunn, K. P., Rose, D., Morris, A., Ouvrier, R. A., & Varghese, J. (2007). Conversion disorder in Australian pediatric practice. *Journal of the American Academy of Child and Adolescent Psychiatry, 46,* 68–75.

Krasnik, C. E., Meaney, B., & Grant, C. (2013). A clinical approach to paediatric conversion disorder: VEER in the right direction. Retrieved April 20, 2014, from *www.cpsp.cps.ca/uploads/publications/RA-conversion-disorder.pdf.*

Kugler, B., Bloom, M., Kaercher, L., Truax, T., & Storch, E. (2012) Somatic symptoms in traumatized children and adolescents. *Child Psychiatry and Human Development, 43,* 661–673.

Larrabee, G. J. (2003). Detection of malingering using atypical performance patterns on standard neuropsychological tests. *The Clinical Neuropsychologist, 17*, 410–425.

Lu, P. H., & Boone, K. B. (2002). Suspect cognitive symptoms in a 9-year-old child: Malingering by proxy? *Clinical Neuropsychology, 16,* 90–96.

McCaffrey, R. J., & Lynch, J. K. (2009). Malingering following documented brain injury: Neuropsychological evaluation of children in a forensic setting. In J. E. Morgan & J. J. Sweet (Eds.), *Neuropsychology of malingering casebook* (pp. 377–385). New York: Psychology Press.

McCune, N., & Hynes, J. (2005). Ten-year follow-up of children with school refusal. *Irish Journal of Psychological Medicine, 22,* 56–58.

Merikangas, K. R., & Rasmussen, B. K. (2000). Migraine comorbidity. In J. Olesen, P. Tfelt-Hansen, & K. M. Welch (Eds.), *The headaches* (pp. 235–242). Philadelphia: Lippincott Williams & Wilkins.

Mittenberg, W., DiGiulio, D. V., Perrin, S., & Bass, A. E. (1992). Symptoms following mild head injury: Expectation as aetiology. *Journal of Neurology, Neurosurgery, and Psychiatry, 55,* 200–204.

Mittenberg, W., Patton, C., Canyock, E. M., & Condit, D. (2002). Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experiential Neuropsychology, 24,* 1094–1102.

National Federation of Youth Sports Associations. (2010). 2009–2010 high school athletics participation survey results. Retrieved from *www.nfhs.org/content.aspx?id=3282*.

National Institute of Child Health and Human Development. (2001). Bullying widespread in U.S. schools, survey finds. Retrieved December 18, 2013, from *www.nih.gov/news/pr/apr2001/nichd-24.htm*.

Phillips, D. A. (2007). Punking and bullying: Strategies in middle school, high school, and beyond. *Journal of Interpersonal Violence, 22,* 158–178.

Ponsford, J., Cameron, P., Fitzgerald, M., Grant, M., Mikocka-Walus, A., & Schönberger, M. (2012). Predictors of postconcussive symptoms 3 months after mild traumatic brain injury. *Neuropsychology, 26,* 304–313.

Randolph, C., & Kirkwood, M. W. (2009). What are the real risks of sport-related concussion, and are they modifiable? *Journal of International Neuropsychological Society, 15,* 512–520.

Reynolds, C. R., & Kamphaus, R. W. (2004). *Behavior assessment system for children* (2nd ed.). Circle Pines, MN: American Guidance Service.

Rhee, H., Holditch-Davis, D., & Miles, M. S. (2005). Patterns of physical symptoms and relationships with psychosocial factors in adolescents. *Psychosomatic Medicine*, *67*, 1006–1012.

Roberts, K. P., Lamb, M. E., & Sternberg, K. J. (2004). The effects of rapport-building style on children's reports of a staged event. *Applied Cognitive Psychology, 18*, 189–202.

Schulte, I. E., & Petermann, F. (2011). Familial risk factors for the development of somatoform symptoms and disorders in children and adolescents: A systematic review. *Child Psychiatry and Human Development, 42,* 569–583.

Siegman, A. W., & Reynolds, M. (1984). The facilitating effects of interviewer rapport and the paralinguistics of intimate communications. *Journal of Social and Clinical Psychology, 2,* 71–88.

Slick, D. J., & Sherman, E. M. S. (2012). Differential diagnosis of malingering and related clinical presentations. In E. M. S. Sherman & B. L. Brooks (Eds.), *Pediatric forensic neuropsychology* (pp. 113–135). New York: Oxford University Press.

Suhr, J. A., & Gunstad, J. (2002). "Diagnosis threat": The effect of negative expectations on cognitive performance in head injury. *Journal of Clinical and Experimental Neuropsychology, 24,* 448–457.

Thieme, K., Turk, D. C., & Flor, H. (2004). Comorbid depression and anxiety in fibromyalgia syndrome: Relationship to somatic and psychosocial variables. *Psychosomatic Medicine, 66*, 837–844.

Tombaugh, T. N. (1996). *Test of Memory Malingering.* Toronto, Ontario, Canada: Multi-Health Systems.

Turner, S. M., Hersen, M., & Heiser, N. (2003). The interviewing process. In M. Hersen & S. M. Turner (Eds.), *Diagnostic interviewing* (3rd ed., pp. 3–20). New York: Kluwer Academic/Plenum.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children* (4th ed.). San Antonio, TX: Psychological Corporation.

Wechsler, D. (2011). *Wechsler Abbreviated Scale of Intelligence* (2nd ed.). Bloomington, MN: Pearson.

8

Managing Noncredible Performance in Pediatric Clinical Assessment

AMY K. CONNERY
YANA SUCHY

This chapter offers guidance on managing issues of performance validity in pediatric evaluations from the beginning of the exam, *before* noncredible presentation is identified, and through the completion of the evaluation, when feedback is provided both in person and in written formats. Specifically, we focus on three issues. First, we discuss providing instruction to patients at the beginning of the evaluation session so as to elicit their full effort. We address the level of specificity given in this instruction and the implications of a general versus specific warning about the inclusion of performance validity tests (PVTs) in the test battery. Second, we review decisions the clinician must make when a patient presents noncredibly. Considerations for continuing versus discontinuing the exam and intervening versus not intervening with the patient when validity concerns arise are addressed. Last, we provide a model for giving feedback to caregivers and children after noncredible performance has been determined during a clinical evaluation. Sample language for the clinician to utilize during the feedback session and in written reports is provided.

INSTRUCTIONS AT THE BEGINNING OF THE EVALUATION

Given that the ethics code of the American Psychological Association (APA, 2002) explicitly calls for informed consent for assessment or diagnostic services, clinicians should carefully consider whether patients need to be informed (i.e., "warned") about the inclusion of PVTs in the assessment battery. Although some clinicians hold the view that, from an ethical standpoint, all aspects of an evaluation (including the usage of PVTs) should be shared in detail and a priori with patients, others feel that warning patients about the use of PVTs is tantamount to warning patients about, for example, tests of incidental memory or rule changes on the Wisconsin Card Sorting Test. Such warnings would, of course, render these tests uninterpretable, as they would fundamentally change the construct that is being assessed. The APA ethics code states only that "Informed consent includes an explanation of the nature and purpose of the assessment, fees, involvement of third parties and limits of confidentiality" (Standard 9.03), leaving it up to clinicians to decide how much detail should be provided about the "nature and purpose" of the evaluation.

In line with the general guidance offered by the APA ethics code, no professional consensus exists about the level of specificity that should be provided to patients regarding the administration of PVTs. In the absence of clear guidelines, clinical practice spans the gamut (Sharland & Gfeller, 2007). Some clinicians offer a general "warning," which commonly consists of an instruction to the patient to put forth full effort throughout the exam. Others offer a more specific warning, which may include an instruction to provide good effort with an explanation that "effort and truthfulness will be assessed" (Bush, 2009, p. 532). Others yet may provide an even more specific explanation that tests designed to detect feigning or exaggeration will be administered (Sharland & Gfeller, 2007). Additionally, although some clinicians include some degree of warning regarding the administration of PVTs in the consent form signed at the beginning of the evaluation, others do not (Bush, 2005; Carone, Iverson, & Bush, 2010).

Although no pediatric studies have examined whether providing a specific warning reduces noncredible performance, several adult analogue studies have focused on the topic, with mixed results. On the one hand, Johnson and Lesniak-Karpiak (1997) reported that when simulators were warned, performance improved relative to those who were not warned, although not to the level of nonsimulating controls. This finding suggests that simulators continued to "malinger," with their feigned performance becoming more subtle and, consequently, possibly more difficult to detect. On the other hand, Schenk and Sullivan (2010) found that a specific warning to simulators resulted in an improvement of performance to the levels of controls, suggesting that the warning caused participants to abandon their feigning attempts altogether. Another study (Sullivan & Richer, 2002) failed to find any impact of a specific warning on subsequent performance.

Although the inconsistencies in the preceding findings alone make it difficult to draw conclusions for clinical practice, additional limitations also need to be considered. The degree to which participants' responses in analogue studies parallel those of actual patients is not known. Additionally, the degree to which study participants or actual patients alter their behavior in response to warnings likely depends on various situational factors, such as the exact study design for research participants and a host of psychological/intrapersonal and social/interpersonal factors for patients. Therefore, it is virtually impossible to utilize available research to predict how a given patient would respond to a specific warning.

Although relatively few neuropsychologists report providing a specific warning in practice surveys (22%), most state that they *always* provide a general instruction to patients to give their best effort during the assessment (Sharland & Gfeller, 2007; Slick, Tan, Strauss, & Hultsch, 2004). Such nonspecific instruction is, of course, in line with good clinical practice and is included in most test manuals. Because even fairly specific warnings seem to have *no* measurable impact on test performance in healthy volunteers giving full effort (Etherton & Axelrod, 2013), it is believed that general warnings may at most facilitate good effort in nonfeigning examinees and are unlikely to alter the performance of feigning examinees. Importantly, general warnings are also unlikely to affect test sensitivity (Suhr, 2002).

Because research has shown that a specific warning may have the potential to alter examinee behavior (Johnson & Lesniak-Karpiak, 1997; Schenk & Sullivan, 2010), clinicians should be thoughtful about whether this is an acceptable outcome within the context of a given referral question. In some situations, a suspected exaggeration of impairment may itself be the target of potential clinical intervention. In such situations, obtaining a snapshot of the patient's typical presentation outside of the exam room (even if it involves invalid cognitive performance) may well be desired. For example, in a clinical exam of a child with a suspected somatization disorder whose school absences and somatic/cognitive complaints occur in the context of a distant history of an uncomplicated mild traumatic brain injury (mild TBI), the patient's performance on objective measures of performance validity may itself be useful in guiding clinical management decisions. In such situations, clinicians may prefer to provide only a very general warning.

In other instances, valid performance on neuropsychological measures is absolutely paramount for clinical decision making. For example, valid data are critical in presurgery epilepsy evaluations or in evaluations of the effects of a specific medical intervention. In such situations, providing a specific warning that may help induce the patient to put forth maximum effort could well be indicated.

While the debate continues as to the most appropriate level of specificity in the warning regarding the administration of validity tests, a general consensus exists that this instruction should be provided at the beginning of the testing session. A warning directly before a PVT is administered would likely alert the patient to which tests are PVTs, thereby compromising test security and reducing the tests' utility (Youngjohn, Lees-Haley, & Binder, 1999). Such a warning could be viewed as being in violation of the

APA ethics code, which states that "Psychologists make reasonable efforts to maintain the integrity and security of test materials and other assessment techniques" (Standard 9.11).

In brief, whereas providing a general warning is in line with good clinical practice and grossly aligns with the informed consent requirements put forth by the APA, provision of a specific warning is open to some question. Clinicians should be aware that specific warnings may alter subsequent test behavior in an unpredictable manner, and the impact on performance in any given patient cannot be known. Additionally, specific warnings may result in "more sophisticated malingering" rather than bona fide normalization of performance and effort (Youngjohn et al., 1999; p. 512), thereby leading to reduction in the sensitivity of PVTs (Schenk & Sullivan, 2010). Last, for purposes of test security, any warning, whether specific or general, should be provided prior to and not during the testing session.

DECISIONS TO BE MADE WHEN VALIDITY CONCERNS ARISE

Clinicians commonly alter the original evaluation plan when questions of performance validity arise. In a Slick et al. (2004) survey of forensic practices, when validity was in question, 58% of respondents stated that they changed their assessment plan in some way. Decisions regarding the course of action in cases of noncredible performance should be made prior to the start of the evaluation (Bush, 2009). Making these decisions before the exam affords the clinician sufficient time to carefully consider the many complex issues involved in an alteration of the original assessment plan and allows such alterations to proceed more smoothly. The typical decisions that need to be made include whether the exam should be continued or discontinued and whether or not the clinician should intervene in the moment with a patient providing noncredible data.

Continuing versus Discontinuing the Exam

For a number of reasons, some experts advocate continuing the exam as planned even if a patient performs noncredibly. Continuing the exam allows clinicians to establish, at a minimum, the lower threshold of neuropsychological capacity, which may offer valuable information (Bush, 2009). In addition, continuing the exam without an interruption avoids alerting patients as to which tests were PVTs, thereby also minimizing the potential for more sophisticated feigning (Youngjohn et al., 1999). Finally, continuing the exam potentially provides additional information regarding the validity of the patient's presentation and performance, which could not be accomplished if testing were discontinued (Bush, 2009; Carone et al., 2010).

Other experts believe that the purpose of the exam is *not* to establish a minimum level of neurocognitive ability based on results of questionable or unknown validity, but rather to ascertain *optimal* performance. From that standpoint, one can argue that there is no clinical justification to continue with the exam (Bush, 2009). In fact, if the exam is continued as planned, the examinee is exposed to more neuropsychological tests, which could potentially limit the utility of those tests in a future evaluation.

Several factors should be considered when deciding whether or not to continue with an exam. First, one needs to consider the *context* (e.g., clinical vs. forensic) of the evaluation. For example, in a forensic exam, it may not be sensible for the clinician to veer from the original assessment plan (see Donders, Chapter 13, this volume, for comprehensive information about validity issues in forensic examinations). In contrast, in a purely clinical evaluation, the clinician may have more flexibility. The *length* of the planned exam might also be a consideration. In comparison to a many-hour comprehensive evaluation, it may make more sense during an abbreviated evaluation to continue with the originally planned evaluation, thereby potentially gaining additional clinically relevant information while limiting patient exposure to measures that may be useful in a subsequent evaluation.

The principal *goal* of the evaluation should also be considered. As discussed previously, although in some situations the clinician needs to obtain information about the patient's motivations or fluctuations in effort (e.g., a suspected somatization disorder), in others a true picture of the patient's cognitive status is paramount (e.g., epilepsy presurgery evaluations). Thus, although continuing the exam in the former case may be indicated, in the latter case the clinician may opt to stop the exam in order to preserve the integrity of a possible future evaluation.

Intervening versus Not Intervening with the Patient

If concerns arise about test behavior or task engagement after the evaluation has begun, some clinicians advocate for a brief intervention with the patient. Most pediatric practitioners would agree that an intervention with a child who seems tired, hungry, or oppositional and consequently is not giving his or her best effort is typically warranted. In such situations, a frank discussion with the patient, providing a break, snack, or whatever else may be needed to address the underlying motive for the noncredible effort is common practice. However, in some situations the reasons for the poor performance may not be entirely clear or may be attributed to an effortful attempt to feign impairment. In such situations, the decision to intervene is more complex.

In support of intervention, some anecdotal evidence, as well as several published case studies of adults and children, suggests that intervening may lead patients to subsequently give valid performances (Bush, 2009; Donders, 2011; Osmon & Mano, 2009). Similar findings are suggested by research that has more systematically examined the effect of an intervention on test performance in group studies (Montague, Long, Stanford, & Pulsipher, 2014; Suchy, Chelune, Franchow, & Thorgusen, 2012). In the first case–control study of this issue in adults, Suchy et al. (2012) noted that after confronting patients regarding noncredible performance, subsequent performance on PVTs and memory tests improved to the level of controls in two-thirds of the cases. Importantly, normalization of scores on the failed PVT when it was administered again was associated with apparent normalization of subsequently administered memory tests. Thus normalization of PVT scores in a given adult patient provides some degree of certainty that subsequent memory performance may also be valid. Similarly, in a recent study with adolescents, Montague et al. (2014) demonstrated that intervention with pediatric patients after initial validity concerns resulted in improved subsequent performance. There is no research to date on the impact of intervention on younger children.

Despite these encouraging findings, intervention or confrontation with patients during testing has been criticized as potentially having the same result as provision of a highly specific warning (discussed previously). In fact, many of the same issues in determining whether to provide a general or specific warning before testing are relevant here. In particular, clinicians who choose to confront patients need to take care to avoid linking their concerns to specific tests. Relatedly, clinicians who plan to confront patients after failed PVTs need to be careful about the type of information they provide during the informed consent procedures; that is, they should avoid telling their patients that specific tests of efforts are embedded in the assessment battery. Rather, they should provide only a fairly general vague warning about effort, and their confrontation should be equally general or vague.

Although performance does improve in some patients, as noted anecdotally and in the aforementioned studies, evidence suggests that it does not improve in *all* patients (Suchy et al., 2012). Even among those who do evidence an improvement, it is not certain that performance reaches a level that is fully comparable to those for whom no concerns about validity have been present or that it is truly representative of the patient's "best" performance. That is, the possibility always exists that an intervention can result in more sophisticated feigning or exaggeration (Bush, 2009; Youngjohn et al., 1999). Thus, following a confrontation, continued monitoring of performance validity, both by repeating the failed PVT (see Suchy et al., 2012) and by including additional PVTs within the remainder of the assessment battery, is strongly indicated.

In brief, when a PVT has been failed, clinicians need to carefully consider the context, length, and purpose of the evaluation when deciding whether to continue or discontinue the evaluation and whether to intervene or not intervene by confronting the patient. Issues to consider include (1) the possibility that test security may be threatened by stopping the evaluation or confronting the patient (i.e., the patient will have surmised which tests are PVTs), (2) the degree to which additional exposure to tests may limit the opportunity of future unbiased evaluations, (3) the degree to which continued assessment may or may not yield clinically relevant information, and (4) the degree of certainty one can assign to performance following a confrontation. These issues need to be carefully considered within the context of a *specific patient* and a specific *referral question*. Finally, the timing and the verbiage utilized, as well as the synergistic impact of specific a priori warnings in combination with discontinuing or confrontation, should be carefully considered (Suchy et al., 2012; Youngjohn et al., 1999).

FEEDBACK AFTER NONCREDIBLE PERFORMANCE

General Considerations for Providing Feedback

Feedback after noncredible data is identified is especially complex and delicate because the nature of the information being communicated (i.e., that test results are invalid due to a certain approach to the task by the child) can create confusion or potential conflict with (or within) the family. This potential for conflict, as well as the possible lack of regular or extensive experience providing this type of feedback, can result in clinician discomfort and failure to administer PVTs in order to avoid these circumstances (Carone et al., 2010). Careful consideration in establishing rapport will help the feedback session proceed more smoothly, reduce instances of overt conflict with patients and families, and help to make the feedback a clinically useful intervention.

Developing good rapport with the child and caregivers from the beginning of the evaluation is essential to a productive feedback session. If caregivers feel that the clinician is knowledgeable regarding the presenting condition, as well as willing to listen in a nonjudgmental way to the current symptoms and the complete story provided by the patient, they are generally more receptive to the message that the child's test performance was noncredible.

In this regard, clinicians should avoid expressing doubt or disbelief regarding the child's reported symptoms or presentation until testing is completed and all information and data are collected. Caregivers can be less receptive to the message that the child failed PVTs if they perceive that the clinician's mind was made up regarding the validity of the child's presentation before the exam even began. Thus, although the information the family provides can be appropriately queried and information about the natural clinical history of a disorder or an injury can be provided if applicable, direct communication of disbelief should be avoided (Carone et al., 2010).

Because children and adolescents are often unwilling or unable to discuss their noncredible performance openly, it can be easier to explain test results to caregivers and to explore the potential reasons for PVT failure without the child present. Therefore, at the initial intake appointment, it is often helpful to state that the feedback session may begin with caregivers alone if the child's inclusion in the feedback session has been planned or is desired by the family. This way, if the child provides noncredible data, the family will not view starting the feedback without the child as a deviation from the original plan, and the feedback can begin without conflict or mistrust. Feedback that is directly accusatory of the child (e.g., "your child is lying or faking") or too indirect (e.g., "your child was not engaged with the testing") should be avoided, as these approaches can lead to defensiveness on the part of the caregiver or misunderstanding of test results (Carone et al., 2010).

Feedback Model

There is a small literature on providing feedback to adult patients regarding noncredible neuropsychological performance (Carone et al., 2010; Bush, 2009; Tombaugh, 1996). Carone et al. (2010) outline a model for providing such feedback and offer both general and specific ideas for how to frame results to the patient. Although some aspects of the model are applicable to children, a direct downward extension of the model is not possible in the pediatric context because test results must be explained to the caregiver, who was not the one who gave the noncredible performance. This presents unique challenges not encountered in the adult context.

Connery, Baker, Peterson, and Kirkwood (2013) have developed a pediatric model for conducting a caregiver feedback session after noncredible performance has been determined during a clinical evaluation. This model can be implemented when a decision has been made not to intervene with the child during the evaluation. As highlighted in Figure 8.1, in this model, the feedback session usually begins with a general opening statement to the caregiver introducing the idea of validity and providing an initial explanation of the child's noncredible performance. The child is not present, as discussed earlier. Four common caregiver responses to the opening statement are described. For each response, Figure 8.1 offers an example of a possible clinician reply that, in our clinical experience, can help support caregiver understanding of what occurred during the evaluation. Potential reasons for the noncredible performance are then explored with caregivers. An explanation of how to understand this finding is provided, and recommendations for management are made.

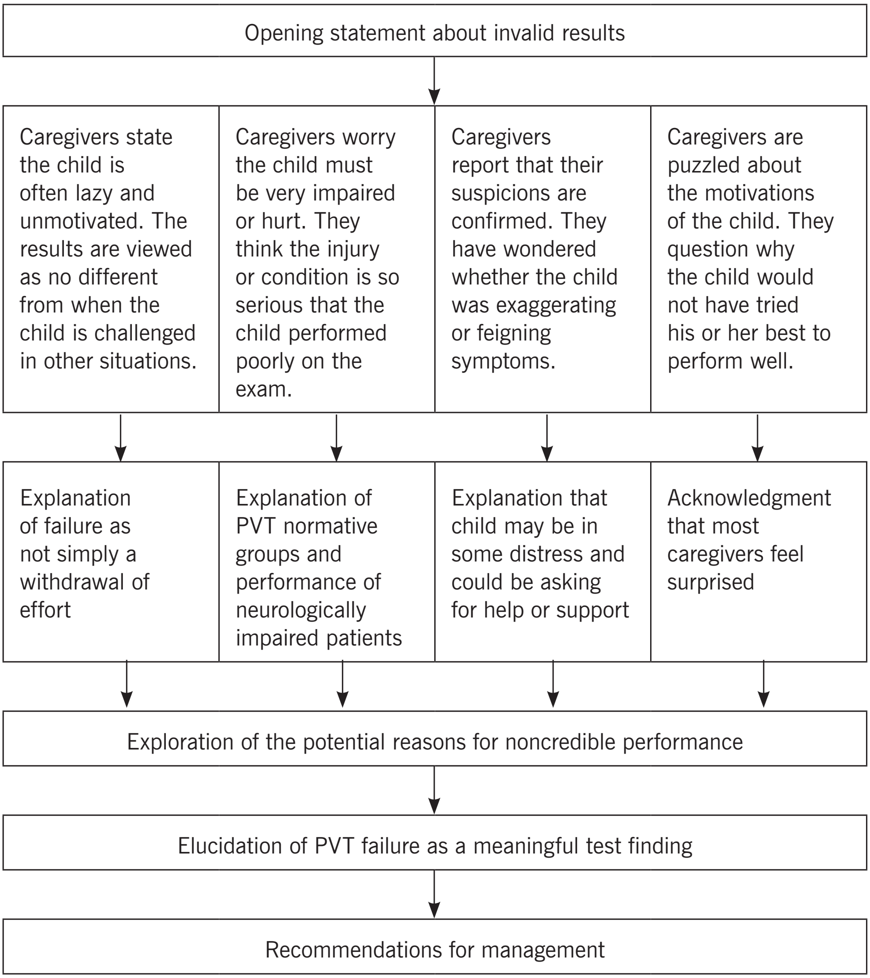

FIGURE 8.1. A pediatric feedback model for addressing noncredible performance.

Opening Statement

The model begins with an opening statement that sets the stage for the feedback by providing a message to caregivers that is direct and nonaccusatory and that gives a preliminary explanation of test results:

> "Whenever we do these evaluations, we give tests that measure whether children are performing to the best of their ability in order to make sure the test results are valid. In other words, when children do not do well on testing, we want to make sure that it is due to an actual

weakness rather than to children not trying their best. During today's evaluation, these tests showed that your child was not always trying his [her] best to do well."

Common Caregiver Responses and Clinician Explanations

Although of course responses are varied, we have found that there are four common ways in which caregivers respond to this opening statement. The first is that the caregiver reports that the child is indeed "lazy" and "never tries his [her] best." In this circumstance, the clinician's goal is to help the caregiver appreciate that the child's general apathy or inconsistent effort in everyday tasks is unlikely to produce scores at the level seen in PVT failure. If the caregiver persists in conceptualizing the failure as a result of the child's "laziness," it can be beneficial to speak more directly to the potential increased effort an examinee might employ in PVT failure. As Slick and Sherman (2012) note, the examinee who fails PVTs may actually be exerting more effortful test-taking behavior than those who pass PVTs. They note:

> In addition to the normal demands of test-taking, feigning examinees may simultaneously engage in a variety of additional cognitive activities, such as consciously keeping a specific feigning strategy in mind, monitoring and assessing item difficulty, deciding when to make errors, deciding what type of errors to make, inhibiting correct responses, monitoring their performance over time in order to maintain a target error rate, and monitoring the examiner's reaction for clues about the believability of the feigning performance. (p. 117)

Case Example

A 15-year-old female presented for neuropsychological evaluation due to concerns about unusual symptoms of potential neurological origin. A few months prior to the evaluation, she experienced several episodes in which she reportedly was suddenly unaware of how she ended up in a certain place. She reported losing awareness for extended periods of time during these episodes. Notably, these episodes often involved her wandering off from an undesired activity to a desired one (e.g., leaving school to hang out at a friend's house). The patient had a comprehensive neurological workup, which was normal. Academic performance had been variable, with frequent complaints from parents and teachers regarding inconsistency in performance, not working at her potential, and not following through with assignments and tasks. The mother had been particularly resistant to medical providers' feedback that there was a non-neurological explanation for the symptoms.

During a neuropsychological exam, the patient failed multiple validity measures. The clinician made the opening statement to the mother at the feedback session. The mother sighed and stated that, as usual for her daughter, she simply did not try her best.

If the feedback session had ended there, the mother would likely have continued to believe that the test results had provided no valid or useful information regarding her daughter's functioning, just as her grades typically did not reflect her ability. It was important to support the mother in appreciating that this was not simply a withdrawal of effort and that there was now objective evidence, not just doctors' opinions, that helped to demonstrate that non-neurological factors were at play. This information helped the mother accept this explanation for the episodes. She halted their continued pursuit of expensive medical tests, put into place more academic supports for the child, and sought psychological support. With this better understanding, the episodes eventually stopped.

Another common response from caregivers to the opening statement is the concern that the child must be so impaired, injured, or in such extensive pain that performance was poor during the exam. When the child presents with a condition that would not be expected to produce such severe impairment, reassurance can be provided about the expected level of impairment and associated expectations for recovery. For example, after a mild head injury, the clinician might explain that the child's performance was lower than would occur with someone with more serious neurological injury or impairment (Carone et al., 2010).

Caregivers may persist in believing that this test finding must indicate that the injury was more severe than originally thought. At this point, it can be helpful to explain that many studies have shown that PVT performance is typically not related to neurological status or injury (Carone, 2008; Donders, 2005; Green, 2003; MacAllister, Nakhutina, Bender, Karantzoulis, & Carlson, 2009), intellectual problems (Green, Flaro, Brockhaus, & Montijo, 2012), or pain (Etherton, Bianchini, Greve, & Ciota, 2005; Iverson, LePage, Koehler, Shojania, & Badii, 2007). Given these research findings, the clinician can explain to caregivers that there is very likely a non-neurological or noninjury explanation for the low scores (Carone et al., 2010).

Case Example

A 10-year-old male sustained a concussion during a soccer game. Over the subsequent weeks, he reported ongoing headaches, decreased attention, memory problems, sleep difficulties, and a variety of other symptoms. His parents, recently separated, were quite concerned about his symptoms and had been reading alarming reports in the media about the seriousness of concussion and long-term problems, such as "chronic traumatic encephalopathy."

On a neuropsychological exam, the child failed multiple PVTs.

The parents responded to the opening statement with alarm, believing that the poor performance must mean that the injury was more severe than originally thought and that their fears regarding the child's long-term prognosis were realized.

The clinician helped the parents appreciate the natural clinical history of mild TBI and presented research regarding typical recovery. The clinician then provided additional detailed information regarding populations with more severe TBI and the performance of these patients on PVTs to help the parents understand that it was not plausible that a much less severe injury would result in worse performance.

As the parents came to understand that symptoms were likely not neurologically driven, they were able to explore the other potential reasons for the persistence of the child's reported symptoms. The clinician reassured the parents about the child's prognosis and supported them in providing the child with messages of expected health and recovery rather than worry and concern regarding permanent brain damage.

After the opening statement, a third common response from caregivers is that they knew all along that the child was "faking." In these instances, it can be important to explain that some children who present noncredibly may be communicating distress and could be asking for help or support. A primary goal for the clinician in this situation is to ensure that caregivers not dismiss the finding and become angry or annoyed at their child for wasting time or money on the evaluation. Again, the finding can be framed in the context of a call for help that it is now the responsibility of the adults to appropriately address.

Case Example

A 13-year-old female presented with complaints of headaches, ankle and back pain, fogginess, memory problems, and fatigue after ankle surgery to correct a fracture a few months previously. The parents were concerned that there was an unreported event during surgery or that the child was experiencing adverse effects from general anesthesia. She had not been back in school full time since the surgery. The child had a history of gastrointestinal problems of unknown etiology, headaches, and prolonged recovery from minor colds and illnesses. Surgical records did not demonstrate any adverse event during the ankle surgery.

On a neuropsychological exam, the child presented with noncredible performance on PVTs. Upon presentation of this information, the caregivers expressed frustration with the child and stated that they suspected the child was feigning but had not wanted to communicate to the child that they did not believe her.

In this case, it was important to help the caregivers appreciate that the child was likely in some psychological distress. Communication of this distress through somatic complaints was a concern and necessitated adult attention and not disregard. Understanding the child's symptoms as more complex than "faking" resulted in the caregivers' attention to the psychological aspects of the child's presentation. A psychological intervention was then planned to help reduce the severity of somatic symptoms, teach coping strategies, and support school reintegration.

The last most common response from caregivers to the opening statement is one of puzzlement: "Wow, I wonder why my child would do that. . . ." This response from caregivers seems to imply an understanding of the message of the opening statement. A general acknowledgement here that most parents are surprised regarding the finding and that it is not an uncommon finding (if that is appropriate in the context of the particular evaluation) is probably most helpful.

Case Example

A 17-year-old male presented for evaluation due to concerns about possible attention-deficit/hyperactivity disorder (ADHD). He had been successful academically over the years and had been taking all honors and advanced placement classes. He was consistently on the high honor roll. During his current, 11th-grade, year, he began to complain of problems focusing in the classroom, completing his homework, and resisting distractions during test taking. His grades had not suffered, but he expressed concerns to his parents that he may have ADHD. Symptoms of inattention were not noted before this year. Teachers had not expressed concerns.

On a neuropsychological exam, the patient failed PVTs. When his parents were informed of the test results, they appeared to clearly understand that the performance was not credible and was not their son's "best." They were able to explore with the examiner the possibility that this adolescent was under significant stress this school year with academic demands and college applications. They wondered whether he was interested in the potential of stimulant medication in helping him meet all these demands or give him a "leg up" when competing for grades and college placement. The parents were then able to think through how they might lower stress for this young man and provide him with some other ways of coping and managing his workload.

Exploration of the Potential Reasons for Noncredible Performance

The next phase of the model involves an exploration with the caregivers of the potential reasons that the child provided noncredible performance. This conversation can begin when the caregiver seems to have an adequate understanding of what occurred and can start with a simple question posed to the caregiver asking why he or she believes this might have occurred. Clinicians can then offer hypotheses about potential reasons for the noncredible performance or the child's complaints. A more detailed discussion of possible reasons for PVT failure by Baker and Kirkwood appears in Chapter 7 of this volume and can offer additional guidance in managing this part of the caregiver feedback session.

PVT Failure as a Meaningful Test Finding

After clinicians and caregivers collectively explore the potential reasons for noncredible test results, the next phase in the model is to explain to caregivers that PVT failure is a meaningful test finding. Clearly, this is not the usual outcome of neuropsychological assessment in which detailed information regarding neurocognitive functioning is obtained. However, if the clinician is tasked with helping to figure out what factors might be driving symptoms and playing a role in the severity or persistence of problems, then critical information to help answer this question has been acquired. The clinician can be reasonably certain that non-neurological factors are at least playing some role in the examinee's presentation and current reported symptomatology.

Recommendations for Management

In clinical settings, determination of noncredible presentation can help guide the development of recommendations for management. For example, the child with an undiagnosed learning disability who is receiving injury accommodations at school that are in fact beneficial in managing school demands may be referred for more extensive learning evaluation when appropriate. The child who appears to be overwhelmed and is seeking a reprieve from a demanding academic and sports schedule can be advised to find ways to cut back on current demands and expectations. Psychological support to address underlying psychological factors is also often indicated after noncredible presentation has been determined. Finally, it may be important to deemphasize further medical consultations and expensive follow-up medical tests if appropriate. In the context of the presenting problem, it can be helpful to promote positive expectations for the child's overall functioning going forward.

Feedback with the Child

When it seems clear that caregivers have a good, basic understanding of what occurred, why it occurred, and the plan going forward, the child is asked to join the feedback session. An opening statement from the clinician to the child is provided, which is an adaptation of suggested language outlined by Carone et al. (2010) for adult evaluations.

> "There were a number of good scores and also some poor scores. However, we do not believe that the low scores are due to your injury (or illness or condition), but rather appear to be due to problems consistently putting forth your best effort to do well. So, the good news is that we believe that brain damage or something bad like that is not causing your low scores and that if you had more consistently tried your best during the evaluation, you would have been capable of performing much better. What are your thoughts about that?"

In our experience, a child rarely admits to feigning or exaggeration. Occasionally, a child will not admit to feigning but will become upset during the feedback, offering tacit acknowledgement of what occurred and further evidence of the extent of distress. However, what most commonly occurs is that children shrug, indicate that they generally understand what the clinician has communicated, but do not have a comment on it or, alternatively, claim to have no idea what the clinician is talking about.

On rare occasions, the child may challenge the clinician and state that the symptoms being experienced are "real." Although previous research has shown that children who fail PVTs self-report more frequent and more severe symptomatology than those who provide valid performance (Kirkwood, Peterson, Connery, Baker, & Grubenhoff, 2014), it can be difficult for the clinician to accurately comment on self-reported symptom veracity (e.g., the frequency or severity of headaches). Therefore, engaging in this argument with the child often proves unproductive. If accurate and appropriate, clinicians can acknowledge that the message is not about the veracity of symptoms. The message is that the clinician believes that the low scores obtained on testing are not likely to be neurologically driven. Therefore, non-neurological factors are likely playing a role in the presentation, and they must be addressed to support resolution or improvement of the presenting concerns.

The feedback with the child can be terminated at this point, even if the child has not engaged in a direct discussion of the PVT failure, as most commonly occurs. In this model, the more important factor is that caregivers have a solid understanding of the exam results, because caregivers are typically responsible for planning and guiding treatment and conveying messages of health or sickness/injury to the child and to others involved in the child's education and care. Because the child was the one who engaged in the test behavior, he or she most likely already understands it to a certain extent even if the noncredible behavior is not acknowledged. To avoid having an accusatory or confrontational interaction with the child, a brief explanation of test results, as described earlier, is typically all that is necessary during the feedback session.

COMMUNICATION IN WRITTEN REPORTS

Iverson (2006) notes that because invalid test results are an important exam finding, neuropsychologists have an ethical responsibility to address these results in written communications. These can include an explanation that results are not consistent with injury severity, that data are invalid, that no firm conclusions can be drawn, or that results may indicate exaggeration or malingering (Sharland & Gfeller, 2007). Overall, clear language stating that the patient gave a noncredible performance should be used without attempting to minimize or confuse the issue (Iverson, 2006; Iverson & Slick, 2003). Three examples of sample language are provided below for this type of written communication in a clinical context:

1. "The patient's objective test performance ranged from well below average to intact. However, there is evidence that the patient did not perform consistently to the best of his or her ability, rendering the results invalid and the interpretation of poor performance uncertain. Overall, the neuropsychological test performance was difficult to characterize, largely because it did not cluster in a fashion consistent with what would be expected after this condition or any other type of neurological insult" (Connery et al., 2013).
2. "Performance was poor on tests and embedded indicators designed to measure the degree of effort applied to testing. Therefore, the test results are not considered an accurate reflection of the patient's optimal cognitive abilities" (Carone et al., 2010).
3. (sample language for clinical evaluation in which objective neuropsychological performance was valid, but symptom report was noncredible) "Regardless of these results, the current reported symptoms (e.g., memory loss for several years of the child's life) are unlikely to be explained neurologically, as this type of difficulty is not seen in the context of this type of injury/illness/developmental problem. Non-neurological factors (e.g., stress, anxiety) are almost certainly playing a role in these presenting (e.g., memory) complaints at this point" (Connery et al., 2013).

CONCLUSIONS AND FUTURE DIRECTIONS

There is a growing consensus regarding the utility of PVTs in pediatric neuropsychological evaluations. Although they provide invaluable information regarding the validity of the test data obtained, the use of PVTs must be approached thoughtfully, and many decisions regarding management of validity issues should be made before the exam begins. With strong considerations for the context, length, and goals of the evaluation and thoughtfulness about the complexity of these decisions, the clinician must decide whether or not to provide a general or specific warning about PVT administration at the beginning of the exam, whether to continue or discontinue the exam if validity concerns arise, and whether or not to intervene with a patient presenting noncredibly. Future research on the impact of these decisions on subsequent test performance in pediatric populations would help to maximize the clinicians' ability to make informed, evidence-based choices.

Although there is some research on providing feedback to adult patients after noncredible performance (Carone et al., 2010; Bush, 2009; Tombaugh, 1996), literature on this issue in a pediatric context is sorely lacking. Here we have attempted to provide some guidance to the clinician in conducting these feedback sessions with parents and children. Future research focused on how this type of feedback affects patient outcome and treatment management decisions is clearly warranted.

REFERENCES

American Psychological Association. (2002). American Psychological Association ethical principles of psychologists and code of conduct. Retrieved January 21, 2014, from *www.apa.org/ethics/code/index.aspx*.

Bush, S. (2005). Independent and court-ordered forensic neuropsychological examinations: Official statement of the National Academy of Neuropsychology. *Archives of Clinical Neuropsychology, 20,* 997–1007.

Bush, S. (2009). What to do after diagnosis? In J. E. Morgan & J. J. Sweet (Eds.), *Neuropsychology of malingering casebook* (pp. 530–539). New York: Psychology Press.

Carone, D. A. (2008). Children with moderate/severe brain damage/dysfunction outperform adults with mild-to-no brain damage on the Medical Symptom Validity Test. *Brain Injury, 22,* 960–971.

Carone, D. A., Iverson, G. L., & Bush, S. S. (2010). A model to approaching and providing feedback to patients regarding invalid test performance in clinical neuropsychological evaluations. *The Clinical Neuropsychologist, 24,* 759–778.

Connery, A. K., Baker, D. A., Peterson, R. L., & Kirkwood, M. W. (2013). *A pediatric feedback model for addressing noncredible performance*. Unpublished manuscript, Department of Physical Medicine and Rehabilitation, Children's Hospital Colorado, Aurora, CO.

Donders, J. (2005). Performance on the Test of Memory Malingering in a mixed pediatric sample. *Child Neuropsychology, 11,* 221–227.

Donders, J. (2011). Pediatric traumatic brain injury: Effects of questionable effort. In J.E. Morgan, I.S. Baron, & J. H. Ricker (Eds.), *Casebook of clinical neuropsychology* (pp. 183–190). New York: Oxford University Press.

Etherton, J. L., & Axelrod, B. N. (2013). Do administration instructions alter optimal neuropsychological test performance? Data from healthy volunteers. *Applied Neuropsychology, 20,* 15–19.

Etherton, J. L., Bianchini, K. J., Greve, K. W., & Ciota, M. A. (2005). Test of Memory Malingering performance is unaffected by laboratory-induced pain: Implications for clinical use. *Archives of Clinical Neuropsychology, 20,* 375–384.

Green, P. (2003). *Green's Word Memory Test: User's manual.* Edmonton, Alberta, Canada: Green's Publishing.

Green, P., Flaro, L., Brockhaus, R., & Montijo, J. (2012). Performance on the WMT, MSVT & NV-MSVT in children with intellectual disabilities and in adults with mild traumatic brain injury. In C. R. Reynolds & A. Horton (Eds.), *Detection of malingering during head injury litigation* (2nd ed., pp. 201–219). New York: Plenum Press.

Iverson, G. L. (2006). Ethical issues associated with the assessment of exaggeration, poor effort, and malingering. *Applied Neuropsychology: Adult, 13,* 77–90.

Iverson, G. L., LePage, J., Koehler, B. E., Shojania, K., & Badii, M. (2007). Test of Memory Malingering (TOMM) scores are not affected by chronic pain or depression in patients with fibromyalgia. *The Clinical Neuropsychologist, 21,* 532–546.

Iverson, G. L. & Slick, D. J. (2003). Ethical issues associated with the psychological and neuropsychological assessment of persons from different cultural and linguistic backgrounds. In I. Z. Schultz & D. O. Brady (Eds.), *Psychological injuries at trial* (pp. 2066–2087). Chicago: American Bar Association.

Johnson, J. L., & Lesniak-Karpiak, K. (1997). The effect of warning on malingering on memory and motor tasks in college samples. *Archives of Clinical Neuropsychology, 12,* 231–238.

Kirkwood, M. W., Peterson, R. L., Connery, A. K., Baker, D. A., & Grubenhoff, J. A. (2014). Postconcussive symptom exaggeration after pediatric mild traumatic brain injury. *Pediatrics, 133*(4), 643–650.

MacAllister, W. S., Nakhutina, L., Bender, H. A., Karantzoulis, S., & Carlson, C. (2009). Assessing effort during neuropsychological evaluation with the TOMM in children and adolescents with epilepsy. *Child Neuropsychology, 15,* 521–531.

Montague, E., Long, E. A., Stanford, L. D., & Pulsipher, D. T. (2014, February). *Optimizing test performance with feedback about effort for adolescents with mild TBI.* Poster presented at the meeting of the International Neuropsychological Society, Seattle, WA.

Osmon, D. C., & Mano, Q. R. (2009). Malingered attention deficit disorder: Effort, depression, and dependence in the pursuit of academic accommodations. In J. E. Morgan & J. J. Sweet (Eds.), *Neuropsychology of malingering casebook* (pp. 386–397). New York: Psychology Press.

Schenk, K., & Sullivan, K. A. (2010). Do warnings deter rather than produce more sophisticated malingering? *Journal of Clinical and Experimental Neuropsychology, 32,* 752–762.

Sharland, M. J., & Gfeller, J. D. (2007). A survey of neuropsychologists' beliefs and practices with respect to the assessment of effort. *Archives of Clinical Neuropsychology, 22,* 213–223.

Slick, D. J., & Sherman, E. M. S. (2012). Differential diagnosis of malingering and related clinical presentations. In E. M. S. Sherman & B. L. Brooks (Eds.), *Pediatric forensic neuropsychology* (pp. 113–135). New York: Oxford University Press.

Slick, D. J., Tan, J. E., Strauss, E. H., & Hultsch, D. F. (2004). Detecting malingering: A survey of experts' practices. *Archives of Clinical Neuropsychology, 19,* 465–473.

Suchy, Y., Chelune, G., Franchow, E. I., & Thorgusen, S. R. (2012). Confronting patients about insufficient effort: The impact on subsequent symptom validity and memory performance. *Clinical Neuropsychology, 26,* 1296–1311.

Suhr, J. A. (2002). Malingering, coaching, and the serial position effect. *Archives of Clinical Neuropsychology, 17,* 69–77.

Sullivan, K., & Richer, C. (2002). Malingering of subjective complaint tasks: An exploration of the deterrent effects of warning. *Archives of Clinical Neuropsychology, 17,* 691–708.

Tombaugh, T. (1996). *Test of Memory Malingering.* Los Angeles: Western Psychological Services.

Youngjohn, J. R., Lees-Haley, P. R., & Binder, L. M. (1999). Comment: Warning malingerers produces more sophisticated malingering. *Archives of Clinical Neuropsychology, 14,* 511–515.

9

Ethical Considerations in Pediatric Validity Testing

WILLIAM S. MACALLISTER
MARSHA VASSERMAN

The issue of validity testing in psychology and neuropsychology, particularly as it pertains to the care of children, often seems like a heated and controversial topic. The mere mention of "effort testing" in professional circles (e.g., neuropsychology listservs) is occasionally met with strong opinion, ire, and perhaps incredulousness that benevolent clinicians would doubt the intentions and motivation of their clients. Some individuals view practitioners who employ such tests as being motivated not by a desire to help their clients but rather by a desire to "catch" dishonest patients. We assert that this view is not only inaccurate but also ill considered. As discussed in prior chapters, it is now well established that children may fail to provide adequate effort in cognitive evaluations for a myriad of reasons. As such, the routine use of performance validity tests (PVTs) in pediatric assessments is not only wise, given that it provides a means to objectively check the accuracy of the data collected, but also consistent with standards set forth by the organizations overseeing the profession. Further, it is in alignment with the ethical guidelines governing the larger field of psychology.

The purpose of this chapter is to review professional standards and ethical considerations and guidelines as they pertain to validity testing in pediatric assessment. The chapter highlights relevant American Psychological Association (APA) ethics code standards as they apply to this issue and argues that validity testing can serve individual clients while simultaneously protecting the integrity of the field at large. The chapter concludes with an illustrative case example in which validity testing played a major role in data interpretation, case conceptualization, treatment recommendations, and outcome. APA ethics code guidelines relevant to the case are highlighted.

PROFESSIONAL STANDARDS FROM THE FIELD OF NEUROPSYCHOLOGY

The field of pediatric neuropsychology has grown exponentially in the past several decades. As part of this growth and evolution, there has been a growing recognition of the need to ensure accuracy and validity in the conclusions we draw from assessments. A major determinant of this validity is ensuring that appropriate effort was put forth by examinees. The field of forensic neuropsychology led the charge in the adoption of objective validity testing, but this trend quickly trickled down to clinical evaluations in which the motivations to put forth less than optimal effort are often less clear.

In 2005, the National Academy of Neuropsychology (NAN) published a position statement on the utility of validity testing in neuropsychology (Bush et al., 2005). The statement candidly begins by asserting that clinicians "are responsible for making determinations about the validity of the information and test data obtained during neuropsychological evaluations" (p. 419). The report goes on to clarify that the specific manner in which the determination of validity is confirmed can vary as a function of numerous factors.

As discussed in prior chapters, there are many methods by which validity is determined (e.g., clinical observation, mismatch between test performance and what is known about an individual's abilities outside of the test setting, and formal validity testing). The consensus statement on the assessment of effort, response bias, and malingering published by the American Academy of Clinical Neuropsychology (AACN; Heilbronner et al., 2009) provided further guidance on this issue but clearly stated that the use of psychometric indicators (i.e., formal tests rather than subjective clinical impressions) are the "most valid approach" (p. 1106). Although this report lists numerous tenets, an important point is that effort may vary across an evaluation, so multiple effort indices should be employed throughout the evaluation.

Essentially, both the position statement of NAN (Bush et al., 2005) and the consensus statement issued by AACN (Heilbronner et al., 2009) set the expectation that performance validity measures be employed routinely in all neuropsychological evaluations; this is true whether they are being conducted for legal/forensic purposes or are part of one's clinical practice. In short, validity testing should no longer be considered "optional."

It is important to note that neither statement specifically mentions how validity should be evaluated in pediatric assessments. However, this omission should not be interpreted to suggest that effort issues do not pertain to pediatric evaluations, or even that they are of lesser importance. Rather, the lack of specific direction about children merely reflects the fact that our knowledge of the utility and effectiveness of PVTs in younger populations was in its infancy when these statements were drafted. Fortunately, as has been elucidated in previous chapters, the last decade has seen a minor explosion of research documenting the effectiveness of PVTs in school-age children and adolescents. As such, there is an evolving acceptance that their use should be standard practice in the assessment of this young population. As we discuss, the use of validity testing is also in keeping with APA ethical guidelines.

PEDIATRIC VALIDITY TESTING AND THE APA ETHICS CODE

The APA's *Ethical Principles of Psychologists and Code of Conduct* (aka APA ethics code) consists of five general principles followed by more specific ethical standards (APA, 2002). Members of APA are committed to compliance with the APA ethics code as it applies to their scientific, educational, and professional roles as psychologists. Validity testing as part of cognitive evaluations is not specifically referenced in this treatise, but the routine use of such is clearly in keeping with the document's aspirations.

APA Ethics Code General Principles and Validity Testing in Children

General Principle A of the APA Ethics Code concerns beneficence and nonmaleficence. Broadly speaking, this principle states that psychologists should strive to provide benefit to those they work with and "do no harm" (which is further described in specific Standard 3.04, Avoiding Harm). The principle goes on to state that psychologists should strive to maintain the welfare and rights of those that they work with (APA, 2002). It may be difficult for one to envision a circumstance in which a failure to recognize invalid effort would result in harm being done to one's client; such a circumstance is more likely to be viewed as someone "getting away with something," gaining financially, or perhaps obtaining access to services that they really do not need. However, situations in which harm occurs certainly can, and do, arise.

Pediatric neuropsychologists frequently work with children with neurological disease, learning disabilities, attention deficits, brain injuries, and so forth. Such problems have clear ramifications for children's educational trajectories. Thus pediatric neuropsychological and psychoeducational assessments frequently help school systems appropriately place children and adolescents with these issues. However, if insufficient effort and engagement during testing are not identified, placement recommendations and services rendered to these students may be inappropriate and may serve them poorly in the long run, resulting in educational underattainment. For example, a child failing to provide adequate effort during intelligence testing may be inappropriately placed in a self-contained class. As such, the instructional level would likely be below the child's capabilities, and the attainment of new knowledge and skills would proceed along a shallower trajectory than would otherwise be the case. Taken to the extreme, this inappropriate placement could, in the long run, disallow access to educational and occupational opportunities that would otherwise be at the child's disposal. A teen receiving a special education diploma would not be eligible to take the SAT and, as a consequence, would be unlikely to be able to apply to college. In such a case, the ethical principle of beneficence and nonmaleficence was clearly not upheld, and validity testing could have ensured a more favorable outcome.

Perhaps a more dramatic example of where this ethical principle is relevant involves situations in which major medical decisions are guided by neuropsychological assessment. For example, in the evaluation of epilepsy surgery candidates, cognitive status is a determinant of surgical outcomes, with more cognitively impaired individuals being at lower risk for cognitive declines following surgical resection of the epileptogenic zone (Sherman et al., 2011). Typically, if neuropsychological functioning is preserved, these individuals are at risk for cognitive loss following surgery. As such, the neurosurgeon may opt for a more conservative resection in cases in which the risk is high; for example, a child with intact memory function in the context of intractable temporal lobe epilepsy may undergo a limited hippocampectomy in the interest of preserving postsurgical memory. The tradeoff in these cases is that an incomplete resection of the epileptogenic zone may decrease the likelihood that he or she will be seizure-free thereafter. Prior work has demonstrated that, although a rare occurrence, children with epilepsy undergoing neuropsychological evaluations do occasionally put forth inadequate effort (MacAllister, Nakhutina, Bender, Karantzoulis, & Carlson, 2009). A child not putting forth adequate effort during an assessment undertaken as part of a presurgical workup would often be deemed less likely to decline postsurgically. Therefore, a less conservative resection could potentially result in major cognitive deficits that could have been avoided had proper validity testing been employed during evaluation. As in our earlier example, this situation clearly illustrates an instance in which a clinician's failure to properly employ validity testing results in clear harm to the patient.

A less dramatic, but equally compelling, example of maleficence involves testing for attention-deficit/hyperactivity disorder (ADHD). A common referral question for most pediatric-oriented clinicians, ADHD evaluations are commonplace in neuropsychology and psychoeducational practice settings. Unfortunately, this is also an area in which the motivation for youth to underperform on cognitive testing may be quite high (see Harrison, Chapter 10, this volume). In such assessments, there is often the potential for secondary gain. Perhaps the strongest indication of this potential is the fact that over 20% of adults being evaluated for ADHD may exaggerate deficits, as evidenced by PVT failure (Marshall et al., 2010). For these individuals, the most likely impetus for symptom exaggeration on testing was the motivation to obtain stimulant medications. It stands to reason that teens undergoing evaluations may have similar motivations. Stimulant medications are often used as study aids by students, even in those with no clear medical need for such a medication (Teter, McCabe, Cranford, Boyd, & Guthrie, 2005). It is also well known that students with ADHD occasionally resell their medications to other students, thus leading to a "black market" supply of stimulant medications. Though these medications are largely considered safe, they are not completely without risk. Teens using such medications without medical need often have substance abuse issues more generally (Hartung et al., 2013). Practitioners not routinely employing validity testing in their assessments of children and teenagers being evaluated for ADHD may inadvertently provide access to medications that the youth do not actually need for medical purposes, increasing the risk for substance abuse and medication side effects, including, in rare cases, significant cardiac complications (Daly, Custer, & McLeay, 2008; Nymark, Hovland, Bjornstad, & Nielsen, 2008).

This latter example also speaks to principles B and C of the APA ethics code. Principle B discusses fidelity and responsibility of psychologists, noting that psychologists establish relationships of trust with those they work with, as well as maintain an awareness of their professional and scientific responsibilities to communities and science at large. Principle C discusses integrity, asserting that psychologists promote accuracy, honesty, and truthfulness in the science, teaching, and practice of psychology. In the ADHD assessment example, the failure to identify those feigning or exaggerating attention impairment provides a disservice to society, in that it artificially inflates the rate of ADHD in the community and contributes to the public's perception that ADHD is an overdiagnosed disorder and that medications are overprescribed in children and adolescents. In the extreme case in which a teen may resell his or her prescribed medications, the clinician failing to provide a validity assessment as part of his or her routine clinical work may also inadvertently contribute to illegal activities, clearly running counter to the ethical responsibilities of fidelity, responsibility, and integrity.

We would also encourage practitioners to consider the impact their assessments have on a broader audience. That is, when considering Principle B and interpreting the clause that insists that we "establish relationships of trust with those [we] work with," we should consider not only the child and family at hand but also the larger system in which the child is ensconced, including the school system and the greater academic community. When considering advocating for services in school (thus incurring costs for the school systems), it is incumbent on clinicians to make recommendations that are supported by valid data. Accordingly, validity testing becomes an important tool in ensuring a trustworthy relationship between school systems and clinicians. The same can be said when advocating for testing accommodations on standardized examinations such as the SAT, American College Testing (ACT) assessment, Graduate Record Examination (GRE), and so forth. In fact, validity testing is required as part of an accommodations assessment for the bar examination. In all likelihood, this requirement will eventually trickle down to other high-stakes exams such as the SAT and ACT (as it should). We would encourage practitioners to be "ahead of the curve" of this inevitable trend.

Principle C also highlights that psychologists do not "engage in fraud, subterfuge or intentional misrepresentation of fact" (p. 1062). Principle D, which speaks to justice, insists that clinicians take precautions to ensure that potential biases do not unduly influence their work. Though only the most disingenuous of psychologists would knowingly falsify data (i.e., commit outright fraud) in order to gain unwarranted services for a client (e.g., extended time on SATs), clinicians do occasionally feel pressure from well-meaning parents to overstate the needs of their patients when advocating for accommodations and services. In the interest of appeasing these parents and advocating for their clients, clinicians may unwittingly be biased. Thus validity testing in this group not only protects the larger community (in keeping with Principle B) but also keeps their own biases in check when making treatment recommendations (in keeping with Principle C).

APA Ethics Code Specific Ethical Standards and Validity Testing in Children

In addition to the general principles discussed previously, the APA ethics code includes specific ethical standards to guide psychologists (APA, 2002). Most relevant to the field of neuropsychology are the standards addressing assessments, which are discussed in ethical standards 9.01 through 9.11.

Ethical standards 9.01 and 9.06 discuss the bases of assessments and the interpretation of assessment results, respectively. These standards make several points that are relevant to validity testing in pediatric evaluations. Standard 9.01a states that psychologists must base their opinions "on information and techniques sufficient to substantiate their findings" (p. 1071). As the preceding chapters have clearly illuminated, the findings from neuropsychological or psychoeducational testing may not be considered "substantiated" if a formal assessment of the validity of test results was not undertaken. Standard 9.01c further states that psychologists may "provide opinions on the psychological characteristics of individuals only after they have conducted an examination of the individuals adequate to support their statements or conclusions" (p. 1071). Standard 9.06 speaks most directly to this issue. It states that, in interpreting the results of assessments, psychologists should consider the "purpose of the assessment as well as the various test factors," which include, among other aspects, test-taking abilities that may affect judgments or "reduce the accuracy" of test interpretation (p. 1072). We would assert, again, that the results from testing without objective assessment of validity must be considered tentative. Consistent with standards 9.01c and 9.06, if the evaluation lacks a formal assessment of validity, this fact should be noted, and the results should be considered only preliminary.

Interpretation of Standard 9.02 (Use of Assessments) is far less black-and-white, considering the relatively nascent nature of validity testing as it pertains to children. Standard 9.02a asserts that tests are used for purposes that are "appropriate in light of the research on, or evidence of, the usefulness and proper application of the techniques" (p. 1071). Standard 9.02b goes on to insist that psychologists should strive to use assessment instruments for which validity and reliability have been established in the population being assessed (APA, 2002).

Although major strides have been made in recent years to provide data on PVTs in various populations, there are clear gaps in our current knowledge. Tasks that work well in one population may not work well in others. This issue has been discussed at length elsewhere in this book (see Kirkwood, Chapter 5), but briefly, PVTs such as the Test of Memory Malingering (TOMM), Word Memory Test (WMT), and the Medical Symptom Validity Test (MSVT) have proven utility in numerous pediatric populations (Blaskewitz, Merten, & Kathmann, 2008; Carone, 2008, 2014; Constantinou & McCaffrey, 2003; Courtney, Dinkins, Allen, & Kuroski, 2003; Donders, 2005; Kirk et al., 2011; Kirkwood, Peterson, Connery, Baker, & Grubenhoff, 2014; Loughan & Perna, 2014; MacAllister et al., 2009; Perna & Loughan, 2014), even in populations in which one may expect major cognitive impairment such as epilepsy (MacAllister et al., 2009), congenital brain disease (Carone, 2014), and moderate to severe traumatic brain injuries (TBI; Carone, 2008). Thus these instruments can be considered reliable and valid in most pediatric and adolescent populations.

Other instruments, however, have shown utility in some populations, but not in others. For example, the embedded effort index of Reliable Digit Span has shown reasonable sensitivity and specificity in detecting suboptimal effort in a population with mild TBI, albeit at a lower cutoff than used in adult populations (Kirkwood, Hargrave, & Kirk, 2011). However, Reliable Digit Span has an unacceptably high false-positive rate when used in populations for which more substantial cognitive impairment is expected, such as epilepsy (Welsh, Bender, Whitman, Vasserman, & MacAllister, 2012) or in children with ADHD or learning disabilities (Vekaria, Vasserman, & MacAllister, 2013). Thus, in keeping with ethics code standards 9.02a and 9.02b, practitioners should choose their PVTs judiciously, considering research that fits the population characteristics most closely. In keeping with Standard 9.02b, when working with a population in which there is a paucity of data to guide the use of validity tests, one must take reasonable precautions to describe the limitations of test interpretations that can be made.

It is also important to consider age characteristics and requisite abilities needed to perform specific tasks. Currently, objective performance validity measures have been validated for school-age children and older. As illustrative examples, the TOMM has shown utility in children as young as 5 years of age (Donders, 2005), whereas a reading level of grade 3 or above is needed to perform adequately on the WMT (Green & Flaro, 2003). Accordingly, in keeping with standards 9.02a and 9.02b, patient characteristics should be considered carefully when choosing appropriate measures of effort. Unfortunately, no measures of performance validity are currently available for use in children under the age of 5. As such, clinicians evaluating these very young individuals will need to rely on clinical observation and judgment in determining whether or not test results can be considered valid, consistent with AACN guidelines, while recognizing the limitations of such an approach (Heilbronner et al., 2009).

Ethical issues pertaining to informed consent are discussed in Standard 9.03. This standard states that psychologists should solicit informed consent for assessments and that this consent should include explanation of the nature and purpose of the assessment. In discussing the nature of the evaluation, it is recommended that examinees be informed that, as part of the assessment, their effort will be assessed. Nevertheless, this should be done in broad strokes, without undue detail. It is advised that clinicians routinely mention that effort is being assessed and that examinees should perform at their best. More specific details regarding validity testing, however, are not only unwarranted but also ill advised (as discussed by Connery & Suchy, Chapter 8, this volume). Consistent with Standard 9.11 (Maintaining Test Security), practitioners must take steps to maintain the security and integrity of their assessment tools (APA, 2002). This point is further elaborated in the AACN position statement that discusses validity testing. The statement recommends that examiners explain the nature of their tests in a manner that allows them to describe the bases of their opinions but that they should avoid inclusion of "specific information pertaining to these measures that could preclude valid future use" (Heilbronner et al., 2009, p. 1106). Similarly, pursuant to standards 9.04 and 9.11, which discuss the release of data and the maintenance of test security, in giving data summary sheets to other clinicians, schools, and families, care should be taken to provide these data in a manner that does not compromise the future utility of these instruments. Release of the validity test protocols should be avoided under all circumstances, unless required by law (e.g., in the case of a formal court order). It is also worth noting here that researchers have curtailed the practice of including specific cutoff scores in reports and publications on PVTs in an effort to maintain the integrity of these instruments.

In nonforensic situations, in which feedback to parents and children or adolescents is expected, validity test failure is a potentially thorny issue. Standard 9.10 of the APA ethics code indicates that psychologists should take steps to explain test results to relevant parties. Of course, in instances in which validity test failure invalidates cognitive test results, a more nuanced approach to feedback may be necessary and may involve a discussion of the potential reasons a child or teen may not have put forth effort or may have been motivated to feign or exaggerate impairment. We recommend utilizing these important data to make meaningful recommendations to guide the client in the future. Although a detailed discussion of how to provide feedback to children and families is outside the scope of this chapter, this issue is addressed elsewhere in this book (Connery & Suchy, Chapter 8), as well as in earlier works (e.g., Postal & Armstrong, 2013).

In discussing validity testing with clinicians who have been reluctant to include freestanding PVTs in their routine pediatric assessments, one theme consistently emerges: Many are reluctant to use such instruments due to time constraints. In a field that is increasingly pressured to shorten evaluation time, this is not a trivial concern. Although we do not find this argument particularly compelling given the importance of ensuring accurate test results, coupled with the fact that validity can be checked fairly quickly, another solution is to rely on embedded validity tests, which we advocate that future editions of tests include. Standard 9.05 of the APA ethics code discusses test construction. This standard asserts that those who develop tests use appropriate psychometric procedures, as well as "current scientific and professional knowledge" relevant to "test design, standardization, and validation," to reduce bias (APA, 2002, p. 1072). Given the research that clearly demonstrates that children and adolescents do, at times, put forth less than optimal effort or engage in deception during evaluations, it seems wise for test developers to develop embedded validity indices for all tests when possible. In addition to potentially shortening overall testing time, doing this will allow ongoing assessment of effort throughout testing, as recommended by the AACN (Heilbronner et al., 2009), rather than at merely one or two time points.

Other Relevant Standards

Aside from the preceding ethical principles that relate directly to clinical assessments, psychologists are active in other areas in which validity testing is relevant. Likewise, there are ethical guidelines to help psychologists learn how to conduct themselves in research and education.

In conducting research on various diseases, validity testing should serve an important role. Standard 8.10 discusses the reporting of research results and firmly asserts that "psychologists do not fabricate data" (APA, 2002, 1070). It has been shown that a nontrivial number of children and adolescents undergoing cognitive evaluations will provide invalid effort. For example, in a sample of children with mild TBI, noncredible effort and symptom exaggeration are not uncommon (Kirkwood & Kirk, 2010; Kirkwood et al., 2014). Studies attempting to document potential long-term consequences of concussions should be informed by this research. Of course, failing to assess effort and outright fabrication of data are not equivalent. However, in the interest of accurately characterizing neurobehavioral outcomes, it certainly seems prudent to consider the role of effort in all research endeavors that involve neuropsychological testing in populations for which insufficient effort is a known factor.

Standards 7.01 and 7.03 raise relevant issues with respect to the training of future psychologists. Standard 7.01 (Design of Education and Training Programs) insists that psychologists developing training programs should ensure that their course sequences provide appropriate knowledge and proper experiences for trainees such that they are competent practitioners in the future. Standard 7.03 (Accuracy in Teaching) goes on to state that psychologists should ensure that course syllabi are accurate regarding the subject matter to be covered (APA, 2002). Given our increased knowledge base related to symptom exaggeration, noncredible effort, and the utility of validity testing in pediatric populations, it is essential that all pediatric training programs include coursework and training experiences that discuss these issues. Table 9.1 details ethical standards relevant to validity testing in children.

CASE EXAMPLE

The following case example is presented to highlight several important ethical considerations related to validity testing in pediatric neuropsychological assessments. Justin (pseudonym) was a young teenage boy at the time of his initial neuropsychological assessment. He was referred secondary to a history of academic, behavioral, and mood concerns. As Justin was adopted, biological family history and early developmental history were

TABLE 9.1. APA Ethical Principles and Relation to Validity Testing in Pediatric Neuropsychology

Ethical principle no.	Main tenet(s) of ethical principle	Relevance to validity testing
7.01	Psychologists developing training programs should ensure that their course sequences provide appropriate knowledge and proper experiences for trainees, such that they are competent practitioners in the future.	Given increased knowledge of symptom exaggeration, noncredible effort, and the utility of validity testing in pediatric populations, pediatric neuropsychology training programs should include related coursework and training experiences.
7.03	Psychologists should ensure that course syllabi are accurate regarding the subject matter to be covered.	Students should receive instruction and training regarding base rates of noncredible effort in various clinical groups and appropriate methods to detect this and how to respond when this occurs.
8.10	Reporting of research results and accuracy of results	Inclusion of validity testing in research protocols will ensure the accuracy of data.
9.01	Psychologists base their opinions on information and techniques sufficient to substantiate their findings.	Without validity testing to determine the accuracy of results, findings cannot be definitively considered "substantiated."
	Psychologists may provide opinions on the psychological characteristics of individuals only after they have conducted an examination of the individuals adequate to support their statements or conclusions.	Without assessment of validity, neuropsychologists cannot definitively determine that their findings are adequately supported, and findings should be considered only tentative.
9.02	Tests are used for purposes that are appropriate in light of the research on, or evidence of, the usefulness and proper application of the techniques. Psychologists should strive to use assessment instruments for which validity and reliability have been established in the population being assessed.	In choosing performance validity tests, neuropsychologists should consider validation research relevant to the tools they use. For example, tools such as the WMT, MSVT, and TOMM have been validated in children and adolescents with severe cognitive impairment, whereas embedded indices (e.g., Reliable Digit Span) have shown utility in some populations (e.g., mild TBI) but not more neurologically/neuropsychologically involved populations.
9.03	Psychologists should solicit informed consent/assent prior to assessments. Consent/assent should explain the nature and purpose of the assessment.	As part of consent process, examinees should be informed that the evaluation will include assessment of their effort. (However, see also 9.04 and 9.11 below.)

(*continued*)

TABLE 9.1. (*continued*)

9.04	Release of data	As part of their report and/or data summary sheets, neuropsychologists should ensure that specific information is not included regarding PVTs that would violate test security and/or preclude valid future use. (See also 9.11 below.)
9.05	Test construction	Test developers are encouraged to construct embedded validity indices for tests where possible. This will allow ongoing assessment of effort throughout testing.
9.06	In interpreting results, psychologists should consider the purpose of the assessment, as well as the various test factors, which include test-taking abilities that may affect judgments or "reduce the accuracy" of test interpretation.	Results from neuropsychological testing must be considered tentative until validity of these findings is determined. If the evaluation lacks assessment of validity, this should be noted, and the results should be considered only preliminary.
9.10	Psychologists should take steps to explain test results to relevant parties.	In cases in which PVT failure suggests invalid results, feedback may involve a discussion of the potential reasons a child or teen may not have put forth effort or have been motivated to feign or exaggerate deficits.
9.11	Psychologists take steps to maintain the security and integrity of their assessment tools.	Though neuropsychologists should explain the nature of their tests as part of their informed consent process, they should avoid specific information about PVTs that could preclude valid future use.

largely unavailable. It was known that he achieved developmental milestones on time but struggled with fine motor skills.

Justin has attended private schools since entering kindergarten. His grades were largely average. However, he required significant support from parents and tutors. Further, teachers expressed long-standing concerns regarding his attention, distractibility, poor organization, and time management. Difficulties with independent work completion, as well as in mathematics, were also noted by parents and teachers. He was diagnosed with ADHD, predominantly inattentive subtype (ADHD-I), by a psychiatrist when in grade school and has been treated with a stimulant medication since that time. Further, Justin's parents also reported multiple concerns about his mood.

Due to ongoing academic and attention difficulties, Justin was referred for a neuropsychological evaluation to determine appropriate educational placement and interventions. During the evaluation session, Justin quickly engaged in conversation and was open and forthright with information about his academic difficulties, mood symptoms, and self-harm behaviors. However, upon initiation of testing, concerns about performance validity quickly surfaced. His performances on the first few subtests of an intelligence measure, which were at the level expected for a teen with mild intellectual disability, were deemed implausible. At that time, PVTs were administered, with results falling substantially below passing. For example, his performance on the TOMM was far below chance (Trial 2 = 13 correct), and he failed the Dot Counting Test. Thereafter, testing was discontinued, and his mother was consulted regarding his performance. Upon questioning, Justin admitted that he was underperforming. Interestingly, he seemed rather incredulous that we were able to detect this, despite his rather unsophisticated approach (i.e., below-chance performances). A short-term plan was formulated that involved discussing the assessment with his current therapist.

Given the limited data that were obtained on the first day, coupled with the fact that his parents still wanted information regarding his skills and, more specifically, whether or not he had a mathematics disorder, it was decided that testing be reattempted about a month later. At that time, a briefer intelligence screen was administered, as were mathematics tasks and more tests of validity. Unfortunately, as before, performances across multiple freestanding (TOMM, the Rey 15-Item Test, Dot Counting) and embedded effort indices (California Verbal Learning Test for Children, Recognition) indicated that he was not participating in the testing in a straightforward manner. It is worth noting, however, that his approach on the second testing day was slightly more sophisticated; for example, though his TOMM performance was still a solid fail, it was at least above chance level.

In interpreting testing results, we noted that Justin was clearly not participating in the neuropsychological evaluation in a valid manner. As such, it was difficult to draw firm conclusions regarding his cognitive strengths and weaknesses. What could be said (based on test results) was that his verbal intelligence was *at least* average, though likely higher. By history and a thorough review of records, as well as standardized behavioral ratings completed by parents and teachers, he continued to meet diagnostic criteria for ADHD-I. With respect to the question of a mathematics disorder, it was noted that he did, indeed, show weaknesses on mathematics tasks, but results were questionable given his poor engagement in the testing. However, given reported math difficulties seen across his educational records, this remained a viable possibility. Psychological functioning was clearly the most salient feature of his presentation. Clinical interview and symptom report forms demonstrated a long history of depressed mood, anxiety, emotional lability, and some self-harm (i.e., superficial cutting). After further discussion with Justin, his parents, and his therapist, we recommended a different therapeutic approach. Specifically, we recommended dialectical behavioral therapy (DBT) in addition to increased tutoring and support for his attentional issues and apparent math difficulties.

Following this assessment, Justin entered a DBT-driven individual treatment program that focused on emotional regulation, limit setting, and acceptance of responsibility. He did very well in this setting and made excellent progress. After a year of treatment, Justin underwent neuropsychological reevaluation. At that time, Justin performed well across all freestanding and embedded validity indices and was deemed to be putting forth appropriate effort. Interestingly, the overall pattern of results was similar to what was seen in the first assessment (e.g., verbal intellectual skills were significantly stronger than perceptual reasoning skills), though all scores were substantially higher overall. Notably, Justin continued to struggle with mathematics, and the suspected mathematics disorder diagnosis was confirmed. He continued to show poor attention and executive functions, consistent with his long-standing and well-documented ADHD-I diagnosis.

This case exemplifies several ethical issues. As discussed previously, one of the overriding ethical principles is that of beneficence and nonmaleficence (Principle A), a principle that was clearly upheld by the inclusion of PVTs as part of the assessment. Had the initial results of Justin's evaluation been considered valid without examining effort and engagement, erroneous decisions regarding his overall level of functioning would have been made. This, in turn, would have led to inappropriate decisions regarding educational programming. Based on his projected results, placement in an overly restrictive setting, one geared toward students with far lower intellectual abilities than Justin possesses, would have been considered. Such a placement would potentially have resulted in negative consequences to Justin's self-concept and self-esteem, as well as to his social and academic growth.

Another important consideration is that, without inclusion of validity testing, the severity of his psychopathology may have been underestimated. For Justin, his psychopathology was clearly manifesting as manipulative behaviors. In fact, conversations with Justin's mother and therapist after the initial consultation reflected limited knowledge of Justin's capability to be misleading. As such, the data from validity tests provided concrete evidence of his manipulative tendencies, with his below-chance performance on the TOMM considered a "smoking gun of intent" (Pankratz & Erickson, 1990, p. 385). Accordingly, it is reasonable to assume that optimal treatment recommendations, which included a very specific therapeutic approach for individuals with these personality characteristics, could not have been made without validity testing. In short, the optimal outcome that was achieved for Justin was a direct result of these data.

APA ethics code principles B (fidelity and responsibility) and C (integrity) are also relevant in this case. When working with children like Justin, one of the primary goals of assessment is to accurately describe their functioning with valid data (consistent with standards 9.01 and 9.02) in order to make the most appropriate recommendations for academic placement, interventions, and accommodations. Information reported is utilized to advocate for services with the school system. However, in Justin's case, had validity testing not been included as part of this assessment, the information provided to the school personnel (which would have suggested a low level of intelligence) would be quite discrepant with their experience of Justin as a student maintaining respectable grades, despite his apparent disabilities. In this case, the inclusion of validity testing not only protected the relationship of trust between the neuropsychologist and the school system but also protected the integrity of the field by not presenting unsubstantiated data as truth (in keeping with general principles B and C).

APA ethical standard 9.06 highlights that, when interpreting assessment results, psychologists must take into account all factors, including the purpose of the assessment, person characteristics, and situational factors that may affect their interpretation. In the case of Justin, exaggerated deficits were related to a situational factor (i.e., emotional distress). In therapy, it came to light that he exaggerated his deficits on testing for fear that performing at a level consistent with his true capabilities would have led to a change in his school setting. Ironically, the opposite was actually true; had his results been accepted at face value, it is possible that we would have advocated for a different school environment. In sum, this case exemplifies the fact that the proper use of validity testing with school-age children is in line with relevant ethical principles and standards. Moreover, in certain cases, the incorporation of validity testing will actually lead to better clinical outcomes. In this case, we were able to make targeted therapy recommendations (i.e., a DBT program) that led to an optimal outcome. At the time of this writing, Justin was doing well in high school (with appropriate academic supports) and had many friends and a stronger relationship with his family. He continued to see his DBT therapist for occasional "check-ins" but no longer required weekly support.

CONCLUSIONS

Despite the fact that pediatric practitioners have been somewhat slow to adopt validity testing as part of their routine assessments, particularly outside of the forensic arena, it now seems prudent that we do so. The last decade has seen a small explosion of research demonstrating that symptom exaggeration and noncredible effort are not uncommon in children and teens, and such effort is not always detectable by behavioral observation.

Moreover, PVTs have shown utility in detecting these response biases in this young group.

The use of formal validity testing as part of the routine assessment of children and adolescents should no longer be considered optional, as it is in alignment with the professional guidelines of the field (e.g., NAN, AACN) and consistent with the ethical guidelines for psychologists (APA, 2002). Integration of performance validity data into neuropsychological practice reflects the current state of the field.

As we move forward, providing training regarding issues of effort, motivation, and symptom exaggeration in pediatric assessments should become part of a common educational curriculum. Moreover, it is strongly recommended that future tests be designed to include embedded validity measures whenever feasible, as this will serve to shorten overall testing time (as it will obviate the need for freestanding validity tests in some cases) and will allow an ongoing assessment of task engagement rather than just validity testing at a few points in time. The routine use of validity testing in pediatric evaluations will ensure the accuracy of data collected, thereby ensuring that appropriate diagnostic conclusions can be drawn. Only after the determination that data are accurate and valid can appropriate treatment recommendations be made. Validity testing in children and adolescents undergoing neuropsychological or psychoeducational evaluation allows the intelligent utilization of our health care resources.

REFERENCES

American Psychological Association. (2002). Ethical principles of psychologists and code of conduct. *American Psychologist, 57*(12), 1060–1073.

Blaskewitz, N., Merten, T., & Kathmann, N. (2008). Performance of children on symptom validity tests: TOMM, MSVT, and FIT. *Archives of Clinical Neuropsychology, 23*(4), 379–391.

Bush, S. S., Ruff, R. M., Troster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., et al. (2005). Symptom validity assessment: Practice issues and medical necessity: NAN Policy and Planning Committee. *Archives of Clinical Neuropsychology, 20*(4), 419–426.

Carone, D. A. (2008). Children with moderate/severe brain damage/dysfunction outperform adults with mild-to-no brain damage on the medical symptom validity test. *Brain Injury, 22*(12), 960–971.

Carone, D. A. (2014). Young child with severe brain volume loss easily passes the Word Memory Test and Medical Symptom Validity Test: Implications for mild TBI. *The Clinical Neuropsychologist, 28*(1), 146–162.

Constantinou, M., & McCaffrey, R. J. (2003). Using the TOMM for evaluating children's effort to perform optimally on neuropsychological measures. *Child Neuropsychology, 9*(2), 81–90.

Courtney, J. C., Dinkins, J. P., Allen, L. M., III, & Kuroski, K. (2003). Age-related effects in children taking the Computerized Assessment of Response Bias and Word Memory Test. *Child Neuropsychology, 9*(2), 109–116.

Daly, M. W., Custer, G., & McLeay, P. D. (2008). Cardiac arrest with pulseless electrical activity associated with methylphenidate in an adolescent with a normal baseline echocardiogram. *Pharmacotherapy, 28*(11), 1408–1412.

Donders, J. (2005). Performance on the Test of Memory Malingering in a mixed pediatric sample. *Child Neuropsychology, 11*(2), 221–227.

Green, P., & Flaro, L. (2003). Word Memory Test performance in children. *Child Neuropsychology, 9*(3), 189–207.

Hartung, C. M., Canu, W. H., Cleveland, C. S., Lefler, E. K., Mignogna, M. J., Fedele, D. A., et al. (2013). Stimulant medication use in college students: Comparison of appropriate users, misusers, and nonusers. *Psychology of Addictive Behaviors, 27*(3), 832–840.

Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., Millis, S. R., & Conference Participants. (2009). American Academy of Clinical Neuropsychology Consensus Conference Statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist, 23*(7), 1093–1129.

Kirk, J. W., Harris, B., Hutaff-Lee, C. F., Koelemay, S. W., Dinkins, J. P., & Kirkwood, M. W. (2011). Performance on the Test of Memory Malingering (TOMM) among a large clinic-referred pediatric sample. *Child Neuropsychology, 17*(3), 242–254.

Kirkwood, M. W., Hargrave, D. D., & Kirk, J. W. (2011). The value of the WISC-IV Digit Span subtest in detecting noncredible performance during pediatric neuropsychological examinations. *Archives of Clinical Neuropsychology, 26*(5), 377–384.

Kirkwood, M. W., & Kirk, J. W. (2010). The base rate of suboptimal effort in a pediatric mild TBI sample: Performance on the Medical Symptom Validity Test. *The Clinical Neuropsychologist, 24*(5), 860–872.

Kirkwood, M. W., Peterson, R. L., Connery, A. K., Baker, D. A., & Grubenhoff, J. A. (2014). Postconcussive symptom exaggeration after pediatric mild traumatic brain injury. *Pediatrics, 133*(4), 643–650.

Loughan, A. R., & Perna, R. (2014). Performance and specificity rates in the Test of Memory Malingering: An investigation into pediatric clinical populations. *Applied Neuropsychology: Child, 3*(1), 26–30.

MacAllister, W. S., Nakhutina, L., Bender, H. A., Karantzoulis, S., & Carlson, C. (2009). Assessing effort during neuropsychological evaluation with the TOMM in children and adolescents with epilepsy. *Child Neuropsychology, 15*(6), 521–531.

Marshall, P., Schroeder, R., O'Brien, J., Fischer, R., Ries, A., Blesi, B., et al. (2010). Effectiveness of symptom validity measures in identifying cognitive and behavioral symptom exaggeration in adult attention-deficit/hyperactivity disorder. *The Clinical Neuropsychologist, 24*(7), 1204–1237.

Nymark, T. B., Hovland, A., Bjornstad, H., & Nielsen, E. W. (2008). A young man with acute dilated cardiomyopathy associated with methylphenidate. *Vascular Health and Risk Management, 4*(2), 477–479.

Pankratz, L., & Erickson, R. C. (1990). Two views of malingering. *The Clinical Neuropsychologist, 4*(4), 379–389.

Perna, R., & Loughan, A. R. (2014). The influence of effort on neuropsychological performance in children: Is performance on the TOMM indicative of neuropsychological ability? *Applied Neuropsychology: Child, 3*(1), 31–37.

Postal, K., & Armstrong, K. (2013). *Feedback that sticks: The art of effectively communicating neuropsychological assessment results*. New York: Oxford University Press.

Sherman, E. M., Wiebe, S., Fay-McClymont, T. B., Tellez-Zenteno, J., Metcalfe, A., Hernandez-Ronquillo, L., et al. (2011). Neuropsychological outcomes after epilepsy surgery: Systematic review and pooled estimates. *Epilepsia, 52*(5), 857–869.

Teter, C. J., McCabe, S. E., Cranford, J. A., Boyd, C. J., & Guthrie, S. K. (2005).

Prevalence and motives for illicit use of prescription stimulants in an undergraduate student sample. *Journal of American College Health, 53*(6), 253–262.

Vekaria, P., Vasserman, M., & MacAllister, W. (2013). The utility of the TOMM and Reliable Digit Span as performance validity measures in children and adolescents [Abstract]. *Archives of Clinical Neuropsychology, 28*(6), 600.

Welsh, A. J., Bender, H. A., Whitman, L. A., Vasserman, M., & MacAllister, W. S. (2012). Clinical utility of Reliable Digit Span in assessing effort in children and adolescents with epilepsy. *Archives of Clinical Neuropsychology, 27*(7), 735–741.

Part III

Validity Testing across Evaluative Settings

10

Child and Adolescent Psychoeducational Evaluations

ALLYSON G. HARRISON

Children and adolescents with bona fide disabilities are entitled to a variety of academic and other supports to assist them as they progress through the educational system. Sometimes, however, children, even young children, may not invest their full effort during psychoeducational evaluations, producing invalid scores that interfere with accurate data interpretation and diagnosis. During psychoeducational assessments, examiners have typically relied on subjective judgments rather than objective validity tests to determine whether data are valid or not. Because of the potential academic and/or economic benefits provided to children (and/or families) who are given a disability diagnosis such as attention-deficit/hyperactivity disorder (ADHD) or a learning disability (LD), concerns have been raised regarding the ease with which unimpaired children or adolescents could feign such disorders in order to receive test accommodations, stimulant medication, or disability benefits. Even when a child is not purposefully feigning, a lack of engagement in the testing process can also negatively affect the validity of obtained test scores.

Much evidence has been presented recently regarding the need for performance validity tests (PVTs) and symptom validity tests (SVTs) in the assessment of college-age students undergoing psychoeducational assessments, yet no research has investigated the extent to which these issues require attention in a pediatric population. This chapter, therefore, reviews the nature and the typical contexts in which psychoeducational assessments occur; reasons that children and adolescents might produce invalid test scores; the existing literature regarding symptom exaggeration, feigning, and test-taking motivation in this context; methods to detect noncredible performance or exaggerated symptom reporting; and suggestions for further research.

THE NATURE OF PSYCHOEDUCATIONAL EVALUATIONS AND TYPICAL CONTEXTS IN WHICH THESE OCCUR

Psychoeducational assessments are typically undertaken because a child or adolescent is experiencing difficulty participating equally in an educational environment despite instructional changes (Edmunds & Edmunds, 2008). Whereas physical or sensory disabilities that interfere with equal participation in an educational setting are diagnosed easily, neurodevelopmental disabilities such as LD or ADHD are more difficult to diagnose, as no lab test or genetic marker can verify whether a child truly has that particular disability. As such, these diagnoses often involve a certain amount of clinical judgment and rely on either patterns of performance on tests or documenting the presence of a specific number of reported symptoms.

According to the Centers for Disease Control and Prevention (2014), the most commonly diagnosed developmental disabilities in children in the United States are LD (7.66%) and ADHD (6.69%). In fact, children with LD represent 42% of all students with disabilities identified in the K–12 school system (National Center for Learning Disabilities, 2014). A pan-Canadian survey in 2006 estimated a prevalence rate of 3.2% in children ages 5–14 years, and 69.3% of all children with disabilities were said to have LD (Statistics Canada, 2007). It therefore seems reasonable to conclude that children with LD and ADHD account for more than half of all students with disabilities in the K–12 school system.

Psychoeducational assessments are typically undertaken by school-based personnel (e.g., the school psychologist, occupational therapist, speech and language pathologist) or privately by a licensed professional (e.g., school/clinical psychologist or medical specialist) if the parents choose to obtain an outside opinion. If a specific disability is diagnosed by a qualified professional and it has sufficient educational impact, the school is legally required to provide appropriate accommodations and supports (e.g., Individuals with Disabilities Education Act [IDEA], 2004; Section 504 of the Rehabilitation Act of 1973 [Public Law 93-112, 87 Stat. 394; September 26, 1973]) so children with disabilities may participate equally within the educational system (Gregg, 2009). Parents of children with education-related disabilities may also be eligible for additional government funding and tax benefits, reflecting the fact that such children often require expensive therapies, tutoring, equipment, and/or personal support workers. These are all positive consequences of the previously noted special education laws.

Unfortunately, children and adolescents who do not meet diagnostic criteria for a disability may, due to either exaggerated symptoms or low test-taking motivation, obtain a disability diagnosis and qualify for accommodations and supports or disability benefits or may gain access to potentially addictive medications. Because the use of PVTs and SVTs is "virtually nonexistent among school psychologists" (DeRight & Carone, 2013, p. 3), no data exist regarding the percentage of children who produce noncredible data during such assessments. This is unfortunate, as we know that a large percentage of normal children engage in some type of lying or deception on a regular basis (e.g., see Peterson & Peterson, Chapter 3, this volume).

The limited research examining rates of feigned or exaggerated educational disabilities has focused almost exclusively on college-age students but demonstrates clearly that disabilities such as ADHD and LD can be feigned or exaggerated easily (e.g., Harrison, Edwards, & Parker, 2007, 2008; Harrison, Edwards, Armstrong, & Parker, 2010; Jachimowicz & Geiselman, 2004; Quinn, 2003; Suhr, Hammers, Dobbins-Buckland, Zimak, & Hughes, 2008; Sullivan, May, & Galbally, 2007). Such exaggeration is also apt to go undetected by assessing professionals who rely exclusively on clinical judgment to detect noncredible responding (Faust, Hart, & Guilmette, 1988; Faust, Hart, Guilmette, & Arkes, 1988). The base rate for suspected symptom exaggeration or feigning of ADHD in college students ranges from 15 to 47% (Harrison & Edwards, 2010; Suhr et al., 2008; Sullivan et al., 2007). Rates for feigned LD are estimated to be around 15% (Harrison, Edwards, & Parker, 2008; Sullivan et al., 2007). Although rates of such behavior in pediatric populations are unknown, multiple case reports of noncredible performance or malingering of ADHD, LD, or intellectual disability exist in pediatric neuropsychological evaluations (e.g., Cassar, Hales, Longhurst, & Weiss, 1996; Harrison, Green & Flaro, 2012; Kirkwood, Kirk, Blaha, & Wilson, 2010; Lu & Boone, 2002).

Practical Aspects of Psychoeducational Testing

Most school psychologists have limited testing time, and, in many cases, a portion of the educational tests are administered at a different time by other educational professionals (National Association of School Psychologists, 2014). Therefore low effort or noncredible performance in one part of the assessment process may not be identified unless PVTs or SVTs are given by each examiner. School teams initiate assessments for academic underperforming, so exaggeration or feigning of symptoms is usually not one of the differential diagnostic options considered in school-based assessments. However, not all students identified as having a disability receive evaluations from school-based personnel. Indeed, due to a backlog in service provision, lack of personnel, or even disagreement about whether the child requires an assessment at all, a sizable percentage of students receiving school-based accommodations for LD and ADHD are evaluated by private practitioners. Unlike school-based evaluations, those conducted by private diagnosticians are typically paid for by the parents of the child being assessed. Private practitioners derive their livelihood from referrals and word of mouth, so they may be reluctant to include any assessment measure in their test battery that might indicate malingering or deception (Harrison, Lovett & Gordon, 2013).

REASONS FOR NONCREDIBLE PERFORMANCE ON PSYCHOEDUCATIONAL TESTS

Contrary to popular belief, many primary and secondary gains may be obtained if a child or adolescent is diagnosed as having a disability that interferes substantially with a major life activity, such as learning or attention. A number of positive benefits that can result from a disability diagnosis in the school system are outlined in the following sections.

Extra Time

Students with impaired reading, processing, or test-taking speed require extra time in order to participate equally on timed tests (Cahalan, Mandinach, & Camara, 2002; Cahalan-Laitusis, King, Cline, & Bridgeman, 2006). The rationale for providing extra time has been that students with certain cognitive disabilities such as LD or ADHD benefit from extended test-taking time, whereas students without disabilities do not (e.g., Sireci, Scarpati, & Li, 2005); however, no precise, research-derived method exists to determine exactly how much extra time any given child with a disability should be awarded (Ofiesh, Hughes, & Scott, 2004) so that accommodations do not provide an unfair advantage. Contrary to clinical lore, it is also true that additional time for writing tests helps all students, not just those with disabilities (Sireci et al., 2005). Recent studies (e.g., Lewandowski, Cohen, & Lovett, 2013; Miller, Lewandowski, & Antshel, 2013) show that under severe time pressure students without disabilities benefit more from extra test-taking time than do students with either reading disabilities (RD) or ADHD. Indeed, Lewandowski et al. (2013) conclude that providing disabled students with any more than 25% extra time actually gives an advantage relative to those taking the test under regular time conditions. If all students benefit from extra time but only some receive this accommodation, then extra time becomes an unfair advantage.

More students appear to be obtaining disability diagnoses in order to take advantage of the benefits offered by extra test-taking time. In 2000, the office of the California State Auditor released the results of an extensive statewide audit intended to determine whether a significant number of students were unfairly receiving extra time to complete the SAT due to questionable learning disability diagnoses (California State Auditor, 2000). Rates of disability-related accommodation on the annual statewide achievement test (Standardized Testing and Reporting program, or STAR) were compared with rates of accommodation on the SAT. Students with bona fide disabilities require accommodations for all examinations, regardless of test importance; however, rates of accommodation were substantially higher for students taking the SAT, on which a high score increases a student's chance for college admission, than for students taking the STAR, for which no rewards or advantages accrue if a student obtains a high score. The California State Auditor Board (2000) reported that the basis for extra time accommodations on the SAT was questionable for 18.2% of cases. Further, they noted that accommodations were provided disproportionately to white or affluent families or to those who attended private schools, whereas the number of accommodations provided in inner-city Los Angeles schools was zero. Complaints of this sort are not new, as Lester and Kelman (1997) found that the rates of LD diagnoses in schools across the United States were correlated significantly with socioeconomic status, whereas prevalence of other physical and medical disabilities had no meaningful correlation with parental education or income.

Anecdotal evidence from the popular press also implies that savvy parents and their offspring are gaming the system to ensure that their children are diagnosed as disabled in an attempt to improve performance on high-stakes tests such as the SAT or the American College Test (ACT). Indeed, recent reports (e.g. Mitchell, 2012; Tapper, Morris, & Setrakian, 2006) highlight concerns about the excessively high percentage of students from affluent American neighborhoods whose children have been diagnosed as disabled and granted extra time for taking college entrance tests. Dubbed "the rich kid's loophole" (Tapper et al., 2006), multiple reports quote tutors, guidance counselors, and school principals who accuse affluent parents of encouraging their children to fail diagnostic tests in order to obtain a disability label and qualify for test accommodations such as extra time (Mitchell, 2012; Tapper et al., 2006).

Parents may in fact be motivated to seek a diagnosis for their children well in advance of the day they must take such college admissions tests. The Americans with Disabilities Act Amendments of 2008 (ADAAA) and the recent regulations provided by the Department of Justice regarding interpretation of this law as it currently applies to test accommodations (ADA National Network, 2014) strongly suggest that organizations that administer high-stakes admissions examinations apportion "considerable weight" to a documented history of accommodations and modifications when determining whether to grant accommodations on such admission tests (ADA National Network, 2014). Hence, although the new ADAAA regulations uphold the previous benchmark for determining a disability, these new guidelines also seem to imply that a history of accommodation trumps questionable diagnostic documentation. As such, parents may be motivated to establish an academic accommodation history for their child in order to ensure later accommodations on the SAT or ACT (see Moore, 2010).

Problems of this sort are not unique to North America. Indeed, a recent report from Greece (GR Reporter, 2012) notes that hundreds of fake dyslexia certificates were sold to parents to allow their children special accommodations when taking university entrance exams, apparently a known route to taking an easier untimed oral entrance test. Additionally, students are searching the Internet for ways to fake an LD. Figure 10.1 provides one example of a student trying to find out how to beat the system, but many more may be found online. Clearly, certain adolescents and their parents view obtaining extra time as a decided competitive advantage when vying for acceptance into postsecondary institutions.

Computers, Tuition Rebates, and Disability Tax Credits or Other Benefits

At all levels of education, students with specific disabilities may be entitled to use technology (e.g., laptop or tablet, text-to-speech software) both in class and when taking tests. In an age in which the majority of students frequently use computer technology for communication, one can imagine that access to a computer to take tests and exams would be desirable, especially when your peers must hand write and cannot utilize options such as spell check, grammar check, or cut and paste. The desire to obtain a computer for both home and school may provide students (or their parents as proxy agents) with a motivation to underperform on educational tests of handwriting quality, spelling, or written expression.

Canadian postsecondary students diagnosed with a disability such as LD or ADHD can receive provincial and federal grants for purchasing computers and assistive technology software (Harrison, 2004, 2006; Harrison, Edwards, & Parker, 2007). Parents of Canadian students with diagnosed disabilities may also be eligible for a disability tax credit (Canada Revenue Agency, 2014), and students with disabilities may qualify to have their student loans forgiven (Harrison, 2006). Additionally, Canadian students diagnosed with a permanent disability are eligible for a substantial postsecondary tuition rebate (up to $2,000 per year) (Canada Access Study Grant for Students with Permanent Disabilities, 2014), and up to $10,000 per year in bursary funds (Canada Study Grant for the Accommodation of Students with Permanent Disabilities, 2014). As such, there appear to be many secondary gain incentives for students (or their parents) to feign a disability in an educational context.

Although the United States does not offer a comparable tax credit,

I want to have Dyslexia please?

i have my final year exams this year and i know a guy who has dyslexia and because of this he qualifies for extra marks in his tests.

so i thought that if i could claim dyslexia then i could also get these extra marks to help me on my way to college. but i found out i need to have a doctors note to qualify for these marks.

so i was wondering how i would go about faking a dyslexia test which would get me my doctors note. if the test is just writing a paragraph then all i have to do is mix up a couple of letters here and there but my worry is if there is a more complex test (brain scans) which i may not fail and in turn make me look very foolish.

so firstly what im wondering is, what test will i have to complete to be diagnosed as a dyslexic. and secondly, how will i cheat on this test to allow myself to look dyslexic.

i am aware this is mean and dishonest but if i felt guilty i wouldn't be writing this :D

looking forward to hearing from ye.

Update : ok i am studying quite hard for my tests, but im also looking for extra ways to improve my overall score.

i would look at it as beating the system as opposed to cheating.

FIGURE 10.1. Actual excerpt from an online discussion board.

parents of children with disabilities may opt to itemize deductions on their federal income tax return to recoup expenses related to their children's disability, such as private school tuition, tutoring, specialized materials (e.g., books, software, and instructional materials), diagnostic evaluations (by a private practitioner), therapy, and education-related transportation expenses (U.S. Internal Revenue Service, 2013).

Disability Benefits

Not all cases of pediatric symptom exaggeration or feigning are instigated by the children themselves. In many cases there are reasons to suspect malingering by proxy, a term used when a vulnerable individual acts under the guidance, direction, or control of others to produce false symptoms or underperform on a test (Slick & Sherman, 2013). Cassar et al. (1996) were the first to describe such a case involving a mother who insisted on having her 13-year-old daughter diagnosed with and treated for both ADHD and a bipolar disorder. These authors submit that the mother fabricated her daughter's symptoms and suggest that the mother's motivation was to secure disability benefits for a child with chronic disabilities. Given that many educational disabilities such as ADHD are diagnosed in large part on the basis of parental report, these authors warn how easily a parent could malinger by proxy and obtain disability or other benefits. Similar claims were made in a BBC news article (Goldberg, 2011) with respect to exaggerated ADHD symptoms by parents in the United Kingdom who wanted to claim a disability living allowance and in the United States by Chafetz and his colleagues (e.g. Chafetz, Abrahams, & Kohlmaier, 2007; Chafetz & Prentkowski, 2011) regarding parents encouraging their children to feign in order to obtain Social Security Disability benefits (see Chafetz, Chapter 14, this volume).

Clearly, there is reason to believe that parents may encourage their children to feign symptoms during disability evaluations or even exaggerate symptoms regarding their children when filling out symptom-report inventories. Further, research suggests that, with parental encouragement, children are more likely to lie when they believe that they won't be blamed for the transgression (Talwar, Lee, Bala, & Lindsay, 2004). Unfortunately, no research has been undertaken in school or educational settings to document how frequently malingering by proxy may occur.

Stimulants

Although they are useful in the treatment of ADHD, stimulant medications such as Ritalin (methylphenidate), Dexadrine (dextroamphetamine sulfate), and Adderall (a mixture of amphetamine salts) have the potential to benefit students regardless of the presence or absence of attentional dysfunction, as they assist in improving alertness and attention. Stimulants may be used by individuals without disabilities as a study aid, presumably improving attention and alertness on demand so that students can focus when they want and study more effectively (White, Becker-Blease, & Grace-Bishop, 2006).

Studies of college-age students demonstrate the high rate of stimulant abuse (e.g., Advokat, Guidry, & Martino, 2008; McCabe, Teter, & Boyd, 2006), but less research has investigated the rates of abuse in pediatric populations. Two recent surveys (one in Canada and the other in the United States) report that 20% of teens admit to abusing or misusing prescription drugs that were not prescribed to them directly (Ontario Student Drug Use and Health Survey [OSDUHS], 2013; Partnership for Drug-Free Kids (2012) Partnership Attitude Tracking Study [PATS]), with misuse and abuse of stimulant medications showing the largest increase since 2008. Indeed, 13% of individuals surveyed in the 2012 PATS admitted to misusing or abusing stimulant medications. Additionally, 26% of respondents believed that drugs can be used as study aids, and 29% of parents believed that ADHD medications could improve their child's academic or test performance even if the child did not have ADHD. Finally, more than one-quarter (27%) of teens mistakenly believed that misusing or abusing prescription drugs to get high was safer than using street drugs.

Setlik, Bond, and Ho (2009) also reported disturbing data regarding calls to poison control centers across the United States between 1998 and 2005. Whereas the rates of reporting for other substances had remained fairly constant or decreased over this 8-year span, the rates of calls related to abuse or misuse of ADHD medication in preteens and teenagers (ages 13–19) rose 76%. In particular, over this 8-year period, the study found that calls related to amphetamine/dextroamphetamine-related abuse (Adderall and Dexadrine) increased by 476%, whereas rates of prescription for these medications had increased by only 141% over the same time period for children ages 10 to 19 years. The authors concluded that the dramatic increase in calls regarding abuse or misuse of ADHD medications is out of proportion to rates of misuse of any other illicit drug, suggesting a rising problem with teenagers misusing stimulants.

Multiple websites such as *Adderall tips: How to convince your shrink you have ADD/ADHD* provide enterprising students with helpful step-by-step instructions regarding how to feign symptoms of ADHD in order to obtain stimulant medication. These sites typically describe the ease with which one can obtain stimulant medications from physicians and offer plausible ways to counter any skepticism shown by the physician in order to ensure that the medications will be prescribed. Further, they also coach students regarding what to say upon return visits in order to obtain an increase in the prescription dosage or even a change from a stimulant with less abuse potential to one with a greater effect. Many other online chatrooms (see Figure 10.2 for one example) offer suggestions regarding where and how to access medications for illicit use.

What about the Role of Effort and Motivation?

Even when a student does not purposely set out to feign or exaggerate symptoms, there are still many other reasons that the scores they produce in a psychoeducational assessment may not be accurate. For instance, although not due to deliberate feigning or premeditated deception, low motivation or unwillingness to put forth good effort on tasks in an assessment that are frustrating or unappealing, may also invalidate test results and reduce diagnostic accuracy.

Howard Adelman and his colleagues first identified the influence of low effort on obtained test scores and correct diagnosis of LD more than 20 years ago (Adelman, Lauber, Nelson, & Smith, 1989). They found that almost 80% of children (ages 12–17 years) diagnosed as having an RD were actually capable of performing reading-related tasks if properly motivated

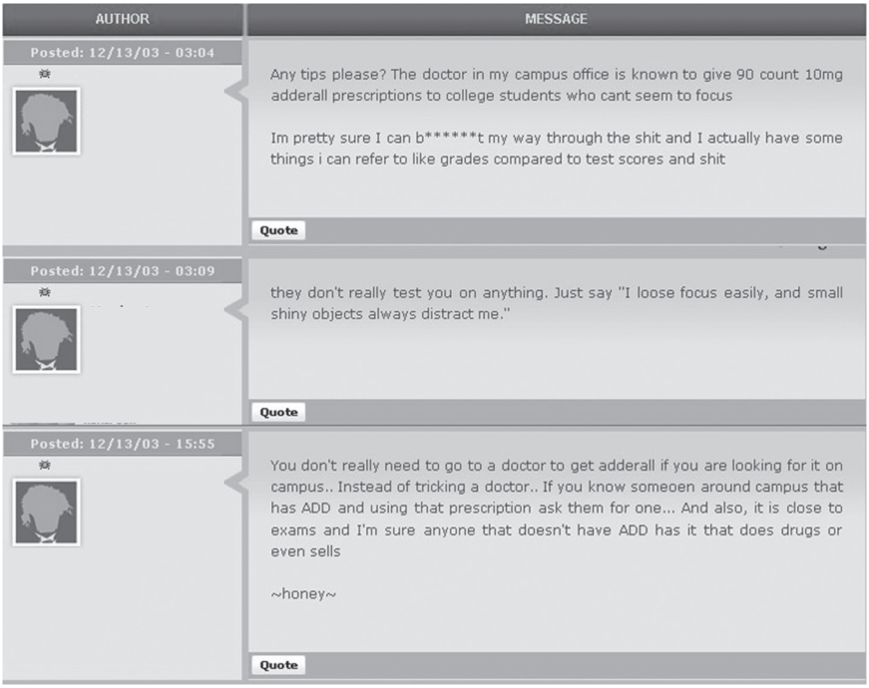

FIGURE 10.2. Online chatroom quotes regarding ways to obtain stimulant medication.

or engaged, noting that the low scores they had previously produced on reading and decoding tests were due more to lack of engagement or avoidance than to an actual inability to complete the tasks. These authors cautioned clinicians not to make disability diagnoses based on data gathered under low or avoidance motivation conditions (e.g., when the client is not engaged in testing and is investing low effort) and suggested that test developers produce assessment instruments that are intrinsically appealing and of interest to younger children.

Children with ADHD often perform poorly on a variety of academic and psychological tasks and demonstrate significant motivation difficulties, even though there is evidence of adequate cognitive capacity (e.g., Flaro & Green, 2000). In the United States, upward of 40% of high school students are chronically disengaged from learning and therefore invest little effort on schoolwork or school-based tests (National Research Council, 2004). Unmotivated students often perform poorly on tests of academic functioning (Usher & Kober, 2012). Additionally, Diller (2006) describes what he terms “the valley of motivational fatigue” (p. 78), a period in adolescence (ages 13–17) during which young males in particular have difficulty being motivated to perform well on school-related tasks. He notes that lack of motivation at this age, coupled with decreased persistence and lower academic performance, may lead assessors to mistakenly believe that the cause is ADHD.

Many children undergoing pediatric brain injury assessments fail effort tests even when no apparent primary or secondary gains are identified (e.g., Kirkwood & Kirk, 2010). In such cases, children may not invest full effort for many reasons. Children may have come to believe that they cannot do certain tasks simply as a function of iatrogenic illness. Alternatively, children may have oppositional-defiant traits and choose not to comply with the evaluation (e.g., Green & Flaro, 2003; Harrison, Green, & Flaro, 2012). Other causes for low test-taking motivation include boredom, uninterest in testing, school avoidance, or ensuring continued attention from parents or caregivers (Constantinou & McCaffrey, 2003; Donders, 2005; Mantynen, Poikkeus, Ahonen, Aro, & Korkman, 2001). Although these are not cases of deliberate malingering, they nevertheless influence a clinician's ability to obtain accurate information about the true capabilities and deficits of a child.

THE EASE OF FEIGNING EDUCATIONAL DISABILITIES

As noted previously, almost no research has examined symptom exaggeration within a pediatric educational context, so we must turn to research at the postsecondary level that suggests that it is fairly easy for young adults to feign symptoms of most behaviorally defined disabilities. Indeed, in a recent survey, Canadian postsecondary disability services office staff agreed that LD and ADHD were the disabilities most easily feigned in an educational context. Hearing and visual impairments were judged to be the most difficult to feign (Harrison & Wolforth, 2012).

A clinical diagnosis of ADHD is based mainly on interview data and symptom report checklists (Jachimowicz & Geiselman, 2004), and such methods are extremely susceptible to feigned impairment. Indeed, multiple studies (e.g., Booksh, Pella, Singh & Gouvier, 2010; Harrison, Edwards, & Parker, 2007, 2008; Jachimowicz & Geiselman, 2004; Quinn, 2003; Sollman, Ranseen, & Berry, 2010) have shown how easily both naïve and coached college-age students can feign believable symptoms on all major ADHD self-report inventories, including exaggeration of historical, as well as current, symptoms.

In clinical samples, Sullivan et al. (2007) found that college students presenting for assessment of ADHD who failed a PVT reported a significantly higher number of ADHD symptoms than did those who were deemed to have invested good test-taking effort. Suhr et al. (2008) found that students presenting for ADHD assessments and suspected of symptom exaggeration could not be discerned from those with true ADHD on the basis of symptom report scores alone. Harrison and Edwards (2010) found that Canadian college students undergoing ADHD evaluations who failed objective effort tests were indistinguishable clinically from those who passed such tests when examining their self-reported ADHD scores on the Conners' Adult ADHD Rating Scales (CAARS; Conners, Erhardt, & Sparrow, 1999).

Additionally, although there is a general assumption that it is more difficult to get collateral sources to lie or exaggerate on symptom questionnaires, this assumption have not been investigated empirically (Jasinski & Ranseen, 2011). No research exists regarding the rate at which parents may lie or exaggerate on observer reports or in interviews. Recent research, however, suggests that some parents do interfere with the assessment and accommodation process. For instance, Harrison and Wolforth (2012) found that the majority of postsecondary disability service providers (62%) believe that parents of students with disabilities interfere in their children's academic environment more than is reasonable. When asked to elaborate, one-quarter of this subset reported that these parents insist on getting the diagnosis they want for their child, 11% that parents go "diagnosis shopping," 19% that parents coach their children to exaggerate their deficits in order to receive accommodations, and 48% that parents insist on the accommodations that they want (rather than what is reasonable based on the documentation). Hence parents may be motivated to exaggerate their children's symptoms, so we cannot assume that observer ratings are less vulnerable to overreporting or feigning than are self-report scales.

By contrast, most studies to date suggest that symptom validity scales on commonly employed personality assessment inventories are insensitive to feigned ADHD/LD and other pediatric conditions (Kirk, Hutaff-Lee, Connery, Baker, & Kirkwood, 2014; Sullivan et al., 2007). It is useful, however, to consider that children and adolescents attempting to feign mental health problems might well overendorse symptoms on such measures, and so self-report inventories that include validity scales are a useful adjunct to any psychoeducational assessment.

In an attempt to obtain more objective evidence of ADHD symptoms, computerized vigilance tests are often used by clinicians to diagnose ADHD, even though their specificity and sensitivity are weak (e.g., Schatz, Ballantyne, & Trauner, 2001). Research shows that college students can easily feign symptoms of ADHD on these tests (e.g., Leark, Dixon, Hoffman & Huynh, 2002; Quinn, 2003). Indeed, Leark et al. (2002) showed that college students with no history of ADHD or psychological problems could produce a profile on a computerized vigilance test, the Test of Variables of Attention (TOVA; Greenberg, 1991; Greenberg, Kindschi, Dupuy, & Corman, 1996), which was interpreted to indicate abnormal attention abilities. They concluded that clinicians should consider response bias when patients obtained excessively elevated scores on the TOVA.

Quinn (2003) found that another commonly used continuous performance test (the Integrated Visual and Auditory Continuous Performance Test [IVA CPT]) was also vulnerable to feigning but suggested that abnormal score patterns might indicate symptom exaggeration. Both Sollman et al. (2010) and Booksh et al. (2010) reported similar findings using the Conners' Continuous Performance Test (CPT; Conners, 1995), suggesting that students could easily feign believable attention problems on the CPT, too. Finally, Suhr, Sullivan, and Rodriguez (2011) found that individuals who failed two or more Word Memory Test (WMT; Green, 2003) indices performed similarly to those with a diagnosis of ADHD and worse on most CPT variables than those endorsing clinical levels of psychological symptoms. Hence, research to date indicates that most commonly used CPTs are easily feigned in clinical settings.

A number of researchers have also shown how easily students feigning ADHD can manipulate performance on various cognitive and achievement tests. For instance, Harrison, Edwards, and Parker (2007) demonstrated that students asked to feign ADHD could produce scores on tests of reading and processing speed on the Woodcock–Johnson III inventory (Woodcock, McGrew, & Mather, 2001) consistent with significant academic impairment. Other authors (e.g., Booksh et al., 2010; Frazier, Frazier, Busch, Kerwood, & Demaree, 2008; Sollman et al., 2010; Suhr et al., 2008) have demonstrated that college students feigning ADHD can easily obtain impaired scores on a variety of commonly used tests of executive functioning, academic achievement, and/or speeded reading comprehension. In all cases, students feigning ADHD were able to return test scores that were equivalent to or more impaired than those produced by students with genuine ADHD, suggesting the ease with which a motivated student could manipulate a psychoeducational assessment and qualify as being disabled.

Although no research has been undertaken to examine the ease with which pediatric samples could feign or exaggerate symptoms of LD, case reports describing such behavior do exist (e.g., Harrison et al., 2012; Lu & Boone, 2002). Further, it has been demonstrated that college students motivated to feign RD can do so easily on tests of phonological awareness, word decoding, reading, and processing speed (Frazier et al., 2008; Harrison, Edwards, & Parker, 2008; Harrison, Edwards, et al., 2010; Lindstrom, Coleman, Thomassin, Southall, & Lindstrom, 2011), rapid naming and phonological processing (Lindstrom et al., 2011), and other tests of academic fluency (Sullivan et al., 2007). Similar to the findings with feigned ADHD, studies all find that students asked to feign RD can achieve test scores equal to or more impaired than those returned by college students with genuine reading impairments. In fact, Lindstrom et al. (2011) concluded that students feigning RD returned test score profiles that were "disturbingly sophisticated" (p. 316), easily meeting commonly used diagnostic criteria such as performing below average on psychoeducational tests. As such, it is likely that students actively attempting to feign RD could produce test scores that would be interpreted as indicating a substantial impairment in academic functioning and qualify for test accommodations or disability benefits.

METHODS TO DETECT NONCREDIBLE PERFORMANCE

Given that pediatric clients may feign many symptoms of both ADHD and LD successfully, the question that remains is how best to identify such behavior when it occurs, so as to accurately interpret test scores obtained in a psychoeducational evaluation. Recent reviews by DeRight and Carone (2013) (as well as Kirkwood, Chapter 5, this volume) report that, in general, school-age children are capable of passing most freestanding PVTs using adult cutoffs.

Although Kirkwood (Chapter 5, this volume) provides general information about PVTs and SVTs in pediatric evaluations, it is important to consider which ones may be appropriate for use with clients with LD and ADHD specifically. For instance, although use of PVTs that rely less heavily on reading ability may have intuitive appeal when conducting assessments with individuals suspected of reading impairments, research indicates that measures such as the Test of Memory Malingering (TOMM; Tombaugh, 1996), although highly specific, may be relatively insensitive to low effort both in young adults undergoing evaluations for LD (e.g., Lindstrom, Lindstrom, Coleman, Nelson, & Gregg, 2009) and in children (Chafetz, 2008; Chafetz et al., 2007; Rambo, Callahan, Hogan, Hullmann, & Wrape, 2014). Similarly, the Rey 15-Item Test (FIT; Rey, 1964) has been shown to have weak sensitivity and specificity in differentiating between malingered and actual neurological impairment (e.g., Arnett, Hemmeke, & Schwartz, 1995; Iverson & Franzen, 1996). Furthermore, there is little reason to think that a student attempting to feign or exaggerate a reading or attention problem would necessarily believe that visual picture memory skills constitute core deficits that need to be demonstrated in order to obtain such a diagnosis.

The WMT and the Medical Symptom Validity Test (MSVT; Green, 2004) have demonstrated better sensitivity and specificity in identification of noncredible performance both with children (e.g., Carone, Green, & Drane, 2014; Green & Flaro, 2003; Kirkwood & Kirk, 2010), and with college-age students feigning LD (Lindstrom et al., 2009). Research suggests that the WMT and MSVT are relatively unaffected by reading impairments in children and adolescents so long as overall word decoding skills are better than the first percentile and word reading skills are at a grade

3 level or higher (Green & Flaro, 2003; Larochette & Harrison, 2012). Research also shows that these two tests return low false-positive rates in assessment of children and adolescents with severe ADHD (Harrison, Flaro & Armstrong, 2014).

DeRight and Carone (2013) conclude that freestanding PVTs are less prone to false-positive classification than are embedded measures. For example, although the Reliable Digit Span (RDS; Greiffenstein, Baker, & Gola, 1994) is insensitive to the cognitive difficulties experienced by adolescents or young adults with RD (Harrison & Armstrong, 2014; Harrison, Rosenblum, & Currie, 2010), there is reason to suspect that this measure produces an unacceptably high false-positive rate in younger children (Blaskewitz, Merten, & Kathmann, 2008). Furthermore, the Digit Span Scaled Score (DSSS) has a high false-positive rate in adolescents with LD, particularly if Canadian normative data are employed in scoring (Harrison & Armstrong, 2014). Other embedded tests that show promise in pediatric evaluations are the A-Test (Chafetz & Abrahams, 2006) and the automatized sequencing task (Kirkwood, Connery, Kirk & Baker, 2013), although more research is required to verify the sensitivity and specificity of these tests.

False-Positive Rates

When employing PVTs in an assessment, it is important to ensure that the false-positive rate is not elevated in persons who have genuine neurological impairment (Merten, Bossink, & Schmand, 2007). Because LD and ADHD can interfere with both working memory and vigilance (e.g., Epstein et al., 2003) and RD interferes with word reading, any PVT used in pediatric populations must not falsely identify students with genuine impairments as investing low effort.

Larochette and Harrison (2012) have shown that the WMT and MSVT have low false-positive rates when evaluating children and adolescents with severe RD or severe ADHD. In fact, the Severe Impairment Profile (SIP) analysis proposed by Green (2008a) allows clinicians to determine whether failure on the WMT or MSVT is likely due to a genuine neurological condition rather than poor effort. Employing this profiling strategy, Larochette and Harrison (2012) found that the WMT/MSVT failure rate of adolescents with genuine RD went from 9.5% to less than 1%. Further, all individuals who failed these PVTs had word decoding skills below the first percentile and a documented history of severe reading impairments. More recently, Harrison et al. (2014) also showed that the rate of failure on the WMT and MSVT in a pediatric ADHD sample was halved when employing the SIP, with 4.7% failing the WMT and 2.7% failing the MSVT. Although these authors caution clinicians not to automatically assume that an SIP is a false-positive diagnosis, a review of the historical data and real-world difficulties demonstrated by participants in their study confirmed a longstanding pattern of severely impaired functioning.

Conversely, Harrison et al., (2014) found that the Computerized Assessment of Response Bias (CARB; Conder, Allen, & Cox, 1992) had an unacceptably high false-positive rate in their pediatric sample (27% of the ADHD sample failed this test). Worth noting, too, was the fact that 6.8% of these participants failed the Nonverbal Medical Symptom Validity Test (NV-MSVT; Green, 2008b) even after allowing for profile analysis. More research is therefore needed to determine whether the NV-MSVT is an appropriate test to use when evaluating effort in pediatric ADHD evaluations.

CONCLUSIONS AND FUTURE DIRECTIONS

It is imperative that psychologists conducting psychoeducational assessments for children and adolescents begin evaluating for symptom and performance credibility. Results of psychoeducational assessments are vulnerable to manipulation by both children and their parents, and inaccurate diagnostic conclusions may be drawn based on such flawed data. Neurodevelopmental disabilities such as LD and ADHD seem particularly vulnerable to feigning or malingering, especially because there is no gold standard to verify these diagnoses. Unfortunately, almost no research exists regarding the extent to which exaggeration or malingering of such disorders may be occurring during pediatric psychoeducational assessments; however, data from college-age students suggest that such disorders may be feigned easily and relatively often. Additional research must therefore be undertaken to evaluate the best methods to identify when children, adolescents, or their parents are misrepresenting symptoms both on performance tests and on self-report questionnaires administered during psychoeducational evaluations. As the majority of existing PVTs appear to be appropriate for use with a pediatric sample, school and clinical psychologists should begin using such measures in their assessments.

Clinicians may be reluctant to include PVTs and SVTs in their test batteries, especially when they derive their livelihood from private referrals and word of mouth. However, Carone, Iverson, and Bush (2010) provide an excellent summary of the ways in which clinicians can encourage compliance during testing and also how to address evidence of noncredible performance when such situations arise (see also Connery & Suchy, Chapter 8, this volume). As shown in this chapter, children and adolescents with bona fide LD and ADHD can certainly pass many existing PVTs, except in rare cases in which they have a well-documented history of profound impairment in multiple domains of functioning.

Research is needed into whether more specialized PVTs and SVTs are required to identify exaggeration of academic or attention difficulties in children. For example, are existing PVTs based on recognition memory more effective than domain-specific PVTs (e.g., Osmon, Plambeck, Klein, & Mano, 2006) in identifying exaggerated reading or attention problems? How easily could other education-related disabilities, such as anxiety, depression, or intellectual disability, be feigned or exaggerated?

As with assessments of neurological complaints in adults (e.g., Green, Rohling, Lees-Haley, & Allen, 2001) and even performance of undergraduates volunteering for experiments in a psychology research pool (e.g., An, Zakzanis, Bors, & Joordens, 2012), many of the cognitive and academic problems previously found in samples of children and adolescents with LD and ADHD may be due, in fact, to low test-taking effort or noncredible performance. Because no data exist regarding the validity of test findings reported in pediatric LD and ADHD samples, it is possible that some of our previously held views regarding core symptoms of these disorders may not be entirely accurate. Until we objectively investigate the credibility of symptoms and performance in pediatric psychoeducational assessments, we will not be able to answer such questions.

REFERENCES

ADA National Network. (2014). Exams and courses. Retrieved February 18, 2014, from *http://adata.org/exams*.

Adelman, H. S., Lauber, B. A., Nelson, P., & Smith, D. C. (1989). Minimizing and detecting false positive diagnoses of learning disabilities. *Journal of Learning Disabilities, 22*, 234–244.

Advokat, C. D., Guidry, D., & Martino, L. (2008). Licit and illicit use of medications for attention-deficit/hyperactivity disorder in undergraduate college students. *Journal of American College Health, 56*, 601–606.

An, K., Zakzanis, K. K., Bors, D., & Joordens, S. (2012). Conducting research with nonclinical healthy undergraduates: Does effort play a role in neuropsychological test performance? *Archives of Clinical Neuropsychology, 27*, 849–857.

Arnett, P. A., Hemmeke, T. A., & Schwartz, L. (1995). Quantitative and qualitative performance on Rey's 15-Item Test in neurological patients and dissimulators. *The Clinical Neuropsychologist, 9*, 17–26.

Blaskewitz, N., Merten, T., & Kathmann, N. (2008). Performance of children on symptom validity tests: TOMM, MSVT, and FIT. *Archives of Clinical Neuropsychology, 23*, 379–391.

Booksh, R. L., Pella, R. D., Singh, A. N., & Gouvier, W. D. (2010). Ability of college students to simulate ADHD on objective measures of attention. *Journal of Attention Disorders, 13*, 325–338.

Cahalan, C., Mandinach, E. B., & Camara, W. J. (2002). *Predictive validity of SAT I: Reasoning Test for test-takers with learning disabilities and extended time accommodations*. New York: College Entrance Examination Board.

Cahalan-Laitusis, C., King, T. C., Cline, F., & Bridgeman, B. (2006). *Observational timing study on the SAT Reasoning Test for test-takers with learning disabilities and/or ADHD* (Research Report 2006-4). New York: College Board.

California State Auditor. (2000). *Standardized tests: Although some students may receive extra time on standardized tests that is not deserved, others may not be getting the assistance they need*. Sacramento, CA: Bureau of State Audits. Retrieved from *www.bsa.ca.gov/bsa*.

Canada Access Study Grant for the Accommodation of Students with Permanent Disabilities. (2014). Available at *www.servicecanada.gc.ca/eng/goc/grant_equipmentdisabilities.shtml*.

Canada Access Study Grant for Students with Permanent Disabilities.(2014). Available at *www.servicecanada.gc.ca/eng/goc/grant_disabilities.shtml*.

Canada Revenue Agency. (2014). Form T2201. Retrieved January 27, 2014.

Carone, D. A., Green, P., & Drane, D. L. (2014). Word Memory Test profiles in two cases with surgical removal of the left anterior hippocampus and parahippocampal gyrus. *Applied Neuropsychology: Adult, 21,* 155–160.

Carone, D., Iverson, G., & Bush, S. (2010). A model to approaching and providing feedback to patients regarding invalid test performance in clinical neuropsychological evaluations. *The Clinical Neuropsychologist, 24*(5), 1–20.

Cassar, J. R., Hales, E. S., Longhurst, J. G., & Weiss, G. S. (1996). Can disability benefits make children sicker? *Journal of the American Academy of Child and Adolescent Psychiatry, 35*(6), 700–701.

Centers for Disease Control and Prevention. (2014). Developmental disabilities increasing in US. Retrieved from *www.cdc.gov/features/dsdev_disabilities*.

Chafetz, M., & Abrahams, J. (2006). "A random letter test predicts effort in the disability exam. *The Clinical Neuropsychologist, 20*(3), 596–596.

Chafetz, M. D. (2008). Malingering on the Social Security Disability consultative exam: Predictors and base rates. *The Clinical Neuropsychologist, 22*, 1–18.

Chafetz, M. D., Abrahams, J. P., & Kohlmaier, J. (2007). Malingering on the Social Security Disability consultative exam: A new rating scale. *Archives of Clinical Neuropsychology, 22*, 1–14.

Chafetz, M. D., & Prentkowski, E. (2011). A case of malingering by proxy in a Social Security Disability psychological consultative examination. *Applied Neuropsychology, 18*, 143–149.

Conder, R., Allen, L, & Cox, D. (1992). *Computerized Assessment of Response Bias test manual*. Durham, NC: Cognisyst.

Conners, C. K. (1995). *Conners' Continuous Performance Test*. Ottawa, Ontario, Canada: Multi-Health Systems.

Conners, C. K., Erhardt, D., & Sparrow, M. A. (1999). *Conners' Adult ADHD Rating Scales (CAARS)*. North Tonawanda, NY: Multi-Health Systems.

Constantinou, M., & McCaffrey, R. J. (2003). Using the TOMM for evaluating children's effort to perform optimally on neuropsychological measures. *Child Neuropsychology, 9*(2), 81–90.

DeRight, J., & Carone, D. A. (2013). Assessment of effort in children: A systematic review. *Child Neuropsychology, 12*, 1–24.

Diller, L. (2006). *The last normal child: Essays on the intersection of kids, culture and psychiatric drugs*. Westport, CT: Praeger.

Donders, J. (2005). Performance on the Test of Memory Malingering in a mixed pediatric sample. *Child Neuropsychology, 11*, 221–227.

Edmunds, A., & Edmunds, G. (2008). *Special education in Canada*. Toronto, Ontario, Canada: McGraw-Hill Ryerson Higher Education.

Epstein, J. N., Erkanli, A., Conners, C. K., Klaric, J., Costello, J. E., & Angold, A. (2003). Relations between Continuous Performance Test performance measures and ADHD behaviors. *Journal of Abnormal Child Psychology, 31*(5), 543–554.

Faust, D., Hart, K., & Guilmette, T. J. (1988). Pediatric malingering: The capacity of children to fake believable deficits on neuropsychological testing. *Journal of Consulting and Clinical Psychology*, *56,* 578–582.

Faust, D., Hart, K., Guilmette, T. J., & Arkes, H. R. (1988). Neuropsychologists' capacity to detect adolescent malingerers. *Professional Psychology: Research and Practice*, *19*, 508–515.

Flaro, L., & Green, P. (2000). Symptom validity test results with children: CARB and WMT. *Archives of Clinical Neuropsychology*, *15,* 840.

Frazier, T. W., Frazier, A. R., Busch, R. M., Kerwood, M. A., & Demaree, H. A. (2008). Detection of simulated ADHD and reading disorder using symptom validity measures. *Archives of Clinical Neuropsychology, 23*(5), 501–509.

Goldberg, A. (2011, February 6). Unscrupulous parents seek ADHD diagnosis for benefits. BBC News Health. Retrieved from *www.bbc.co.uk/news/health-12359070*.

GR Reporter. (2012). Hundreds of students with fake certificates of dyslexia. Retrieved from *www.grreporter.info/en/hundreds_students_fake_certificates_dyslexia/7215*.

Green, P. (2003). *The Word Memory Test*. Edmonton, Alberta, Canada: Green's Publishing.

Green, P. (2004). *Green's Medical Symptom Validity Test (MSVT) for Microsoft Windows: User's manual*. Edmonton, Alberta, Canada: Green's Publishing.

Green, P. (2008a). *Advanced Interpretation Program*. Edmonton, Alberta, Canada: Green's Publishing.

Green, P. (2008b). *Green's Nonverbal Medical Symptom Validity Test (NV-MSVT) for Microsoft Windows: User's manual 1.0*. Edmonton, Alberta, Canada: Green's Publishing.

Green, P., & Flaro, L. (2003). Word Memory Test performance in children. *Child Neuropsychology*, *9*(3), 189–207.

Green P., Rohling, M. L., Lees-Haley, P. R., & Allen, L. M., III. (2001). Effort has a greater effect on test scores than severe brain injury in compensation claimants. *Brain Injury 15,* 1045–1060.

Greenberg, L. M. (1991). *Test of Variables of Attention*. Los Alamitos, CA: Tova.

Greenberg, L., Kindschi, C., Dupuy, T., & Corman, C. (1996). *Test of variables of attention: Clinical guide*. Los Alamitos, CA: Universal Attention Disorders.

Greiffenstein, M. F., Baker, W. J., & Gola, T. (1994). Validation of malingered amnesic measures with a large clinical sample. *Psychological Assessment, 6*(3), 218–224.

Gregg, N. (2009). Accommodations: Evidence-based accommodation research specific to the adolescent and adult population with learning disabilities. In J. Taymens (Ed.), *Learning to achieve: A review of the research literature on serving adults with learning disabilities* (pp. 119–180). Washington, DC: National Institute for Literacy.

Harrison, A. G. (2004). An investigation of reported symptoms of ADHD in a university population. *ADHD Report, 12,* 8–11.

Harrison, A. G. (2006). Adults faking ADHD: You must be kidding! *ADHD Report, 14*(4), 1–7.

Harrison, A. G., & Armstrong, I. (2014). WISC-IV unusual digit span performance in a sample of adolescents with learning disabilities. *Applied Neuropsychology: Child, 3,* 152–160.

Harrison, A. G., & Edwards, M. J. (2010). Symptom exaggeration in post-secondary students: Preliminary base rates in a Canadian sample. *Applied Neuropsychology, 17*(2), 135–143.

Harrison, A. G., Edwards, M. J., Armstrong, I., & Parker, K. C. H. (2010). An investigation of methods to detect feigned reading disabilities. *Archives of Clinical Neuropsychology, 25*(2), 89–98.

Harrison, A. G., Edwards, M. J., & Parker, K. C. H. (2007). Identifying students faking

ADHD: Preliminary findings and strategies for detection. *Archives of Clinical Neuropsychology, 22*(5), 577–588.

Harrison, A. G., Edwards, M. J., & Parker, K. C. H. (2008). Identifying students feigning dyslexia: Preliminary findings and strategies for detection. *Dyslexia, 14*(3), 228–244.

Harrison, A. G., Flaro, L., & Armstrong, I. (2014). Rates of effort test failure in children with ADHD: An exploratory study. *Applied Neuropsychology: Child*. Available at *www.pubfacts.com/detail/25257943/Rates-of-Effort-Test-Failure-in-Children-With-ADHD:-An-Exploratory-Study.*

Harrison, A. G., Green, P., & Flaro, L. (2012). The importance of symptom validity testing in adolescents and young adults undergoing assessments for learning or attention difficulties. *Canadian Journal of School Psychology, 27*(1), 98–113.

Harrison, A. G., Lovett, B., & Gordon, M. (2013). Documenting disabilities in postsecondary settings: Diagnosticians' understanding of legal regulations and diagnostic standards. *Canadian Journal of School Psychology, 28*(4), 303–322.

Harrison, A. G., Rosenblum, Y., & Currie, S. (2010). Examining unusual digit span performance in a population of postsecondary students assessed for academic difficulties. *Assessment, 17*, 283–293.

Harrison, A. G., & Wolforth, J. (2012). Findings from a pan-Canadian survey of disability services providers in postsecondary education. *International Journal of Disability, Community, and Rehabilitation, 11*(1). Available at *www.ijdcr.cu/VOL11_01/index.shtml.*

Individuals with Disabilities Education Act. (IDEA 2004). Building the legacy. Retrieved March 23, 2015, from *http://idea.ed.gov.*

Iverson, G. L., & Franzen, M. D. (1996). Using multiple objective memory procedures to detect simulated malingering. *Journal of Clinical and Experimental Neuropsychology, 18,* 38–51.

Jachimowicz, G., & Geiselman, R. E. (2004). Comparison of ease of falsification of attention deficit hyperactivity disorder diagnosis using standard behavioral rating scales. *Cognitive Science Online, 2,* 6–20.

Jasinski, L. J., & Ranseen, J. D. (2011, December). Malingered ADHD evaluations: A further complication for accommodations reviews. *Bar Examiner,* 6–16.

Kirk, J. K., Hutaff-Lee, C., Connery, A., Baker, D. A., & Kirkwood, M. W. (2014). The relationship between the self-report BASC-2 validity indicators and performance validity test failure after pediatric mild traumatic brain injury. *Assessment, 21, 562–569.*

Kirkwood, M. W., Connery, A. K., Kirk, J. W., & Baker, D. A. (2013). Detecting performance invalidity in children: Not quite as easy as A, B, C, 1, 2, 3 but automatized sequences appears promising. *Child Neuropsychology, 20,* 245–252.

Kirkwood, M. W., & Kirk, J. W. (2010). The base rate of suboptimal effort in a pediatric mild TBI sample: Performance on the Medical Symptom Validity Test. *The Clinical Neuropsychologist, 24*(5), 860–872.

Kirkwood, M. W., Kirk, J. W., Blaha, R. Z., & Wilson, P. (2010). Noncredible effort during pediatric neuropsychological exam: A case series and literature review. *Child Neuropsychology, 16,* 604–618.

Larochette, A.-C., & Harrison, A. G. (2012). Word Memory Test performance in Canadian adolescents with learning disabilities: A preliminary study. *Applied Neuropsychology: Child, 1*(1), 38–47.

Leark, R. A., Dixon, D., Hoffman, T., & Huynh, D. (2002). Fake bad test response bias effects on the Test of Variables of Attention. *Archives of Clinical Neuropsychology, 17*(4), 335–342.

Lester, G., & Kelman, M. (1997). State disparities in the diagnosis and placement of pupils with learning disabilities. *Journal of Learning Disabilities, 30*(6), 599–607.

Lewandowski, L. J., Cohen, J. A., & Lovett, B. J (2013). Effects of extended time allotments on reading comprehension performance of college students with and without learning disabilities. *Journal of Psychoeducational Assessment, 31*, 326–336.
Lindstrom, W., Coleman, C., Thomassin, K., Southall, C., & Lindstrom, J. (2011). Simulated dyslexia in postsecondary students: Description and detection using embedded validity indicators. *The Clinical Neuropsychologist, 25*(2), 302–322.
Lindstrom, W. A., Lindstrom, J. H., Coleman, C., Nelson, J., & Gregg, N. (2009). The diagnostic accuracy of symptom validity tests when used with postsecondary students with learning disabilities: A preliminary investigation. *Archives of Clinical Neuropsychology, 24*, 659–669.
Lu, P. H., & Boone, K. B. (2002). Suspect cognitive symptoms in a 9-year-old child: Malingering by proxy? *The Clinical Neuropsychologist, 16*(1), 90–96.
Mantynen, H., Poikkeus, A. M., Ahonen, T., Aro, T., & Korkman, M. (2001). Clinical significance of test refusal among young children. *Child Neuropsychology, 7*, 241–250.
McCabe, S. E., Teter, C. J., & Boyd, C. J. (2006). Medical use, illicit use and diversion of prescription stimulant medication. *Journal of Psychoactive Drugs, 38*, 43–56.
Merten, T., Bossink, L, & Schmand, B. (2007). On the limits of effort testing: Symptom validity tests and severity of neurocognitive symptoms in nonlitigant patients. *Journal of Clinical and Experimental Neuropsychology, 29*(3), 308–318.
Miller, L. A., Lewandowski, L. J., & Antshel, K. M. (2013, April 16). Effects of extended time for college students with and without ADHD. *Journal of Attentional Disorders*. Retrieved from *http://jad.sagepub.com/content/early/2013/04/16/1087054713483308.abstract*.
Mitchell, H. (2012, January 25). Faking ADHD gets you into Harvard. *Daily Beast*. Retrieved from *www.thedailybeast.com/articles/2012/01/25/faking-adhd-gets-you-into-harvard.html*.
Moore, A. S. (2010). Accommodations angst. *New York Times: Education Life*. Retrieved February 18, 2014 from *www.nytimes.com/2010/11/07/education/edlife/07strategy-t.html?pagewanted=all&_r=0*.
National Association of School Psychologists. (2014). Diagnosis and treatment of attention disorders: Roles for school personnel. Retrieved February 18, 2014, from *www.nasponline.org/resources/factsheets/add_fs.aspx*.
National Center for Learning Disabilities. (2014). The state of LD 2014. Retrieved February 18, 2014, from *www.ncld.org/types-learning-disabilities/what-is-ld/state-of-learning-disabilities*.
National Educational Association of Disabled Students. (2014). Financial aid directory. Available at *www.neads.ca/en/norc/funding/page04.php*.
National Research Council. (2004). *Engaging schools: Fostering high school students' motivation to learn*. Washington, DC: National Academic Press.
Ofiesh, N., Hughes, C., & Scott, S. (2004). Extended test time and postsecondary student with learning disabilities: A model for decision making. *Learning Disabilities Research and Practice, 19*(1), 57–70.
Ontario Student Drug Use and Health Survey. (2013). Toronto, Ontario, Canada: Centre for Addiction and Mental Health. Available at *www.camh.ca/en/research/news_and_publications/ontario-student-drug-use-and-health-survey/Pages/default.aspx*.
Osmon, D. C., Plambeck, E. A., Klein, L. & Mano, Q. (2006). The Wording Reading Test in adult learning disability: A simulation study. *The Clinical Neuropsychologist, 20*, 315–324.
Partnership for Drug-Free Kids. (2012). Partnership Attitude Tracking Study. Retrieved from *www.drugfree.org/newsroom/full-report-and-key-findings-the-2012-partnership-attitude-tracking-study-sponsored-by-metlife-foundation*.

Quinn, C. A. (2003). Detection of malingering in assessment of adult ADHD. *Archives of Clinical Neuropsychology, 18,* 379–395.

Rambo, P., Callahan, J., Hogan, L., Hullmann, S., & Wrape, L. (2015). Effort testing in children: Can cognitive and symptom validity measures differentiate malingered performances? *Applied Neuropsychology: Child, 4*, 1–8.

Rey, A. (1964). *L'examen clinique en psychologie*. Paris: Presses Universitaires de France.

Schatz, A. M., Ballantyne, A. O., & Trauner, D. A. (2001). Sensitivity and specificity of a computerized test of attention in the diagnosis of attention-deficit/hyperactivity disorder. *Assessment, 8*, 357–365.

Setlik, J., Bond, R., & Ho, M. (2009). Adolescent prescription ADHD medication abuse is rising along with prescriptions for these medications. *Journal for the American Academy of Pediatrics,* 10.1542(124), 875–880.

Sireci, S. G., Scarpati, S. E., & Li, S. (2005). Test accommodations for students with disabilities: An analysis of the interaction hypothesis. *Review of Educational Research, 75*, 457–490.

Slick, D. J., & Sherman, E. M. S. (2013). Differential diagnosis of malingering. In D. C. Carone & S. S. Bush (Eds.), *Mild traumatic brain injury: Symptom validity assessment and malingering* (pp. 57–72). New York: Springer.

Sollman, M. J., Ranseen, J. D., & Berry, D. T. (2010). Detection of feigned ADHD in college students. *Psychological Assessment, 22*, 325–335.

Statistics Canada. (2007). *Participation and activity limitation survey 2006: Technical and methodological report*. Retrieved from *www.statcan.gc.ca/pub/89-628-x/89-628-x2007001-eng.htm*.

Suhr, J., Hammers, D., Dobbins-Buckland, K., Zimak, E., & Hughes, C. (2008). The relationship of malingering test failure to self-reported symptoms and neuropsychological findings in adults referred for ADHD evaluation. *Archives of Clinical Neuropsychology, 23*(5), 521–530.

Suhr, J. A., Sullivan, B. K., & Rodriguez, J. L. (2011). The relationship of noncredible performance to Continuous Performance Test scores in adults referred for attention-deficit/ hyperactivity disorder evaluation. *Archives of Clinical Neuropsychology, 26*, 1–7.

Sullivan, B. K., May, K., & Galbally, L. (2007). Symptom exaggeration by college adults in attention-deficit hyperactivity disorder and learning disorder assessments. *Applied Neuropsychology, 14*(3), 189–207.

Talwar, V., Lee, K., Bala, N., & Lindsay, R. C. L. (2004). Children's lie-telling to conceal a parent's transgression: Legal implications. *Law and Human Behavior, 28*(4), 411–435.

Tapper, J., Morris, D., & Setrakian, L. (2006, March 30). Does loophole give rich kids more time on SAT? *ABC News*. Retrieved from *http://abcnews.go.com/Nightline/story?id=1787712*.

Tombaugh, T. N. (1996). *TOMM: Test of memory malingering*. North Tonawanda, NY: Multi-Health Systems.

U. S. Internal Revenue Service. (2013). Medical and dental expenses (Publication No. 502). Retrieved January 24, 2014, from *www.irs.gov/pub/irs_pdf/p502.pdf*.

Usher, A., & Kober, N. (2012). Student motivation: An overlooked piece of school reform. Retrieved on January 26, 2014, from *www.cep-dc.org/displayDocument.cfm?DocumentID=405*.

White, B. P., Becker-Blease, K. A., & Grace-Bishop, K. (2006). Stimulant medication use, misuse, and abuse in an undergraduate and graduate student sample. *Journal of American College Health, 54*, 261–268.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *The Woodcock–Johnson III*. Itasca, IL: Riverside.

11

Pediatric Clinical Neuropsychological Evaluations with Medical Populations

BRIAN L. BROOKS

Within medical settings, neuropsychologists provide valuable information on a child's functioning to aid with diagnosis, treatment/rehabilitation, and monitoring of disease (Baron, Fennell, & Voeller, 1995; Yeates, Ris, Taylor, & Pennington, 2009). Typically, hospital-based neuropsychologists function within an interdisciplinary team and are viewed as a key aspect of comprehensive medical care. In many circumstances, neuropsychologists in medical settings function as direct consultants to pediatricians, intensivists, physiatrists, neurologists, and neurosurgeons. However, the information gathered about a child's functioning is also used for non-hospital-based interventions, such as academic accommodations and school planning.

Like most other settings in which neuropsychologists are involved in a child's care, the types of evaluations provided for medical populations are driven by the referral question. Assessments within the medical setting may take the form of rapid screening evaluations to track a patient's response to treatment, or they may involve a more comprehensive evaluation of a child's cognitive, psychological, and academic skills following a change in neurological status (e.g., an acquired injury) or in anticipation of an expected change (e.g., pre- and postneurosurgery). They may be conducted with inpatients during the acute or subacute phases of an illness or with outpatients as part of their general rehabilitation program.

Every setting will be unique regarding the medical populations seen for evaluations. In tertiary hospitals in which many pediatric neuropsychologists practice, common medical diagnoses for the patients referred include epilepsy, traumatic brain injury (TBI), stroke, cancer, and hydrocephalus. Regardless of the specific diagnoses of patients seen for neuropsychological assessments, one common goal underlying all evaluations with medical populations is the need to have confidence in the obtained data as an adequate representation of a child's current functioning.

Confidence in obtained data is achieved through multiple sources, including (but not limited to) behavioral observations, consistency between what is known about a disease and the presentation, and the use of performance validity tests (PVTs) and symptom validity tests (SVTs). Observations and experience with a disease presentation remain as subjective means of determining validity, whereas PVTs provide an objective marker of validity that is measurable, replicable, and rooted in psychometric research. Of course, clinical judgment remains necessary when interpreting PVT performance. This chapter is designed to provide information on the use of PVTs with children and adolescents who have a medical diagnosis. It provides a rationale for using PVTs with these youth, reviews the existing literature on PVTs in various medical diagnoses, and comments on the existing limitations of the literature base for PVTs.

RATIONALE FOR VALIDITY TESTING WITHIN A MEDICAL SETTING

When should clinicians use PVTs and with which clinical populations? The answer to this question is simple: Always and with every assessment should be the goal. There are several broad false conceptualizations about using PVTs that have an impact on using these measures in medical settings. Many pediatric clinicians believe that PVTs (1) are applicable only for forensic referrals and are not needed as part of day-to-day clinical assessments, (2) are necessary only when a clinician already suspects poor engagement in the assessment, (3) are equivalent to "malingering" when failed, and (4) are not needed when evaluating children and are intended only for use with adult neuropsychological assessments.

When considering the use of PVTs in neuropsychological assessments with children and adolescents, clinicians may think that the necessity is present only when cases are litigious or when they already "suspect" that a patient is deliberately underperforming based on behavioral observations. However, it is strongly recommended that PVTs be used regularly in *all* assessments, including those conducted within medical settings, regardless of whether litigation or suspicion of underperforming are present or not. Evaluation of the validity of any assessment should come through multiple sources (as noted previously), with clinical judgment providing the ultimate decision based on careful consideration of all data.

Performing below an established cutoff score on a PVT does not automatically equate with malingering. Malingering is the "intentional production of false or grossly exaggerated physical or psychological symptoms, motivated by external incentives" (American Psychiatric Association, 2013, p. 726). Although people who malinger may be *more likely* to perform below cutoff scores on PVTs, it is not true that those who perform below cutoff scores are necessarily malingering. Performing below cutoff scores on PVTs may occur for various volitional and avolitional reasons, of which one *may* be malingering (Slick & Sherman, 2012; see also Sherman, Chapter 2, this volume). Determination of "malingering" requires the clinician to investigate the reason(s) for underperforming. In medical settings, PVT performance is best used to determine the "validity" or "credibility" of data as a way of providing confidence to the clinician that a child complied with the assessment demands. It is also possible that children who obtain PVT performance below a cutoff score were compliant with the assessment demands but have true cognitive impairment that limits their ability to achieve a higher score. However, based on clinical experience and available research, it would take substantial cognitive impairment to not achieve a "pass" on many of the PVTs with validity evidence for use in young children (e.g., picture-based PVTs).

Traditionally, PVTs have been designed for, standardized on, and used with adult populations. Position papers put forth by the National Academy of Neuropsychology (NAN; Bush et al., 2005) and the American Academy of Clinical Neuropsychology (AACN; Heilbronner et al., 2009) have advocated for the use of PVTs in all neuropsychological assessments, although the message is most often applied to clinical assessments with adults. Although the focus of these position papers was not necessarily geared toward a specific patient age group (i.e., adult versus children), Heilbronner et al. (2009) recommended that "effort measures and embedded validity indicators should be applied to pediatric samples" (p. 1107) as part of the future scientific investigations for neuropsychologists.

STATE OF SETTING-SPECIFIC SCIENTIFIC LITERATURE ABOUT VALIDITY TESTING

The amount of peer-reviewed literature on validity testing in children and adolescents with medical diagnoses has dramatically increased in the past few years (see Figure 11.1). Prior to 2010, there were very few published manuscripts that included pediatric patients with medical diagnoses, although there was increasing evidence of the use of PVTs in healthy children. From 2010 until 2014, there has been a surge of published studies on PVTs in medical populations. Of the 24 studies identified since 2010, 12 of them (50%) have a publication date of 2014 or are in press at the time of writing this chapter.

The majority of peer-reviewed studies examining PVTs in youth with

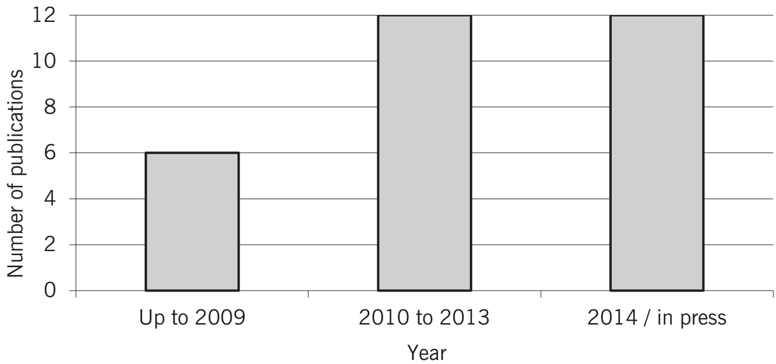

FIGURE 11.1. PVT publications involving pediatric patients with a medical diagnosis. The number of publications includes both case reports and group studies but does not include review articles. Search terms used for articles were consistent with De Right and Carone (2015) and included (pediatric OR child*) AND ("symptom validity test*" OR "malingering" OR "effort test" OR "validity test*" OR "childhood deception").

medical diagnoses have provided evidence based on three populations: (1) mild traumatic brain injury (mild TBI), (2) epilepsy, and (3) heterogeneous medical samples. Although the mild TBI literature may have overlap with reviews of PVTs in sport concussion evaluations (see Rohling, Langrinrichsen-Rohling, & Womble, Chapter 12, this volume) and in forensic neuropsychological evaluations (see Donders, Chapter 13, this volume), it is important to also consider the evidence as part of a medical population because patients with mild TBIs are not always injured in sports, they are not always involved in forensic proceedings, and they often present to the hospital or specialized tertiary clinic for medical care. In the following section, the evidence for using PVTs in these medical populations is presented and discussed. Where possible, the "pass"[1] rates will identify true positives and false positives.

PVTs in Youth with Traumatic Brain Injury

Each year, it is estimated that 6 out of every 1,000 young people will sustain a mild TBI (Cassidy et al., 2004), with young children and adolescents being two of the highest risk age groups (Rutland-Brown, Langlois,

[1] "Pass" being defined as performance above an established cutoff score on a PVT. Scoring below a cutoff does not necessarily mean that a person cannot be deemed to be a false positive, and scoring above a cutoff does not necessarily mean a person cannot be deemed a false negative.

Thomas, & Xi, 2006). Most often mild TBIs in young children are caused by falls, with a fall from a height or a fall on the ground accounting for nearly three-quarters of the injury mechanisms (Andersson, Sejdhage, & Wage, 2012). In adolescence, many mild TBIs are sustained through sport (McCrory et al., 2013). Although many youth may not seek medical care for these milder injuries and the vast majority recover within a relatively brief time period (Barlow et al., 2010), a significant minority require specialized care due to preinjury risk factors and/or protracted recovery. Within the scope of specialized care for youth with mild TBI, neuropsychologists are frequently involved and provide critical information about cognitive and psychological functioning.

In the adult mild TBI literature, the rates of noncredible performance on neuropsychological testing can approach 40–50% in published samples (Armistead-Jehle, 2010; Larrabee, 2003), particularly when there may be external incentives. The existing literature on PVT performance in youth with mild TBI provides the largest number of publications among the medical diagnostic groups. Table 11.1 summarizes the studies to date with youth with mild TBI. Of the 10 studies involving youth with mild TBI who were administered a PVT as a primary outcome, 8 were completed by Kirkwood and colleagues. In addition, a case series published by Kirkwood, Kirk, Blaha, and Wilson (2010) included several children with mild TBI, although the focus of the study was examining specific cases with noncredible performance rather than being a population-based study.

The first published study on PVT use in a group of youth with mild TBI showed that 83% of patients 8–17 years of age passed validity testing (Kirkwood & Kirk, 2010). More than 90% of the failures on the Medical Symptom Validity Test (MSVT; Green, 2004) in this study were considered true positives for noncredible effort. Across studies by Kirkwood and colleagues that have used the MSVT, between 82% (Kirk, Hutaff-Lee, Connery, Baker, & Kirkwood, 2014) and 88% (Kirkwood, Peterson, Connery, Baker, & Grubenhoff, 2014) of youth with mild TBI have obtained performance above cutoff scores originally established in adults, with the authors again considering failure in most cases to be evidence of noncredible effort. The samples have consistently been evaluated at a median of 6–7 weeks postinjury, so this would be considered outside the acute range postinjury and into the slow-to-recover time frame. To date, there has not been information on PVT performance in youth tested soon after mild TBI and whether failure rates in the acute period are consistent with rates in those who are slow to recover.

Most published literature contains evidence for using PVTs in those youth with mild TBIs, with a clear paucity on the performance of those with more substantial TBIs (see Carone, 2008, for an example of a study including children with more severe injuries). As part of a larger study with children and adolescents who have various medical diagnoses, Loughan

TABLE 11.1. Summary of PVT Studies with Pediatric TBI

Population	Author (year)	Sample size	Age range (mean)	PVT used	PVT type	Percent above cutoff score	SN/SP
Mild TBI	Kirkwood & Kirk (2010)	193	8–17 years (14.5)	MSVT TOMM	Stand-alone Stand-alone	83% —[a]	
	Kirkwood et al. (2011)	274	8–16 years (14.2)	Digit Span ACSS RDS MSVT TOMM	Embedded Embedded Stand-alone Stand-alone Stand-alone	81.8% —[a]	.51/.96 .51/.92
	Kirkwood et al. (2012)	276	8–16 years (14.2)	MSVT	Stand-alone	81%	
	Kirkwood et al. (2013)	439	8–17 years (14.7)	AST MSVT	Embedded Stand-alone	83.8%	.55/.90
	Araujo et al. (2014)	382	8–16 years (13.1)	RDS	Embedded	88%	
	Green et al. (2014)	319	8–17 years (14.6)	Rey 15 MSVT	Stand-alone Stand-alone	85%	.55/.91
	Kirkwood et al. (2014)	191	8–17 years (14.5)	MSVT	Stand-alone	88%	
	Kirk et al. (2014)	274	8–17 years (14.7)	BASC-2 MSVT	Questionnaire Stand-alone	88% 82%	
	Baker et al. (2014)	411	8–16 years (14.2)	CVLT-C MSVT	Embedded Stand-alone	83.9%	.55/.91
	Ploetz et al. (in press)	39	< 18 years	TOMM	Stand-alone	100%	
Complicated mild to severe TBI[b]	Ploetz et al. (in press)	35	< 18 years	TOMM	Stand-alone	91%	
TBI (severity unknown)	Loughan & Perna (2014)	19	< 18 years (11.8)	TOMM	Stand-alone	78.9%	

Note. TBI, traumatic brain injury; PVT, performance validity test; SN/SP, sensitivity/specificity based on author-defined cutoff scores; MSVT, Medical Symptom Validity Test; TOMM, Test of Memory Malingering; ACSS, age-corrected scaled score for Digit Span; RDS, Reliable Digit Span; WMT, Word Memory Test; BASC-2, Behavior Assessment System for Children, Second Edition.

[a]The second PVT was administered only to those who did not pass the first PVT, so pass/fail rates would not represent the entire sample.

[b]All patients with a mild TBI had positive neuroimaging, and those with a moderate to severe TBI had positive neuroimaging, Glasgow Coma Scale score of 12 or less, loss of consciousness greater than 30 minutes, and/or posttraumatic amnesia greater than 24 hours.

and Perna (2014) provided information on PVT pass rates in a subsample of youth who had sustained a TBI. On the Test of Memory Malingering (TOMM; Tombaugh, 1996), 79% of youth with a TBI passed based on adult cutoff scores. In this study, however, no information about true versus false positives was reported, the TBI sample was small (*n* = 19), it was uncertain how the TBIs were sustained, the time from injury to assessment was not noted, and information on the severity of the brain injuries was not provided. It is assumed that some of these failures would have been true positives and that this is a mixed sample ranging from mild to severe, but these are just assumptions.

Ploetz, Mazur-Mosiewicz, Kirkwood, Sherman, and Brooks (in press) presented data in a subsample of youth who were identified as having a TBI that ranged in severity from complicated mild to severe (i.e., all had positive neuroimaging findings). On the TOMM, 91% of the sample achieved performance at or above the cutoff scores established for adults. Furthermore, none of those who had scores below the cutoff were deemed to have been a false positive; in other words, each of these TBI cases failing the TOMM was judged to have been providing noncredible effort. Similar to the Loughan and Perna (2014) study, the sample size in Ploetz et al. (in press) was small (*n* = 35), and the time from injury to assessment was not noted, which limits the generalizability of the results.

In addition to the studies looking at stand-alone PVT performance, there is some literature on the use of embedded PVTs in youth with mild TBIs. Araujo et al. (2014) found that 80% of youth with mild TBI were above cutoffs on two embedded markers on Digit Span when tested on average 18–19 days postinjury (2–3 weeks). Specifically, 88% had a Reliable Digit Span (RDS; Greiffenstein, Baker, & Gola, 1994) score above 6, and 83% had an age-adjusted scales score above 5. The authors used these cutoffs to establish credible and noncredible responder groups, suggesting that they believed most of the performances below the cutoffs were true positives for noncredible effort.

Most other studies with youth with mild TBI have focused on the development of cutoff scores for embedded markers, so the individual PVT pass rates are not published. To date, there is literature on the establishment of cutoff scores for the Digit Span Scaled Score and RDS (Kirkwood, Hargrave, & Kirk, 2011), the Automatized Sequences Test (AST; Kirkwood, Connery, Kirk, & Baker, 2013), and the California Verbal Learning Test—Children's Version (CVLT-C; Baker, Connery, Kirk, & Kirkwood, 2014) when using these measures with youth with mild TBI. The proposed cutoff scores specific for use in pediatric assessments have all suggested that sensitivity is approximately 50% when specificity is held at ≥ 90%. Further studies are needed to confirm these suggested cutoff scores, but the existing literature suggests promise with embedded PVTs in youth with mild TBI.

One area of validity research that has drastically lagged behind the adult literature has been the use of embedded *symptom validity tests* in questionnaires. In contrast to PVTs that measure performance-based validity, these SVTs allow the determination of whether subjectively rated problems are credible, often presenting as inconsistent reporting, endorsing excessively negative items to make oneself look bad, and failing to endorse commonly occurring issues to make oneself look overly positive. In youth with mild TBI, Kirk et al. (2014) found that 92% of their sample was determined to have passed all five validity indicators on the self-report version of the Behavior Assessment Scale for Children—Second Edition (BASC-2; Reynolds & Kamphaus, 2004). The percentage of the mild TBI sample that was "flagged" on the individual validity scales ranged from .7% (Response Pattern and V scale) to 5.4% (L scale). Interestingly, however, those youth with mild TBI who failed the MSVT were *not* more likely to provide an invalid profile on the questionnaire. Obviously more research is needed, but these initial findings suggest that the measurement of noncredible performance on cognitive tests may not necessarily correspond with self-report data on the BASC-2, and that noncredible symptom reporting in youth with mild TBI may need to be established through means other than being based on those who do or do not pass stand-alone PVTs (e.g., Araujo et al., 2014; Kirkwood et al., 2014). The findings may also suggest that psychological and behavioral symptoms are less likely than physical symptoms to be over- or underendorsed in those with mild TBI who fail a PVT.

PVTs in Youth with Epilepsy

Neuropsychological assessments play a key role in any comprehensive pediatric epilepsy clinic (Sherman et al., 2011; Westerveld, 2010). Despite the importance of obtaining credible results in neuropsychological evaluations for youth with epilepsy, especially when teams use this information to make critical decisions on surgical resections, there is little research on the use of PVTs in this population (see Table 11.2). MacAllister, Nakhutina, Bender, Karantzoulis, and Carlson (2009) gave the TOMM to 60 youth between 6 and 17 years of age with various epilepsy syndromes; 36 had partial epilepsy; 12 had generalized epilepsy; and the remaining 12 were either mixed or unclassified. As well, 5 of the participants had already undergone a partial resection, with some participants being referred to neuropsychology as part of their presurgical workup. MacAllister et al. (2009) found that 90% of these patients with epilepsy were able to pass this PVT using adult cutoff scores, with performance being positively correlated with intellectual level but not age. Furthermore, of the 6 youth who did not pass the TOMM, the authors believed that 4 of them were false positives due to very low cognitive abilities and interictal electroencephalogram (EEG) discharges during testing.

TABLE 11.2. Summary of PVT Studies with Pediatric Epilepsy

Author (year)	Sample size	Age range (mean)	PVT used	PVT type	Percent above cutoff score
MacAllister et al. (2009)	60	6–17 years (12.1)	TOMM	Stand-alone	90%
Welsh et al. (2012)	54	6–17 years (13.0)	RDS TOMM	Embedded Stand-alone	65% 90%
Ploetz et al. (in press)	98	< 18 years	TOMM	Stand-alone	96%

Note. PVT, performance validity test; TOMM, Test of Memory Malingering; RDS, Reliable Digit Span.

TOMM performance in another sample of youth with epilepsy was presented by Ploetz et al. (in press) as part of their larger sample of children and adolescents with neurological diagnoses. In this group of 98 youth with epilepsy who are followed through a tertiary hospital, 96% were able to achieve scores on the TOMM at or above the cutoff scores established for adults. Of the small proportion who did not perform above the cutoff score for the TOMM, three out of four were deemed to have been false positives.

Welsh, Bender, Whitman, Vasserman, and MacAllister (2012) used 54 of the 60 youth with a diagnosis of epilepsy from the MacAllister et al. (2009) study to look at performance on RDS. Similar to the larger sample of 60 youth, 90% of this sample passed the TOMM based on adult criteria. When using an RDS cutoff of ≤ 6 to suggest poor compliance (from Kirkwood et al., 2011), only 65% of the sample was identified as passing this embedded PVT. Clearly this is much lower than the 88% pass rate reported by Araujo et al. (2014) in youth with mild TBI and is much lower than the desired pass rate of 90% (Babikian, Boone, Lu, & Arnold, 2006). RDS scores in this sample of youth with epilepsy were positively and significantly correlated with both intelligence and age. Based on this study, Welsh and colleagues suggested that embedded measures such as RDS may not have as much utility in patient populations with significant cognitive problems because they are likely to produce higher false-positive rates. The authors also suggested alternative cutoff scores for the RDS in youth with epilepsy, including scores of 4 or less that result in sensitivity of 60% and specificity of 89%.

Although the work by MacAllister and colleagues (2009) and Ploetz and colleagues (in press) explore using PVTs in youth with epilepsy, their work truly is just a start. Interestingly, there is a glaring shortage of PVT evidence in pediatric patients who are undergoing epilepsy surgery. Given the significance of having credible findings on testing when informing decisions on surgical resections, the use of PVTs should be an area of top priority. Stand-alone PVTs may be less affected by true cognitive sequelae;

therefore, these instruments may be appropriate to use in neuropsychological evaluations with youth who have epilepsy. On the other hand, the study with RDS suggests that embedded measures need to be carefully evaluated for use in children, especially when applying adult-based cutoff scores or even values that are established using other pediatric patient populations with very different cognitive functioning.

PVTs in Youth with Heterogeneous Diagnoses

The medical diagnoses reviewed so far (TBI and epilepsy) represent two fairly specific diagnostic groups often seen in medical settings. Although there is actually considerable heterogeneity within each of those diagnoses (e.g., the International League Against Epilepsy identifies 14 classifications of seizure types and 21 electrographic syndromes and other epilepsies in children; Berg et al., 2010), the literature does start to provide support for PVT use in youth specifically with diagnoses of mild TBI or epilepsy. The broader literature base with youth who have various medical diagnoses, however, provides much more evidence for the general use of PVTs in youth.

Table 11.3 presents studies with youth who have mixed diagnoses—that is, samples with heterogeneous diagnoses (e.g., medical disorders, psychological disorders, attentional/learning disorders). For the majority of the studies, the outcomes presented are for the entire mixed diagnosis group and not for the individual diagnoses. Pass rates at the group level on stand-alone and embedded PVTs have ranged from 82 to 97% in these heterogeneous samples (85–97% pass stand-alone PVTs and 82% pass embedded markers). A few subgroups with smaller sample sizes in these studies, notably a group with learning disabilities (LD) from Loughan and Perna (2014) and a group with hydrocephalus from Ploetz et al. (in press), were able to achieve 100% above the established cutoff score for TOMM.

Just over a decade ago, the research on PVT use in children and adolescents with medical diagnoses commenced, which shows the relative infancy of this literature base compared with adult PVT literature that has flourished for a few decades. Green and Flaro (2003) administered the Word Memory Test (WMT; Green, 2003) to youth between 7 and 18 years of age who had various medical and psychiatric diagnoses. Although the WMT requires a third-grade reading level, the administration and determination of cutoff scores were kept the same for all children (regardless of reading level) as would be for adults. In this sample, which included youth with fetal alcohol effects/syndrome, schizophrenia, bipolar mood disorder, attention-deficit/hyperactivity disorder (ADHD), conduct disorder, oppositional defiant disorder, learning disabilities, and a mixed neurological group, 86% were able to successfully pass the WMT based on adult cutoff scores. Green and Flaro (2003) further reported that performance on the

TABLE 11.3. Summary of PVT Studies with Mixed Pediatric Diagnoses

Author (year)	Sample size	Age range (mean)	PVT used	PVT type	Percent above cutoff score	SN/SP
Green & Flaro (2003)	135	7–18 years (12.6)	WMT	Stand-alone	85.9%	
Courtney et al. (2003)	111	6–17 years (11.2)	CARB WMT	Stand-alone Stand-alone	—[a] —[a]	
Donders (2005)	100	6–16 years (11.9)	TOMM	Stand-alone	97%	
Carone (2008)	38	< 18 years (11.8)	MSVT	Stand-alone	94.7%	
Brooks et al. (2011)	53	6–19 years (12.4)	TOMM	Stand-alone	94.3%	
Kirk et al. (2011)	101	5–16 years (10.7)	TOMM	Stand-alone	96%	
Brooks (2012)	100	6–19 years (14.0)	VSVT	Stand-alone	95%	
Loughan et al. (2012)	51	6–18 years (11.8)	TOMM RDS	Stand-alone Embedded	86.2%	.43/.91
Perna & Loughan (2013)	75	6–18 years (11.5)	TOMM	Stand-alone	88%	
Perna & Loughan (2014)	75	6–18 years (11.5)	TOMM	Stand-alone	87%	
Perna et al. (2014)	75	6–18 years (11.5)	Digit Span ACSS CMS Verbal Memory TOMM	 Embedded Embedded Stand-alone	88%	.44/.94 .11/.90
Brooks et al. (2014)	275	7–18 years (13.9)	CNS Vital Signs TOMM VSVT	Embedded Stand-alone Stand-alone	82.2% 88.4%[b] 84.5%[c]	
Loughan & Perna (2014)[d]						
Total sample	86	6–18 years (11.6)	TOMM	Stand-alone	87%	
Affective disorder	44				90.9%	
Attention-deficit/hyperactivity disorder	55				92.7%	
Conduct disorder	22				81.8%	
Intellectual disability	16				68.8%	
Learning disability	20				100%	
Pervasive developmental disability	7				85.7%	

(*continued*)

TABLE 11.3. (*continued*)

Ploetz et al. (in press)[d]					
Total sample	266	5–18 years (13.0)	TOMM	Stand-alone	94%
Stroke	37				86%
Hydrocephalus	13				100%

Note. PVT, performance validity test; SN/SP, sensitivity/specificity based on author-defined cutoff scores; WMT, Word Memory Test; CARB, Computerized Assessment of Response Bias; TOMM, Test of Memory Malingering; MSVT, Medical Symptom Validity Test; RDS, Reliable Digit Span; ACSS, age-corrected scaled score; CMS, Children's Memory Scale; VSVT, Victoria Symptom Validity Test.

[a]Pass rates were not reported.

[b]$N = 189$.

[c]$N = 110$.

[d]Because this study provided the pass rates for each diagnostic group, these subgroups have been presented here. Subgroups from these studies with TBI or epilepsy are presented in Tables 11.1 and 11.2, respectively.

WMT in their mixed pediatric sample was not influenced by age or level of intellectual abilities; however, additional research with the WMT in children does not necessarily support the absence of age effects in younger children (Courtney, Dinkins, Allen, & Kuroski, 2003).

Courtney and colleagues (2003) administered the WMT and the Computerized Assessment of Response Bias (CARB) to a mixed sample of children and adolescents ranging from 6 to 17 years of age who were referred for a clinical evaluation (attentional and learning disorders were the two most common diagnoses, but psychiatric and medical problems were also included in this sample). This study specifically sought to investigate the presence of potential age effects on PVTs, so actual pass rates for their sample on these PVTs were not reported. In contrast to the data reported by Green and Flaro (2003), Courtney et al. (2003) reported moderate positive correlations between age and PVT scores. Furthermore, when considering the adult-established instructions and cutoff scores for these measures, high failure rates on the WMT and CARB were found for children under 10 years of age. The authors cautioned against the use of these specific PVTs in young children, those with lesser reading skills, and those who are developmentally challenged. Unfortunately, those three characteristics are often seen in youth with medical diagnoses who are evaluated by neuropsychologists.

The MSVT, a shortened version of the WMT, has most often been studied in youth with mild TBIs (see previous section). However, Carone (2008) examined the performance on the MSVT in a sample of children and adolescents with moderate to severe acquired brain injuries (not limited to traumatic), and this was done specifically within the context of comparing their performance with that of adults with mild TBI. In those youth with moderate to severe brain dysfunction, 95% were able to pass this PVT using adult cutoff scores (and the two who "failed" the MSVT were deemed to be true positives). This level of pass rate in the youth was contrasted with only 79% of adults with mild TBI who passed the MSVT.

The majority of published studies that support using PVTs in children and adolescents with mixed medical diagnoses has utilized the TOMM. There are no clear reasons that research with the TOMM in youth has flourished compared with other PVTs. However, the absence of a reading component for this test, the appeal of black-and-white drawings to children, and initial research by Donders (2005) showing that children with medical diagnoses as young as 6 years can readily pass may all be relevant factors. Based on the existing literature, the pass rates for the TOMM in medical samples range from 86 to 97% when looking at group studies (subgroups within a few studies had pass rates ranging from 69 to 100%).

Of all the published studies with mixed medical samples, only two provide a further breakdown of PVT performance by the various diagnoses. In Loughan and Perna (2014), the highest pass rates for the TOMM were obtained in those with LD (100%), ADHD (93%), and affective disorder (91%). In comparison, the lowest pass rates were found in those with intellectual disability (69%) and in those with conduct disorder (82%). The sample sizes for some of these subanalyses were small; therefore, larger samples will be needed to replicate these findings. Ploetz et al. (in press) reported the highest pass rates of 100% for those with a concussion and hydrocephalus; those with epilepsy (96%), those with complicated mild to severe TBI (91%), and a mixed neurology group (91%) all had pass rates that exceeded 90%. Only the sample with pediatric stroke had a pass rate below 90% (i.e., 86%), with two-thirds of those below the cutoff deemed to be false positives due to substantial cognitive impairment.

Heterogeneous medical samples have also contributed to the literature on the use of embedded markers of validity. Generally, embedded validity markers in medical samples have lower sensitivity when specificity levels are held at .90 or higher. This has been true for RDS (sensitivity = .43; Loughan, Perna, & Hertza, 2012), for Digit Span Scaled Score (sensitivity = .44; Perna, Loughan, Hertza, & Segraves, 2014), for the Children's Memory Scale (Cohen, 1997), for Verbal Memory Recall > Recognition (sensitivity = .11; Perna et al., 2014), and CNS Vital Signs (sensitivity = .04–.35; Brooks, Sherman, & Iverson, 2014). Although embedded markers represent an optimal option for detecting poor engagement because additional measures are not needed, the lower sensitivity poses a difficulty for accurate positive detection on their own in the presence of true cognitive problems.

RECOMMENDATIONS: HOW VALIDITY TESTS SHOULD BE USED AND INTERPRETED

It is strongly recommended that PVTs be used in all neuropsychological assessments with all children and adolescents, regardless of the length or reason for the evaluation. Although the existing position statements on the use of PVTs in neuropsychological assessments are directed more toward evaluations with adult patients (Bush et al., 2005; Heilbronner et al., 2009), these statements are applicable to pediatric evaluations. The following cautions are provided when using PVTs with children and adolescents in medical settings. Each one of these factors needs to be carefully considered if present and if a child fails one or more PVTs.

Young Age

When using PVTs with children and adolescents, age is an important factor to consider. Across several studies, age has been positively correlated with PVT performance (Brooks, 2012; Brooks, Sherman, & Krol, 2011; Courtney et al., 2003). Younger children are more likely to have trouble remaining engaged with a task and/or complying with test demands. In turn, young children are more likely to be identified as falling below cutoff scores on PVTs and are more likely to be considered false positives. Ploetz et al. (in press) showed that 13% of 5- to 7-year-olds fell below established cutoff scores on the TOMM (over half of these were deemed to be false positives) compared with 8% of 8- to 10-year-olds (6 of the 8 failures were deemed to be false positives), 3% of the 11- to 13-year-olds (2 out of 3 were deemed to be false positives), and 5% of 14- to 18-year-olds (1 out of 5 was deemed to be a false positive). The evidence to date suggests that young children can readily pass PVTs, but age is definitely a factor that needs to be considered regarding pass rates and false-positive rates. As well, PVTs with reduced verbal requirements are most promising for use with the youngest children.

Low Intellectual Functioning or Significant Cognitive Impairment

Youth with lower intellectual abilities are able to pass PVTs based on adult-established cutoff scores, but there may be a higher proportion who falls below cutoff scores compared with youth with broadly normal intellectual abilities. Loughan and Perna (2014) suggest that those with intellectual disability may fall below established cutoffs at quite a high rate (31%), whereas Ploetz et al. (in press) suggest that the failure rate of those with intellectual disability is very low and within acceptable limits (6%, and all were deemed to be false positives). Although PVTs are not intended to be direct measures of cognitive abilities, there is evidence that positive correlations are present between cognition and PVTs in medical samples (Brooks, 2012; Brooks et al., 2011; Ploetz et al., in press). The result can be that PVT performance may be affected by substantial and real cognitive impairment.

Reading Level

Some PVTs require examinees to read words. For example, the WMT and MSVT both stipulate at least a third-grade reading level (Green, 2003, 2004). In neuropsychological assessments within the medical setting, younger children are often developmentally and academically delayed, so that reading level may be an affected ability. It is important that reading level be evaluated and ruled out as a confounding variable when using PVTs with reading components. Although reading may be less of a concern with adolescents, use of these PVTs with younger children can be challenging.

FUTURE RESEARCH AND CLINICAL DIRECTIONS FOR PVTs WITH MEDICAL POPULATIONS

This chapter aimed to provide evidence for the use of PVTs when assessing children and adolescents with medical diagnoses. Although PVT use may be deemed to be for forensic purposes, for those situations in which poor engagement is already suspected, or with adults only, PVT use is clearly appropriate and necessary for all neuropsychological assessments. Future position papers should emphasize the necessity for broad use of PVTs in all pediatric assessments.

In the available literature, PVT performances in youth have been studied in those with mild TBIs and with epilepsy and in heterogeneous diagnostic groups. Generally the published rates of obtaining performances above cutoff scores on PVTs in the medical samples are quite high. In youth with mild TBIs, 82–88% perform above cutoff scores on stand-alone PVTs (many of the children in these studies were true positives for noncredible effort). In youth with epilepsy, there are only a few published studies, but the pass rate is strong at ≥ 90% on a stand-alone PVT and even higher when one accounts for false positives. Pass rates in youth with epilepsy on an embedded marker (RDS) are much lower compared with the sample with mild TBI (65% vs. 80%, respectively), which likely reflects the influence of true cognitive impairment. In the current literature involving large mixed samples of youth with various diagnoses, 85–100% pass stand-alone PVTs and 82% pass embedded markers.

The existing published literature in children and adolescents with medical diagnoses lays the groundwork for using PVTs in neuropsychological assessments, but there are some clear limitations that remind us that this line of research is still in its infancy. First, there are very few published studies. The studies reviewed in this chapter look at only a small selected number of diagnostic groups, notably mild TBI and epilepsy. The studies with heterogeneous samples are generally very mixed, and the sample sizes are too small to provide meaningful information for the different diagnoses. The literature on embedded PVTs is extremely limited as well, with very few markers being developed a priori for use in the detection of noncredible performance.

Second, the evidence for using PVTs in youth with medical diagnoses is limited to only a few PVTs. It is critical that clinicians and researchers have several measures to select during their evaluations, so having evidence on only a few PVTs for medical populations (e.g., WMT, MSVT, TOMM) is insufficient and supports further research with other measures.

Third, only a few groups of researchers are publishing studies on PVT use in pediatric medical populations. Notably, Kirkwood and colleagues, Perna and Loughan and colleagues, MacAllister and colleagues, and Brooks and colleagues account for the vast majority of existing literature. These aforementioned limitations do suggest that the PVT literature in youth with medical diagnoses is far from exhaustive.

The fourth limitation pertains to the PVTs themselves. PVTs have traditionally been developed and validated for use with adults. The research to date has taken adult-developed PVTs and applied them to pediatric samples using adult-based cutoff scores for determining noncredible performance. As a result, PVTs are often made to fit a pediatric population with several caveats for their use and interpretation, such as young age, lower cognitive functioning, and reduced reading level. Much like having cognitive measures that are developmentally appropriate for use in children because they were designed for children, it is time for PVTs to be developed a priori for use with children rather than translating measures from adult to pediatric use.

Despite the limitations of PVT literature in medical populations, it is important to recognize that the field is currently experiencing a surge of publications in this area (see Figure 11.1). The existing literature provides enough of a knowledge base for using PVTs in assessments with children and adolescents who have a medical diagnosis, but clearly more is needed. It is hoped that more researchers will examine PVT use in various diagnostic groups, that pediatric-specific PVTs will be developed, and that there will be an emphasis on the development of embedded markers.

AUTHOR DISCLOSURE

I am coauthor of the Child and Adolescent Memory Profile™, a pediatric memory battery with embedded validity markers (Sherman & Brooks, 2015) and the Memory Validity Profile™, a stand-alone PVT that is designed specifically for use in pediatric samples (Sherman & Brooks, in press).

REFERENCES

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: Author.

Andersson, E. E., Sejdhage, R., & Wage, V. (2012). Mild traumatic brain injuries in children between 0–16 years of age: A survey of activities and places when an accident occurs. *Developmental Neurorehabilitation, 15*(1), 26–30.

Araujo, G. C., Antonini, T. N., Monahan, K., Gelfius, C., Klamar, K., Potts, M., et al. (2014). The relationship between suboptimal effort and post-concussion symptoms in children and adolescents with mild traumatic brain injury. *The Clinical Neuropsychologist, 28*(5), 786–801.

Armistead-Jehle, P. (2010). Symptom validity test performance in U.S. veterans referred for evaluation of mild TBI. *Applied Neuropsychology, 17*(1), 52–59.

Babikian, T., Boone, K. B., Lu, P., & Arnold, G. (2006). Sensitivity and specificity of various Digit Span scores in the detection of suspect effort. *The Clinical Neuropsychologist, 20*(1), 145–159.

Baker, D.A., Connery, A.K., Kirk, J. W., & Kirkwood, M.W. (2014). Embedded performance validity indicators within the California Verbal Learning Test, Children's Version. *The Clinical Neuropsychologist, 28*(1), 116–127.

Barlow, K. M., Crawford, S., Stevenson, A., Sandhu, S. S., Belanger, F., & Dewey, D. (2010). Epidemiology of postconcussion syndrome in pediatric mild traumatic brain injury. *Pediatrics, 126*(2), e374–381.

Baron, I. S., Fennell, E. B., & Voeller, K. K. S. (1995). *Pediatric neuropsychology in the medical setting.* New York: Oxford University Press.

Berg, A. T., Berkovic, S. F., Brodie, M. J., Buchhalter, J., Cross, J. H., van Emde Boas, W., et al. (2010). Revised terminology and concepts for organization of seizures and epilepsies: Report of the ILAE Commission on Classification and Terminology, 2005–2009. *Epilepsia, 51*(4), 676–685.

Brooks, B. L. (2012). Victoria Symptom Validity Test performance in children and adolescents with neurological disorders. *Archives of Clinical Neuropsychology, 27*(8), 858–868.

Brooks, B. L., Sherman, E. M., & Iverson, G. L. (2014). Embedded validity indicators on CNS vital signs in youth with neurological diagnoses. *Archives of Clinical Neuropsychology, 29*(5), 422–431.

Brooks, B. L., Sherman, E. M., & Krol, A. L. (2011). Utility of TOMM Trial 1 as an indicator of effort in children and adolescents. *Archives of Clinical Neuropsychology, 27*(1), 23–29.

Bush, S. S., Ruff, R. M., Troster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., et al. (2005). Symptom validity assessment: Practice issues and medical necessity: NAN Policy and Planning Committee. *Archives of Clinical Neuropsychology, 20*(4), 419–426.

Carone, D. A. (2008). Children with moderate/severe brain damage/dysfunction outperform adults with mild-to-no brain damage on the Medical Symptom Validity Test. *Brain Injury, 22*(12), 960–971.

Cassidy, J. D., Carroll, L. J., Peloso, P. M., Borg, J., von Holst, H., Holm, L., et al. (2004). Incidence, risk factors and prevention of mild traumatic brain injury: Results of the WHO Collaborating Centre Task Force on Mild Traumatic Brain Injury. *Journal of Rehabilitation Medicine, 36*(43, Suppl.), 28–60.

Cohen, M. J. (1997). *Children's Memory Scale.* San Antonio, TX: Psychological Corporation.

Courtney, J. C., Dinkins, J. P., Allen, L. M., III, & Kuroski, K. (2003). Age-related effects in children taking the Computerized Assessment of Response Bias and Word Memory Test. *Child Neuropsychology, 9*(2), 109–116.

DeRight J., & Carone, D.A. (2015). Assessment of effort in children: A systematic review. *Child Neuropsychology, 21*(1), 1–24.

Donders, J. (2005). Performance on the Test of Memory Malingering in a mixed pediatric sample. *Child Neuropsychology, 11*(2), 221–227.

Green, C. M., Kirk, J. W., Connery, A. K., Baker, D. A., & Kirkwood, M. W. (2014). The use of the Rey 15-Item Test and recognition trial to evaluate noncredible effort after pediatric mild traumatic brain injury. *Journal of Clinical and Experimental Neuropsychology, 36*(3), 261–267.

Green, P. (2003). *Green's Word Memory Test*. Edmonton, Alberta, Canada: Green's Publishing.

Green, P. (2004). *Manual for Green's Medical Symptom Validity Test (MSVT)*. Edmonton, Alberta, Canada: Green's Publishing.

Green, P., & Flaro, L. (2003). Word Memory Test performance in children. *Child Neuropsychology, 9*(3), 189–207.

Greiffenstein, M., Baker, W., & Gola, T. (1994). Validation of malingered amnesia measures with a large clinical sample. *Psychological Assessment, 6*, 218–224.

Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., Millis, S. R., & Conference Participants. (2009). American Academy of Clinical Neuropsychology Consensus Conference Statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist, 23*(7), 1093–1129.

Kirk, J. W., Harris, B., Hutaff-Lee, C. F., Koelemay, S. W., Dinkins, J. P., & Kirkwood, M. W. (2011). Performance on the Test of Memory Malingering (TOMM) among a large clinic-referred pediatric sample. *Child Neuropsychology, 17*(3), 242–254.

Kirk, J. W., Hutaff-Lee, C. F., Connery, A. K., Baker, D. A., & Kirkwood, M. W. (2014). The relationship between the self-report BASC-2 validity indicators and performance validity test failure after pediatric mild traumatic brain injury. *Assessment, 21*(5), 562–569.

Kirkwood, M. W., Connery, A. K., Kirk, J. W., & Baker, D. A. (2013). Detecting performance invalidity in children: Not quite as easy as A, B, C, 1, 2, 3 but automatized sequences appears promising. *Child Neuropsychology, 20*(2), 245–252.

Kirkwood, M. W., Hargrave, D. D., & Kirk, J. W. (2011). The value of the WISC-IV Digit Span subtest in detecting noncredible performance during pediatric neuropsychological examinations. *Archives of Clinical Neuropsychology, 26*(5), 377–384.

Kirkwood, M. W., & Kirk, J. W. (2010). The base rate of suboptimal effort in a pediatric mild TBI sample: Performance on the Medical Symptom Validity Test. *The Clinical Neuropsychologist, 24*(5), 860–872.

Kirkwood, M. W., Kirk, J. W., Blaha, R. Z., & Wilson, P. (2010). Noncredible effort during pediatric neuropsychological exam: A case series and literature review. *Child Neuropsychology, 16*, 604–618.

Kirkwood, M. W., Peterson, R. L., Connery, A. K., Baker, D. A., & Grubenhoff, J. A. (2014). Postconcussive symptom exaggeration after pediatric mild traumatic brain injury. *Pediatrics, 133*(4), 643–650.

Kirkwood, M. W., Yeates, K. O., Randolph, C., & Kirk, J. W. (2012). The implications of symptom validity test failure for ability-based test performance in a pediatric sample. *Psychological Assessment, 24*(1), 36–45.

Larrabee, G. J. (2003). Detection of malingering using atypical performance patterns on standard neuropsychological tests. *The Clinical Neuropsychologist, 17*(3), 410–425.

Loughan, A. R., & Perna, R. B. (2014). Performance and specificity rates in the Test of Memory Malingering: An investigation into pediatric clinical populations. *Applied Neuropsychology: Child, 3*(1), 26–30.

Loughan, A. R., Perna, R. B., & Hertza, J. (2012). The value of the Wechsler Intelligence Scale for Children—Fourth Edition Digit Span as an embedded measure of effort: An investigation into children with dual diagnoses. *Archives of Clinical Neuropsychology, 27*(7), 716–724.

MacAllister, W. S., Nakhutina, L., Bender, H. A., Karantzoulis, S., & Carlson, C. (2009). Assessing effort during neuropsychological evaluation with the TOMM in children and adolescents with epilepsy. *Child Neuropsychology, 15*(6), 521–531.

McCrory, P., Meeuwisse, W. H., Aubry, M., Cantu, B., Dvorak, J., Echemendia, R. J., et al. (2013). Consensus statement on concussion in sport: The 4th International Conference on Concussion in Sport held in Zurich, November 2012. *British Journal of Sports Medicine, 47*(5), 250–258.

Perna, R. B., & Loughan, A. R. (2013). Children and the Test of Memory Malingering: Is one trial enough? *Child Neuropsychology, 19*(4), 438–447.

Perna, R. B., & Loughan, A. R. (2014). The influence of effort on neuropsychological performance in children: Is performance on the TOMM indicative of neuropsychological ability? *Applied Neuropsychology: Child, 3*(1), 31–37.

Perna, R. B., Loughan, A. R., Hertza, J., & Segraves, K. (2014). The value of embedded measures in detecting suboptimal effort in children: An investigation into the WISC-IV Digit Span and CMS Verbal Memory subtests. *Applied Neuropsychology: Child, 3*(1), 45–51.

Ploetz, D., Mazur-Mosiewicz, A., Kirkwood, M. W., Sherman, E. M. S., & Brooks, B. L. (in press). Performance on the Test of Memory Malingering in children with neurological conditions. *Child Neuropsychology*. Available at *www.tandfonline.com/doi/abs/10.1080/09297049.2014.986446#preview.*

Reynolds, C. R., & Kamphaus, R. W. (2004). *Behavior Assessment System for Children —Second Edition*. Circle Pines, MN: AGS.

Rutland-Brown, W., Langlois, J. A., Thomas, K. E., & Xi, Y. L. (2006). Incidence of traumatic brain injury in the United States, 2003. *Journal of Head Trauma Rehabilitation, 21*(6), 544–548.

Sherman, E. M., Wiebe, S., Fay-McClymont, T. B., Tellez-Zenteno, J., Metcalfe, A., et al. (2011). Neuropsychological outcomes after epilepsy surgery: Systematic review and pooled estimates. *Epilepsia, 52*(5), 857–869.

Sherman, E. M. S., & Brooks, B. L. (2015). *Child and Adolescent Memory Profile™*. Lutz, FL: Psychological Assessment Resources.

Sherman, E. M. S., & Brooks, B. L. (in press). *Memory Validity Profile™*. Lutz, FL: Psychological Assessment Resources.

Slick, D. J., & Sherman, E. M. S. (2012). Differential diagnosis of malingering and related clinical presentations. In E. M. S. Sherman & B. L. Brooks (Eds.), *Pediatric forensic neuropsychology* (pp. 113–135). New York: Oxford University Press.

Tombaugh, T. (1996). *Test of Memory of Malingering (TOMM)*. North Tonawanda, NY: Multi-Health Systems.

Welsh, A. J., Bender, H. A., Whitman, L. A., Vasserman, M., & MacAllister, W. S. (2012). Clinical utility of Reliable Digit Span in assessing effort in children and adolescents with epilepsy. *Archives of Clinical Neuropsychology, 27*(7), 735–741.

Westerveld, M. (2010). Childhood epilepsy. In K. O. Yeates, M. D. Ris, H. G. Taylor, & B. F. Pennington (Eds.), *Pediatric neuropsychology: Research, theory, and practice* (2nd ed., pp. 71–91). New York: Guilford Press.

Yeates, K. O., Ris, M. D., Taylor, H. G, & Pennington, B. F. (Eds.), *Pediatric neuropsychology: Research, theory, and practice* (2nd ed.). New York: Guilford Press.

12

Pediatric Sports-Related Concussion Evaluations

MARTIN L. ROHLING
JENNIFER LANGHINRICHSEN-ROHLING
MELISSA M. WOMBLE

Over the past 30 years, researchers in several fields of health care (e.g., neuropsychology, pediatrics, family medicine, athletic training, physical therapy) have increased their focus on sports-related concussions, specifically as they relate to children and adolescents (Kirkwood, Yeates, & Wilson, 2006). Initially, sports-related concussion researchers only examined student athletes at the college level (Barth et al., 1989). However, as the field has matured, there has been a broadened focus to include both professional athletes and younger and younger student athletes. In this chapter, we review the literature, which primarily involved adults, examining computerized assessments of sports concussion. We also review the reasons that validity testing is necessary in this setting, with specific attention paid to the influence of development on performance validity. Our review is supplemented by the presentation of findings from our own lab, where we have recently completed data collection that includes several hundred high school and college athletes. We present preliminary data analysis as well as our subjective experiences over the 3 years it took for these data to be collected.

INFLUENCE OF THE NATIONAL FOOTBALL LEAGUE LAWSUIT ON SPORTS-RELATED CONCUSSION IN YOUTH

In the field of sports-related concussion, researchers continue to be divided as to the neurocognitive health risks, both short- and long-term, of playing contact sports such as football (e.g., Karantzoulis & Randolph, 2013; Randolph, 2011, 2014; Randolph & Kirkwood, 2009). Examples of such concerns have often spilled into the courtroom. Perhaps the best example of this is the National Football League (NFL) lawsuit filed by more than 3,000 former players. Ex-players claimed that the league was liable for their assumed brain injuries. The former players' allegations hinged on the assumption that long-term neurocognitive deficits associated with repeated hits to the head had been substantiated by researchers (Farr, 2013). Specifically, players believe that they have suffered cognitive deficits and that such deficits were due to chronic traumatic encephalopathy, or CTE (Omalu et al., 2005). CTE is a freshly coined term that describes a form of progressive dementia that is considered similar to Alzheimer's disease and/or amyotrophic lateral sclerosis (ALS). However, CTE's existence remains controversial (Karantzoulis & Randolph, 2013), and there is a lack of consensus regarding the procedures necessary for making such a diagnosis. Furthermore, the etiology of CTE remains undetermined, and its unique neuropathology has yet to be well articulated.

In the suit, the retired players alleged that the NFL knew of the cognitive risks they faced but did not provide sufficient warning to protect them. The NFL recently agreed to an out-of-court settlement, which distributes funds to retired players based on the severity of their neurocognitive deficits. Many observers of the negotiated settlement have falsely concluded that the NFL has validated players' concerns with respect to long-term cognitive disability caused by CTE in their settlement agreement (Mihoces, 2013). However, the settlement did not include any admission of guilt on the part of the league. In fact, the NFL did not formally endorse the theory of CTE as a consequence of players' history of concussions.

The NFL lawsuit and eventual settlement have increased the scrutiny of various younger amateur athletes. Parents fear that their children are at risk for long-term brain damage caused by even the most minor of blows to the head (Lavigne, 2012). In fact, the NFL, in collaboration with the National Institutes of Health (NIH), has committed over $30 million to research sports-related concussion, with a substantial percentage of that money going to investigators studying sports-related concussion in children and teenagers. Consequently, the demand has quickly increased for computerized neurocognitive testing (CNT) programs to aid researchers, doctors, athletic trainers, coaches, and parents in making return-to-play decisions (Resch, McCrea, & Cullum, 2013). These decisions are best made ipsatively, which means that athletes' cognitive performance during a post-concussion assessment is compared with their performance during baseline testing conducted preseason (Jinguji et al., 2012; McCrory et al. 2013; Moser, Schatz, Neidzwski, & Ott, 2011). An ipsative analysis controls for individual differences in athletes' performance that existed prior to their suffering a concussion. After a concussion, older teen and adult athletes recover relatively quickly, with methodologically rigorous studies failing to identify performance-based decline after about 10–14 days postinjury (Belanger & Vanderploeg, 2005; McCrea et al., 2003), whereas younger athletes may take a bit longer to recover (McCrory et al., 2013). Thus return-to-play decisions obviously require that reliable and valid scores be obtained during baseline testing and that checks of functioning occur across time.

Parents' Concerns and Medical–Legal Involvement

Parents have raised concerns about the risks of concussion to younger athletes. Their concerns are bolstered by research indicating that younger athletes may be at greater risk of sustaining a concussion and may require more time to recover from one (Harmon et al., 2013). However, it is likely that parents' concerns have been inflated by exaggerated accounts of sports-related concussion reported in the popular media (Zuspan, 2013). Regardless of the empirical evidence, parents listening to such reports have increased their demands for action on the part of their local school district, athletic coaches, and park district directors.

Some parents have gone as far as to turn to the criminal justice system for answers, as happened with Peters Township High School in eastern Pennsylvania (Crompton, 2012). Peters Township employed the former Pittsburgh Penguins' athletic trainer, Mark Mortland, who is a 25-year veteran of athletic training, with 12 years of experience at the professional level. Mortland alleged that the head football coach for Peters Township, Rich Piccinni, had engaged in "child endangerment" by pressuring or allowing players to return to play prematurely after a sports-related concussion before Mortland had cleared them to play (Crompton, 2012). The local chief of police and the nine-member school board closed the case without filing charges against Piccinni, despite parents' demands.

Such incidents have driven public officials to seek more information from researchers related to the effects of concussion on young athletes, which includes elementary and middle school, in addition to high school, athletes. A recent ESPN report (Keating, 2012) noted that there are 1,700 NFL players, 66,000 college athletes, 1,100,000 high school athletes, and 250,000 middle school athletes playing football. Clearly, the prevalence of concussions will be higher in high school and middle school athletes, as compared with professional athletes, merely due to the number of individuals involved at the lower levels of play.

Evidence of parents' concern was further documented in another ESPN story (Lavigne, 2012), which reported results from an Internet survey that found that 57% of parents believed that "recent stories about the increase in concussions in football have made them less likely to permit their sons to play in youth leagues" (para. 2) and about two-thirds of parents with children under the age of 15 agree that concussions were a significant problem in youth football (Lavigne, 2012). Finally, 94% of parents agreed that concussions are a serious problem for the NFL.

RATIONALE FOR VALIDITY TESTING WITHIN SETTING

Computerized Neurocognitive Test Batteries

Despite this controversy, clearly some brain injuries do occur during athletic events that include children and adolescents. The focus on concussion and the use of CNT software at baseline and postconcussion assessment is intended to assist the clinician in providing the best services possible to young athletes. CNT and the user must be able to determine whether the resulting scores are valid, so that an ipsative analysis can be properly conducted. If athletes provide invalid scores, either at baseline or postconcussion, the process breaks down.

As the amount of money devoted to sports-related concussion increases, the allocation of that money has progressively gone more toward studies focused on children, gender differences, and sport-specific differences. However, prematurely, there has been a concomitant increase in demand for CNT for clinical purposes, as there remains uncertainty as to whether CNT actually modifies the risks of concussion faced by athletes (e.g., see Randolph, 2011; Kirkwood, Randolph, & Yeates, 2012). Those that include measures that control for poor effort or invalid performance are in higher demand. CNTs are commonly marketed to school systems and athletic associations, as well as parents. The predominant product in the market is ImPACT (Immediate Postconcussion and Cognitive Testing; Lovell, 2007, 2013). The company reportedly has sold its CNT software to over 7,000 professional teams, colleges, and high school athletic programs. ESPN's *Outside the Lines* (Keating, 2012) reported that there have been 3 million ImPACT tests given to high school and middle school athletes to date. Furthermore, several other companies are now competing for a share of the ImPACT market. For example, there is the Automated Neuropsychological Assessment Metrics (ANAM, 2010; Cernich, Reeves, Sun, & Bleiberg, 2007; Reeves, Winter, Bleiberg, & Kane, 2007), the Cognitive Drug Research computerized cognitive test system (CDR System; Keith, Stanislav, & Wesnes, 1998), CogState (1999), CogSport (Collie et al., 2007; renamed by Axon Sports to the Computerized Cognitive Assessment Tool, or CCAT), Concussion Vital Signs (CVS; Gualtieri & Johnson, 2008), and HeadMinder's Concussion Resolution Index (CRI; Erlanger, 2002), to name a few. There are also newer paper-and-pencil instruments, which include the freely available Sport Concussion Assessment Tool 3 (SCAT3; Concussion in Sports Group, 2013), and, for younger athletes, the Child-SCAT3 (Concussion in Sports Group, 2013). Although not computerized assessment batteries in the strictest sense of the definition, they are available for cell phone and computer tablets for ease of data entry and analysis. Most of these CNT batteries are reviewed in greater depth by Rahman-Filipiak and Woodard (2013), as well as by Resch and colleagues (2013).

Invalid Performance: “Sandbagging”

Despite the popular media’s focus on sports-related concussion, many athletes remain skeptical about the alleged risks. Evidence of this skepticism was expressed by a celebrated NFL quarterback, Peyton Manning. Manning admitted to scoring more poorly than he was capable of on his baseline tests (Pennington, 2014). This has been referred to as “sandbagging.” Athletes engage in such strategies to make it less likely that their postconcussion scores will fall below their baseline scores. This provides them with an excuse to return to play sooner than might otherwise be expected by medical staff. Soon after Manning’s admission, *NBC Sports* covered the issue in a segment of its *ProFootball Talk* (Smith, 2011). In the published report, players described how they went about “beating the system.” Sandbagging was a common strategy, and younger athletes were picking up on the strategy as well. Oddly, the journalists who brought attention to this issue at the high school level were some football players’ classmates who worked for *The PBS News Hour: Student Reporting Labs* (Tiedens, 2014). These probing reporters interviewed players at their own high schools. These players readily confessed to sandbagging without hesitation. Apparently, they did not think that they were putting themselves at much, if any, risk by so doing. Hence it is apparent that researchers and clinicians need to employ methods for the detection of invalid performance at both baseline and postconcussion assessments. Such detection methods are necessary to ensure that athletes who suffered a concussion are not returned to play before they are fully recovered.

PERFORMANCE VALIDITY TESTING

Invalid performance on neuropsychological tests has become a common concern among clinicians, with several books devoted to the issue (e.g., Larrabee, 2007; Boone, 2007; Carone & Bush, 2012). In fact, a recently published volume on this topic (Carone & Bush, 2012) devoted an entire chapter to performance validity testing (PVT) for sports-related concussion assessments with adults (Macciocchi & Broglio, 2012). Professional associations within the field of clinical neuropsychology have agreed that the standard of practice requires the use of PVTs when conducting clinical assessments, as well as when conducting research (e.g., Heilbronner et al., 2009). Thus it is expected that such measures will be included when clinicians are involved in conducting any CNT for professional, amateur, or college athletes, as well as high school and middle school athletes. Strong evidence of the importance of such assessments was documented by Green, Rohling, Lees-Haley, and Allen (2001), who found that "poor effort," or invalid performance, accounted for more variance in test scores than did the severity of the participant's head injury.

STATE OF SCIENCE OF PERFORMANCE VALIDITY TESTING

Basic Psychometrics: Reliability and Validity

As already mentioned, for competent service delivery, neuropsychologists must be able to detect invalid performance on CNT. Invalid test scores can be attributed to a variety of things, some of which are intentional or planned whereas others are unintentional or unplanned. Examples of the latter include aspects of the testing environment, examiners' errors in administration and interpretation, and/or the psychometric properties (i.e., reliability and validity) of the CNT selected for administration. We address the more challenging aspects of detecting intentional or planned invalid performance first. Then, some of the unplanned or careless causes of invalid test scores are described.

A frequently repeated tenet of psychometrics is worth noting here. Specifically, a test's "validity is bounded by its reliability." Thus a test cannot be more valid than it is reliable. However, reliability does not guarantee that a test is valid for the purpose that is intended. Nevertheless, factors that reduce a test's reliability will likely cause a concomitant reduction in its validity. Consequently, invalid performance results from anything that influences a test's consistency of measurement.

Invalid test scores are, by definition, evidence of "inconsistent" performance by examinees. Because examinees may willfully produce invalid scores, they might change how much effort they exert from moment to moment, depending upon their willingness to fully engage in the task at hand. Thus performance and effort are orthogonal to one another, as shown in Figure 12.1. Specifically, examinees might be able to perform to the best of their ability while exerting high effort (e.g., effortful encoding or explicit memory). They might also put forth low effort but perform to the best of their ability because of the nature of the task (e.g., automatic processing or implicit memory). Alternatively, on some tasks they might exert low effort and perform well below their ability for reasons that are outside of their control (e.g., inattention, carelessness, excessive noise, overheated rooms, or poor quality of equipment). Finally, examinees might put forth high effort and perform well below what they are capable of (i.e., feigning or malingering). Of these four options, examiners would prefer that examinees put forth good effort and perform to the

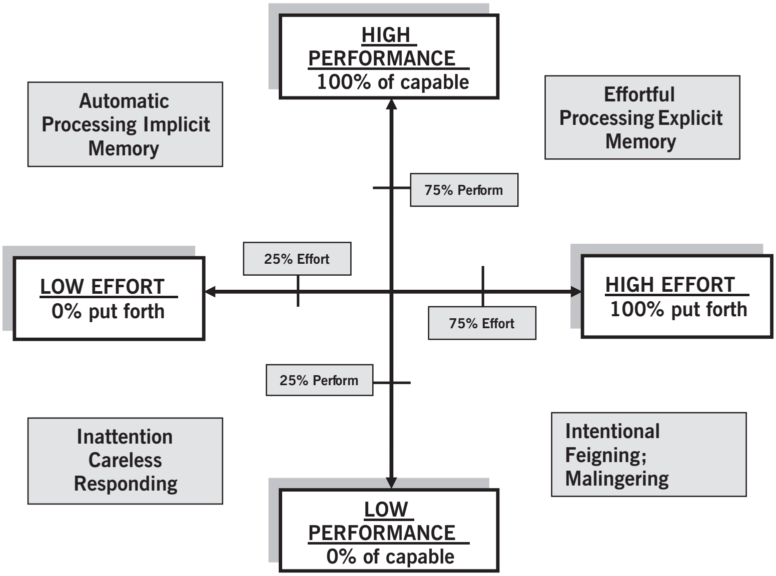

FIGURE 12.1. The orthogonal relationship between effort or intention and performance.

best of their ability. Only this combination will provide examiners with valid scores for competent analysis and good clinical decisions. The three other options cause examiners trouble, especially when it comes to making return-to-play decisions.

Another way of explaining this is that examiners are interested in learning an examinee's "true score variance," while simultaneously reducing "error variance." Some variables that create error variance include prior experience with the test (i.e., practice effects), emotional status, physical status, and preexisting cognitive disorders. Additional variables that can cause reductions in the validity are characteristics of the testing environment (Rahman-Filipiak & Woodard, 2013). For example, the number of examinees in the room, the number of proctors, the noise level, the quality of computers, the institutional support for the evaluation (e.g., coaching staff, school district administrators, athletic trainers), and the quality of instructions provided to examinees can all decrease validity.

Normative Issues for CNTs and PVTs

Reviewing specific PVTs, symptom validity tests (SVTs), or CNTs is beyond the scope of this chapter. However, interested readers are referred to books by Boone (2007) and Larrabee (2007) for comprehensive coverage of validity testing in adults. For coverage of validity issues related to children and adolescents, we recommend DeRight and Carone (2013), Donders and Kirkwood (2012), and Kirkwood (Chapter 5, this volume). Coverage of issues related to mild traumatic brain injury (mild TBI) or concussion in children and adolescents can be found in Kirkwood et al. (2012), Covassin, Elbin, Kontos, and Larson (2010), as well as Covassin, Elbin, Larson, and Kontos (2012).

As mentioned earlier, most professional neuropsychological associations have concluded that the use of PVTs is a standard of practice (Bush et al., 2005). PVTs were previously referred to as symptom validity tests, or SVTs (Larrabee, 2012). However, the label SVT has recently been suggested for use in only those tests used to measure self-reported symptoms, which cannot be objectively verified, as can a sign. Fever is a sign, whereas headache is a symptom. Other psychological symptoms are measures of mood, pain, and self-assessed cognition. This chapter is restricted, for the most part, to the use of PVTs in the area of sports-related concussion in children and adolescents. However, SVTs may soon be as important in this area as PVTs are now.

There are several freestanding PVTs on the market and many other PVTs that are referred to as “embedded” or as “algorithms” that are designed to detect unusual or rare patterns of scores obtained on an instrument. The problem with freestanding PVTs in the case of sports-related concussion assessments is that they add time and expense to the assessment session. Embedded measures add no additional time to the assessment and also no additional cost. However, freestanding measures have generally been shown to be more sensitive and specific to invalid performance than are embedded measures. Results from any single PVT, embedded or freestanding, are not likely to be sufficient for accurately detecting invalid performance. Multiple measures are often necessary, which can mean that embedded and freestanding tests need to be used in combination with one another (Larrabee, 2007, 2014).

A significant problem for clinicians and researchers involved in the area of sports-related concussion is that the most common CNTs and PVTs typically have published norms only for adults (i.e., age $\geq$ 18). Even those that have age-based norms for individuals younger than 18 rarely have norms for children below the age of 14. This limitation makes their use with middle and elementary school-age athletes more challenging. In addition, CNTs often do not have age-based norms that discriminate between those in their 20s and those in their 30s or 40s. Similarly, most of the freestanding PVTs have been designed to be used with adults, and their use with children has challenges. PVTs typically provide no age-based norms, as they are often considered to be so easy that age would not be a factor in successfully obtaining scores above published cutoffs. This is likely true for those children older than 11 years of age. However, the ability of an instrument to aid a clinician in detecting invalid performance can change dramatically when the examinee is 10 years old or younger. Furthermore, embedded algorithms are often not age-based, and these too will generate more false positives as the age of the examinees gets younger. Therefore, clinicians and researchers alike are cautioned when using the available instruments with children younger than 10, as there is limited research on how uninjured children perform on CNTs and PVTs, let alone athletes with concussions.

Literature Review of PVTs in Sports Settings

Just as most of the instruments have been designed for adults and have only limited norms for adolescents, most of the research on PVTs in sports settings has been with adults. Therefore, one has to generalize the findings with adults to children, which may not be entirely accurate. Worthy of note in this discussion are the results of a recent study by Babikian, McArthur, and Asarnow (2013), who assessed children from 8 to 17 years old ($n = 85$; $M = 12.7$; $SD = 2.0$). These researchers reported that

> None of the injury severity indicators or type of injury (head vs. other body part) predicted either 1-month or 12-month cognitive impairment status. Rather, premorbid variables that antedated the injury (parental education, premorbid behavior, and/or learning problems, and school achievement) predicted cognitive impairments. In summary, the best predictor of 1-month impairment classification was school achievement, followed by parent education and premorbid behavioral and academic problems, while the best predictor of 12-month impairment classification was 1-month impairment classification. (p. 152)

On the basis of this and other research, clinicians should expect that the vast majority of athletes who suffer a sports-related concussion will return to baseline reasonably quickly (Hung et al., 2014). Extended lengths of time to recover are often associated with premorbid cognitive dysfunction, psychosocial stressors, or preexisting emotional and/or behavioral disorders.

Rohling recalls attending a workshop over a decade ago that addressed the need for baseline testing of athletes across all levels of sports (Hardey et al., 2001). After the panelists completed their presentations, the audience was encouraged to ask questions. Rohling inquired as to the need to include performance validity testing to ensure that valid scores were obtained. A presenter responded by stating that athletes are a highly motivated and competitive lot. There was no need to include such measures in the battery. Furthermore, if they were included, they would take up too much time, time that could be better spent on cognitive tasks and symptom reporting that provided more information of greater value to the clinician.

With assumptions of this nature, it took time before anyone conducted an empirical study in this area. The first study that we know of that addressed this topic was published by Hunt, Ferrara, Miller, and Macciocchi (2007) 7 years later. These authors found, using some rather insensitive PVTs (i.e., the Rey 15-Item Test and the Dot Counting Test), that 11.1% of the sample failed at least one of these PVTs and that this sample had a standardized mean difference (SMD) effect size of –0.70. These results are presented in Figure 12.2, which shows the effect of passing versus failing these PVTs (see first panel, first bar, marked *P & P* for use of paper-and-pencil measures examining high school student athletes). For comparison purposes, Figure 12.2 includes data in the seventh panel showing the immediate effects of mild TBI (i.e., 1–7 days postinjury, per Rohling et al., 2011) and the residual effects of severe TBI (i.e., 12 months postinjury, per Rohling, Meyers, & Millis, 2003).

In a subsequent simulation study, Erdal (2012) conducted an archival review of ImPACT assessments of college athletes ($n = 269$). Erdal reported that 12.0% of the sample provided invalid scores by failing two or more of the embedded PVTs. These data are illustrated in Figure 12.2 as well, showing the effect of passing versus failing the embedded PVT of the ImPACT battery (see third panel, third bar, marked *ImP,* which included college student athletes). This number is consistent with the results obtained by Hunt et al. (2007) shown in Figure 12.2 (see first panel, first bar). When examining athletes across multiple sessions who failed any two indicators across sessions, 25.0% of this college sample failed the embedded PVT algorithm. An analogue study by Schatz and Glatts (2013) found that of the undergraduate volunteers who were instructed to feign but encouraged not get caught, 25.0–40.0% of them escaped detection by the embedded algorithm for the ImPACT battery. Applying these estimates to the reported 12.0% of student athletes who performed invalidly, the percentage might actually be 15.0–17.0%, with a significant number of student athletes evading detection.

Szabo, Alosco, Fedor, and Gunstad (2013) analyzed data gathered with the ImPACT battery from three seasons of assessments of 159 unique student athletes at the college level. Of these, 17.9% of assessments resulted in two or more invalid scores on a newly developed embedded PVT. Furthermore, 21.4% of athletes had at least one invalid PVT score. For those student athletes who took all three baseline assessments, 28.0% of them had at least one session for which they obtained at least one invalid PVT score. Looking across all three sessions and requiring that at least two invalid scores be produced to consider the entire battery to be invalid revealed 25.0% of the sample as providing invalid test scores. These data are presented in Figure 12.2, which shows the effect of passing versus failing the

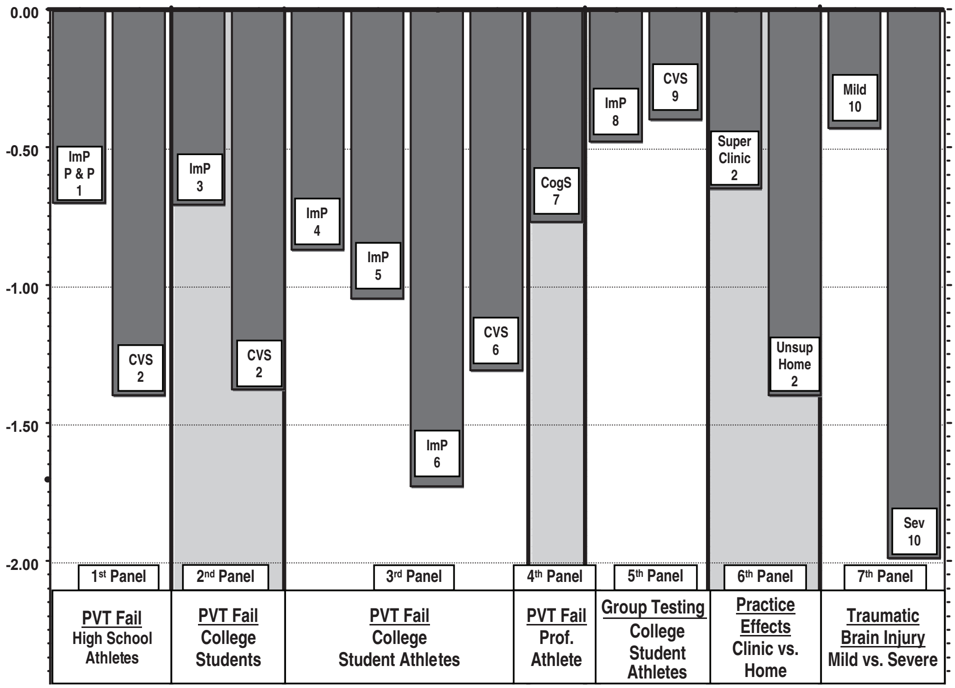

FIGURE 12.2. Standardized mean difference effect sizes (Cohen's *d*) for the various variables examined that influence an examinee's obtained test scores at baseline. PVT, performance validity tests; P & P, paper-and-pencil PVTs administered; ImP, ImPACT (a computerized test battery); CVS, Concussion Vital Signs (a computerized test battery); CogS, Cog Sport (a computerized test battery); Sev, severe traumatic brain injury; Super., supervised testing completed in a clinic or professional setting; Unsup., unsupervised testing (a computerized test battery completed at home). Data from (1) Hunt et al. (2007); (2) Rohling (2011); (3) Erdal (2012); (4) Szabo et al. (2013); (5) Schatz & Glatts (2013); (6) Hill et al. (2014); (7) Darby et al. (2011); (8) Moser et al. (2011); (9) Womble et al. (2012), and (10) Rohling et al. (2003).

embedded PVT of the ImPACT battery (see third panel, first bar, marked *ImP*, which included college student athletes).

Hill, Womble, and Rohling (2015), using a newly developed embedded algorithm, found that 12.9% of college and high school student athletes failed two or more PVTs, whereas 31.6% failed at least one PVT. Furthermore, the effect of failing two or more PVTs on the overall test battery mean, a composite score of all the measures given, equaled an effect size of –1.20. Lower baseline scores make it more difficult for clinicians to detect continued cognitive deficits postconcussion in these athletes. These data are presented in Figure 12.2 showing the effect of passing versus failing the embedded PVT of the CVS battery (see third panel, fourth bar, marked CVS, which included college student athletes).

Darby et al. (2011) analyzed CogSport data collected from 2002 to 2010 from over 17,368 professional athlete assessments across four countries and continents. They utilized a newly developed validity algorithm that included seven separate measures, and reported that 14.1% of athletes provided invalid test scores at baseline. Their findings are shown in Figure 12.2, which shows the effect of passing versus failing the embedded PVT of the CogSport battery (see fourth panel, first bar, marked *CogS,* which included professional athletes). Their results are consistent with those reported by Hunt et al. (2007) of 11.1%, by Schatz and Glatts (2013) of 12.0%, by Erdal of 10.6%., by Szabo et al. (2013) of 17.9%, and by Hill et al. (2015) of 12.9%. Averaging across these five samples, one finds that 13.0% of student athletes failed two or more PVTs in these samples. Using a more liberal criteria to determine the invalidity (i.e., just one PVT) results in 26.0% of these same samples providing invalid test scores. In the Hill et al., (2015) sample, this equated to an effect size of poor effort of –1.20. For example, if an examinee had an overall test battery mean (OTBM) of 100 under good effort, his or her invalid OTBM would equal an 82.

Group Size during Baseline Testing

To make the baseline testing economically feasible, CNT researchers have often recommended that assessments be given in a group setting (Lovell, 2007, 2013). However, group testing may be negatively influenced by the size of the group being tested and the nearness of the various examinees to one another during testing. Moser and colleagues (2011) found that baseline testing of high school athletes in a group format resulted in statistically significant reductions in test scores, with an effect size of –0.48. Although this was not the focus of their study, Moser et al. (2011) reported that of the 15 athletes whose protocols were flagged by an embedded PVT algorithm, 14 of these protocols were from athletes tested in a group setting (8.5%), but only one of these came from an athlete tested in an individual setting (0.5%). The finding suggests that it is 17 times more likely that invalid baseline scores will be obtained in a group setting than in an individual setting. See Figure 12.2 for an illustration of the effect of taking the ImPACT battery individually versus in a group setting (see fifth panel, first bar, marked *ImP,* which included college student athletes).

Comparing our own data (Womble, Rohling, Hill, & Corley, 2012) with those obtained by Moser et al. (2011) revealed similar results. In our own lab, Womble et al. (2012) used the newly developed embedded PVT algorithm to analyze scores from student athletes in middle school, high school, and college with the CVS. There was a significant effect (p = .0005) of group size on the OTBM. When three or fewer athletes were tested in a session the mean OTBM was 96.3. But, when seven or more athletes were tested in a session the OTBM was 90.5 (effect size of –0.40). Additional analyses found 7.5% of athletes provided invalid data in the small groups, whereas 14.4% provided invalid data in the larger groups. It appears from these data that the more examinees there are in a room during baseline testing, the more likely it is that invalid scores will be obtained, and the lower will be the scores for all of the athletes tested in the room. These data are shown in Figure 12.2 (see fifth panel, second bar, labeled *CVS*).

Supervised versus Unsupervised Assessments

Another issue of concern is whether the CNT battery given is proctored by supervisors who are granted some level of authority within the setting. In a study presented by Rohling (2011), which examined CNS Vital Signs (Gualtieri & Johnson, 2008), practice effects on the battery were examined using "in-home" practice sessions versus practice sessions in a supervised setting. Those who were included in the in-home group obtained an OTBM of 81.6 (*SD* = 27.1) versus an OTBM of 103.3 (*SD* = 9.7) for those who took the same practice sessions under supervision. Once again, these data are shown in Figure 12.2 (see sixth panel, both bars, for the effect of supervised vs. unsupervised practice effects, respectively). This resulted in an effect size for taking the test at home equal to –1.40. There was also a difference in the intraindividual variability across subtest scores, with greater variability occurring for the in-home practice group (*SD* = 13.4) as compared with the supervised setting (*SD* = 7.5). Overall, the effect size for intraindividual variability of taking the test at home versus under supervision was –1.70, which was highly significant ($p < .0001$). Finally, the rate of invalid performance of the in-home group was 28.0% compared with 10.0% in the supervised sample, which was also significant ($p < .005$). Summarizing, taking a CNT at home in this adult sample resulted in lower overall scores, greater variability in scores, higher invalid performance rates, and longer lengths of time to complete the battery (47.1 minutes compared with 28.7 minutes, $p < .0001$). Although these results are preliminary and have not been replicated with middle or high school athletes, the evidence suggests that we should not expect valid baseline scores to be obtained when athletes are given a CNT at home rather than in a supervised setting.

Baseline Testing versus Postconcussion Assessments

Another concern that has been raised when employing PVTs in sports-related settings is that they might falsely identify an athlete postconcussion as engaging in invalid responding when in reality he or she is truly cognitively impaired due to a concussion. In the data collected in our lab, again looking at the scores from the CVS, overall, 14.2% of examinees were identified as invalid responders, whereas 15.5% of examinees postconcussion were identified as invalid responders, not significantly different from one another. Nevertheless, the effect size for invalid responses for baseline examinees was just –1.39, as compared with those who were examined postconcussion, which was –2.03. Oddly, there was an interaction effect in these analyses, with valid responders at postconcussion obtaining an OTBM of 99.8, which was significantly higher than valid responders at baseline, who obtained an OTBM of 93.2. However, when considering examinees who had provided invalid scores, those who were examined postconcussion earned lower average scores at 71.8 (15.5% of athletes postconcussion) compared with those who were uninjured at baseline, whose average score was 74.5 (14.2% of baseline assessments). These scores suggest that there was a practice effect (0.45) evident in the valid responders postconcussion that was not evident in the invalid responders postconcussion (–0.20). These analyses also indicate that the criteria used to determine whether invalid performance occurred is accurate and useful for both baseline and postconcussion assessments and that PVTs should be used during each and every assessment.

ASPECTS OF SETTING

Quality and Availability of Computers

Although we have no quantitative data related to the quality and availability of the computers used for data collection, our subjective experiences while supervising baseline and postconcussion sessions are of some import. Most of our experience is supported by the conclusions of Rahman-Filipiak and Woodard (2013), who recently reviewed the literature on these topics. While monitoring athletes during CNT sessions, it was clear that the quality of equipment had a significant effect on the scores examinees obtained. There was tremendous variability in the quality of computers that were available at the schools for testing, as well as the availability of rooms and the times the institutions would allow for testing. Computer quality varied substantially across schools, even within the same public school district, for which data were collected (13 high schools total). School administrators seemed unaware of this variability and often did not ask support staff to assist the sports-related concussion researchers in data collection. Therefore, research assistants were left to download updates to general computer programs, download new programs that were missing, and clean keyboards and mice that were too gummed up to work properly. We were commonly not given advance access to rooms or computers to prepare the equipment prior to testing, and the data collected suffered as a result. Likewise, we lacked the necessary passwords to unlock computers or to access new or updated downloads, or the computers had software installed that prohibited users from downloading programs. Such information needed to be obtained from a "central office," and tracking down the right person at the necessary time was often difficult. Typically, there were few, if any, personnel available to assist in data collection, and some results were lost because some athletes who did not have proctors in their rooms would leave the facility. In summary, at least in some cases, sports-concussion clinicians should anticipate limited assistance from school staff and the designation of faulty equipment for data collection. Such issues are more common in schools with students from lower income households, which makes data collection at baseline more challenging. Furthermore, student athletes at these schools are the individuals who are most in need of such services, because they are the least likely to obtain quality health services related to their sports activity.

These difficulties were not evident in the college sample, where trainers and academic tutors provided exceptional assistance in getting equipment ready. They typically maintained high-quality computers that were made available for CNT administration. The only problem encountered in collecting data from the college sample was that student athletes were typically expected to take baseline tests in the early morning, sometimes as early as 6:30 A.M. Coaches and trainers requested early testing so as to prevent student athletes from missing practice or required classes.

Testing Environment

Room characteristics are another potential influence on test scores. To date, we know of no quantitative data related to this topic, but our qualitative experience suggests that there is a large negative effect on the baseline data collection. Once again, Rahman-Filipiak and Woodard (2013), as well as Moser et al. (2011), reported similar concerns and made similar recommendations regarding this challenge. Specifically, research assistants were often asked to collect data in noisy rooms that either were being remodeled, having their floors waxed, or were too close to outside noise (e.g., railroad tracks with frequent trains passing by). We were even asked to collect data in locker rooms that were damp, dirty, smelly, and with limited ventilation. We were required to bring our own computers into such settings for data collection or it would have been impossible to complete baseline testing. Many of these facilities were not easily accessed by those examinees of the opposite gender. Furthermore, a host of individuals would typically be coming and going during baseline CNT, including teachers, coaches, other student athletes, janitors, and parents. It was common for examinees to be interrupted during testing by individuals asking them unrelated questions, such as permission to borrow a cell phone or eat a snack, or to determine whether a particular school assignment had been completed on the computer. It was rare for those who were interrupting athletes during testing to ask permission beforehand; they often seemed unaware or unconcerned with the potential consequences of their actions. It was also common for baseline sessions to be postponed or canceled due to competing activities. We once had an entire team of examinees (n = 110) cancel the baseline assessment on the day of data collection because a photographer had shown up unexpectedly to take pictures for the football programs and yearbook. In fact, 40% of the schools for which we provided monitoring did not complete baseline data collection until after the season had begun, meaning that it was no longer baseline testing. In addition, 15% of schools did not complete baseline testing until half of the athletic season had been completed.

Again, the college sample was less challenging on these dimensions than were the high school samples. The college sample often had rooms made available that were well insulated to provide noise reduction, that were properly heated or air conditioned with good ventilation, and that had staff provided to keep unnecessary distractions from outside to a minimum by denying access of other student athletes to computer labs during baseline testing.

Assessment Proctors

With respect to the use of room proctors or test administration monitors, we again have no quantitative data, but qualitatively there was variation in the types of proctors used. Athletic trainers and head coaches tended to be the most helpful in proctoring tests, and assistant coaches or high school student volunteers seemed to be the least helpful, frequently making loud comments about an athlete's poor performance or being critical of how well the group was doing compared with the cohort that was tested just minutes before them. Some of these individuals also engaged in laughter, eating and drinking snacks, and making phone calls during testing sessions. High schools did not provide us with additional staff who were not part of the athletic department to assist in data collection.

Coaches, Trainers, Players, and Parents: Reporting Issues

It was common for coaches and athletic trainers to interfere with test administration. As noted before, many would make demands on examiners to assist student athletes in obtaining scores that might make return to play easier. Some also engaged in activities designed to push the interpretive envelope and allow an athlete to earn a score that was sufficient for them to return to play. It was also common for coaches and players to "split" staff; that is, when the neuropsychologist indicated that a player had not adequately recovered to allow him or her to return to play, coaches, players, and/or parents would ask trainers and physicians to clear the player to play without notifying the responsible neuropsychologist. This happened at both the high school and college levels. It also seemed likely that many concussions were not reported in order to avoid return-to-play assessments by the neuropsychologist. Using a base rate of 5.7% of high school and college football players suffering concussions in a single season (Guskiewicz, Weaver, Padua, & Garrett, 2000) and 2,950 baseline tests, we expected 168 concussions to occur. However, we had just 98 concussions reported (3.3%). In fact, two schools that accounted for 1,109 baseline tests reported 49 concussions, for a reporting rate of 4.4%. This left a rate of reporting for the remaining school of just 2.7%. Even more remarkable was that there were five schools in which 524 baseline tests had been completed; yet no concussions were reported to have occurred across an entire season. These "nonreporting" schools might have been exceptionally talented in preventing concussions, but a more likely explanation is that they failed to report concussions that occurred. Such behavior was evident at the college level as well, as neuropsychological staff witnessed concussions during games that were not reported, and student athletes seen during the off season for academic assessments (i.e., learning disabilities or attention-deficit/hyperactivity disorder) occasionally reported suffering concussions that the coaching and/or athletic trainers did not know about or failed to report to the neuropsychological staff. Similar findings were reported by McCrea, Hammeke, Olsen, Leo, and Guskiewicz (2004), who found that 15.3% of high school football players reported suffering a concussion during the season or preseason, while only 47.3% of these concussions were reported to athletic trainers and/or coaches. Thus, as the Institute of Medicine's report noted (2013), it is not easy to change the culture of sports in some regions of the country.

Authority within an Institution

The hierarchy of authority or personnel structure of the institutions in which the sports-related concussion assessments are to occur will significantly influence the validity of the scores obtained. The sports-related concussion consultant must be granted authority within the institution for conducting baseline testing, as well as postconcussion assessments, and to make return-to-play decisions. Undermining of the consultant's authority will certainly increase the frequency and the magnitude of invalid scores obtained. Providing concussion management services to a school system in which the consultant is not an employee limits the consultant's ability to adequately provide services. For example, the neuropsychologist may not be provided with the necessary time, facilities, and overarching authority to provide the clinical services required for valid baselines, or the recommendations the neuropsychologist provides to the system regarding student athletes' health may be ignored. Student athletes can be unaware or ill informed about these issues. When student athletes witness consultants being denied resources or observe clinical recommendations being ignored, the rate of invalid responding can be expected to rise. The student athletes will only take the experience as seriously as do their coaches, school administrators, peers, and parents. In lower income schools, in which athletics are often seen as one of the few ways out of poverty, all of these forces work against the neuropsychologist consultant. Parents, coaches, administrators, and players may be more invested in winning than they are in the health and welfare of student athletes. Coaches in particular often have contract bonuses tied to their teams' success or know that contracts may not be renewed if success is not achieved rapidly enough. Thus maintaining a credible consultation service is difficult, if not at times impossible, unless the consultant is highly motivated or granted outside funding and genuine authority to complete the task (e.g., from departments of education, elected school district officials, superintendents, or college athletic directors or presidents). Without such support, the validity of baseline test scores and the ability of the consultant to assist an athlete who has suffered a concussion to fully recover are limited.

RECOMMENDATIONS

Moser et al. (2011) gave several recommendations worth highlighting here again to improve the quality of baseline assessments. First, test proctors should provide the athletes with a description of the purpose of the evaluation and emphasize the health protective nature of the examination. Second, athletes should be provided with clear instructions for quality test administration, and the proctor should check to be sure that they understood what they were told. Third, athletes should be strongly encouraged to exert their best efforts, and invalid performance checks on the results should be included to make sure proper baseline data has been collected. Fourth, disruptions in the test environment must be minimized. For example, school personnel, peers, and/or electronic devices should not be allowed in the testing room once testing has begun. Fifth, the number of examinees in a room should be kept low (≤ 5) to minimize distractions to all who are in the room. Other recommendations for maintaining a good atmosphere for assessment include providing seating that is comfortable and separating chairs by enough space so that those who are sitting in them cannot distract one another. Ideally, an examinee should not be seated directly across from or next to another examinee. Extraneous sounds and interruptions should be removed, possibly through the use of white noise machines to help examinees remain focused on the CNT. The room's lighting should not cause glare on the computer screen, nor should sun coming in windows be allowed to diminish the quality of the stimuli shown on the screen. Computers, keyboards, and mice should be cleaned and working properly before testing begins. Software should be properly loaded before the athletes arrive for testing. Each room in which testing is being conducted should have multiple proctors who remain present and attentive throughout the assessment session. Examinees should be asked to take a bathroom break before testing begins and not again until all of the testing is completed. Examinees who talk too much, are disruptive, or are not devoting their full attention to the CNT should be excluded. These individuals can all be asked to return at a later time to complete their assessments individually. Individuals should not be assessed when they are exhausted after practice, as this will likely increase the chances that they will respond carelessly and produce invalid data. Athletes should also not be tested too early in the day, too late in the evening, just after academically demanding activities (e.g., a midterm examination), and/or when they are hungry. Examinees should be encouraged to work to the best of their ability so that they do not have to take the test over. Following testing, the assessment team leader, who preferably is a clinical neuropsychologist, should check all examinees' scores for invalid data. This needs to be done both at baseline and following postconcussion assessments.

FINAL THOUGHTS

Most graduate training programs for health service providers are being required to deliver culturally relevant information in their curricula. Multicultural training has become the norm. Trainees are to become "culturally competent" in their service delivery skills, being able to appreciate and navigate issues of gender, race, ethnicity, socioeconomic status, and age. The field of sports concussion management requires practitioners to develop and use these skills on a daily basis. There are obvious health disparities in sports associated with the demographic variables of race, gender, education, and income. Navigating from low-income, predominantly male, African American football players from urban schools to female-oriented, predominantly white, softball or volleyball players from affluent suburban schools can be challenging for the practitioner. Ensuring that the athletes are adequately instructed and properly motivated is very difficult for some health care providers. Adequately interpreting scores with statistical awareness of the influences of culture on the data is equally, if not more, challenging. To then try to help rehabilitate an athlete who has, for whatever reason, had his or her symptoms persist beyond the expected time of recovery takes a tremendous amount of awareness and skill. Finally, and most important as it relates to the current volume, the sports concussion clinician must be capable of moving from elementary school-age to middle school and high school athletes and at times to adult college and professional athletes. Without adequate training and supervision, many individuals are unable to traverse this wide range of cultures, ages, and genders. However, for the clinician to deliver competent services, he or she must not only pay attention to these variables but use his or her clinical skills to incorporate them into service delivery.

REFERENCES

ANAM4. (2010). *Sports Medicine Battery: Automated Neuropsychological Assessment Metrics*. Norman, OK: Center for the Study of Human Operator Performance.

Babikian, T., McArthur, D., & Asarnow, R. F. (2013). Predictors of 1-month and 1-year neurocognitive functioning from UCLA longitudinal mild, uncomplicated, pediatric traumatic brain injury study. *Journal of the International Neuropsychological Society, 19*(1), 145–154.

Barth, J. T., Alves, W. M., Ryan, T. V., Macciocchi, S. N., Rimel, R. W., Jane, J. A., et al. (1989). Mild head injury in sports: Neuropsychological sequelae and recovery of functioning. In H. S. Levin, H. M. Eisenberg, & A. L. Benton (Eds.), *Mild head injury* (pp. 257–275). New York: Oxford University Press.

Belanger, H. G., & Vanderploeg, R. D. (2005). The neuropsychological impact of sports-related concussion: A meta-analysis. *Journal of the International Neuropsychological Society, 11*, 345–357.

Boone, K. B. (2007). *Assessment of feigned cognitive impairment: A neuropsychological perspective*. New York: Guilford Press.

Bush, S. S., Ruff, R. M., Troster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., et al. (2005). Symptom validity assessment: Practice issues and medical necessity: NAN Policy and Planning Committee. *Archives of Clinical Neuropsychology, 20*(4), 419–426.

Carone, D. A., & Bush, S. S. (2012). *Mild traumatic brain injury: Symptom validity assessment and malingering*. New York: Springer.

Cernich, A., Reeves, D., Sun, W., & Bleiberg, J. (2007). Automated Neuropsychological Assessment Metrics sports medicine battery. *Archives of Clinical Neuropsychology, 22*(Suppl. 1), S101–S114.

CogState. (1999). *CogSport* [Computer software]. Parkville, Victoria, Australia: CogState.

Collie, A., Maruff, P., Makdissi, M., McCrory, P., McStephen, M., & Darby, D. (2003). CogSport: Reliability and correlation with conventional cognitive tests used in postconcussion medical evaluations. *Clinical Journal of Sport Medicine, 13*(10), 28–32.

Concussion in Sports Group. (2013). Sport Concussion Assessment Tool 3 (SCAT3). *British Journal of Sports Medicine, 47*(5), 259.

Covassin, T., Elbin, R. J., Kontos, A. P., & Larson, E. (2010). Investigating baseline neurocognitive performance between male and female athletes with a history of multiple concussions. *Journal of Neurology and Neurosurgery, and Psychiatry, 81*, 597–601.

Covassin, T., Elbin, R. J., Larson, E., & Kontos, A. P. (2012). Sex and age differences in depression and baseline sports-related concussion neurocognitive performance and symptoms. *Clinical Journal of Sports Medicine, 22*(2), 98–104.

Crompton, J. (2012). Peters cops open investigation of high school football program. Retrieved May 10, 2014, from *www.post-gazette.com/local/washington/2012/01/26/Peters-cops-open-investigation-of-high-school-football-program/stories/201201260310#ixzz1xAY4d7iy*.

Darby, D., Pereire-Gale, J., Moriarity, J., Kutcher, J., Patricios, J., Turner, M., et al. (2011, June). *Validity of preseason athlete baseline tests for computerized concussion testing*. Poster presented at the 58th annual meeting of the American College of Sports Medicine, Denver, CO.

DeRight, J., & Carone, D. A. (2013). Assessment of effort in children: A systematic review. *Child Neuropsychology, 12*, 1–24.

Donders, J., & Kirkwood, M. W. (2012). Symptom validity assessment with special populations. In D. Carone & S. Bush (Eds.), *Mild traumatic brain injury: Symptom validity and assessment of malingering* (pp. 399–410).New York: Springer.

Erdal, K. (2012). Neuropsychological testing for sports-related concussion: How athletes can sandbag their baseline testing without detection. *Archives of Clinical Neuropsychology*, *27*(5), 473–479.

Erlanger, D. M. (2002). *HeadMinder concussion resolution index (CRI): Professional manual*. New York: HeadMinder.

Farr, D. (2013). NFL, retired players reach $765 million settlement in concussion lawsuit. Retrieved May 10, 2014, from *http://nfl.si.com/2013/08/29/nfl-concussion-lawsuit-settlement*.

Green, P., Rohling, M. L., Lees-Haley, P., & Allen, L. M. (2001). Effort accounts for more neuropsychological test variance then does ability in litigant sample. *Brain Injury, 15*, 1045–1060.

Gualtieri, C. T., & Johnson, L. G. (2008). Computerized test battery sensitive to mild and severe brain injury. *Medscape Journal of Medicine, 10*(4), 90.

Guskiewicz, K. M., Weaver, N. L., Padua, D. A., & Garrett, W. E. (2000). Epidemiology of concussion in collegiate and high school football players. *American Journal of Sports Medicine, 28*(5), 643–650.

Harmon, K. G., Drezner, J. A., Gammons, M., Guskiewicz, K. M., Halstead, M., Herring, S. A., et al. (2013). American Medical Society for Sports Medicine position statement: Concussion in sport. *British Journal of Sports Medicine, 47*(1), 15–26.

Hardey, T. R., Echemendia, R., Collins, M., Albo, R., Chang, S., Martin, R., et al. (2001, November). *Neuropsychological screening in professional, college, and high school sports*. Workshop presented at the 21st annual meeting of the National Academy of Neuropsychology, San Francisco, CA.

Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., Millis, S. R., & Conference Participants. (2009). American Academy of Clinical Neuropsychology Consensus Conference Statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist, 23*(7), 1093–1129.

Hill, B. D., Womble, M. N., & Rohling, M. L. (2015). Logistic regression function for detection of suspicious performance during baseline evaluations using Concussion Vital Signs. *Applied Neuropsychology Adult, 22*(3), 233–240.

Hung, R., Carroll, L. J., Cancelliere, C., Cote, P., Rumney, P., Keightley, M., et al. (2014). Systematic review of the clinical course, natural history, and prognosis for pediatric mild traumatic brain injury: Results of the International Collaboration on Mild Traumatic Brain Injury Prognosis. *Archives of Physical Medicine and Rehabilitation, 95*(3), S174–S191.

Hunt, T. N., Ferrara, M. S., Miller, L. S., & Macciocchi, S. (2007). The effect of effort on baseline neuropsychological test scores in high school football athletes. *Archives of Clinical Neuropsychology, 22,* 615–621.

Institute of Medicine. (2013). *Sports-related concussion in youth: Improving the science, changing the culture*. Washington, DC: Author.

Jinguji, T. M., Bompadre, V., Harmon, K. G., Satchell, E. K., Gilbert, K., Wild, J., et al. (2012). Sport Concussion Assessment Tool—2: Baseline values for high school athletes. *British Journal of Sports Medicine, 46*(5), 365–370.

Karantzoulis, S., & Randolph, C. (2013). Modern chronic traumatic encephalopathy in retired athletes: What is the evidence? *Neuropsychological Review, 23*(3), 350–360.

Keating, P. (2012). Concussion test may not be a panacea. Retrieved May 10, 2014, from *http://espn.go.com/espn/otl/story/_/id/8297794/neuropsychological-testing-concussions-not-panacea*.

Keith, M. S., Stanislav, S. W., & Wesnes, K. A. (1998). Validity of a cognitive computerized assessment system in brain-injured patients. *Brain Injury, 12*, 1037–1043.

Kirkwood, M. W., Randolph, C., & Yeates, K. O. (2012). Sport-related concussion: A call for evidence and perspective amidst the alarms. *Clinical Journal of Sport Medicine, 22*(5), 383–384.

Kirkwood, M. W., Yeates, K. O., & Wilson, P. E. (2006). Pediatric sport-related concussion: A review of the clinical management of an oft-neglected population. *Pediatrics, 117*(4), 1359–1371.

Larrabee, G. J. (2007). *Assessment of malingered neuropsychological deficits*. New York: Oxford University Press.

Larrabee, G. J. (2012). Performance validity and symptom validity in neuropsychological assessment. *Journal of the International Neuropsychological Society, 18*(1), 1–7.

Larrabee, G. J. (2014). False-positive rates associated with the use of multiple performance and symptom validity tests. *Archives of Clinical Neuropsychology, 29*(4), 364–373.

Lavigne, P. (2012). Concussion news worries parents. Retrieved May 10, 2014, from *http://espn.go.com/espn/otl/story/_/id/8297366/espn-survey-finds-news-coverage-concussions-leads-majority-parents-less-likely-allow-sons-play-youth-football-leagues*.

Lovell, M. R. (2007). *ImPACT 2007 (6.0) software user's manual*. Pittsburgh, PA: ImPACT Applications.

Lovell, M. R. (2013). *Immediate post-concussion assessment testing (ImPACT) test: Clinical interpretive manual online ImPACT 2007–2012*. Pittsburgh, PA: ImPACT Applications.

Macciocchi, S. N., & Broglio, S. P. (2012). Symptom validity assessment and sports concussion. In D. A. Carone & S. S. Bush (Eds.), *Mild traumatic brain injury: Symptom validity assessment and malingering* (pp. 366–380). New York: Springer.

McCrea, M., Guskiewicz, K. M., Marshall, S. W., Barr, W., Randolph, C., Cantu, R. C., et al. (2003). Acute effects and recovery time following concussion in collegiate football players: The NCAA Concussion Study. *Journal of the American Medical Association, 290*(19), 2556–2563.

McCrea, M., Hammeke, T., Olsen, G., Leo, P., & Guskiewicz, K. M. (2004). Unreported concussion in high school football players: Implications for prevention. *Clinical Journal of Sports Medicine, 14*(1), 13–17.

McCrory, P., Meeuwisse, W., Aubry, M., Cantu, B., Dvorak, J., Echemendia, R., et al. (2013). Consensus Statement on Concussion in Sport—The 4th International Conference on Concussion in Sport Held in Zurich, November 2012. *Clinical Journal of Sport Medicine, 23*(2), 89–117.

Mihoces, G. (2013). Documentary: For years NFL ignored concussion evidence. Retrieved May 10, 2014, from *www.usatoday.com/story/sports/nfl/2013/10/07/frontline-documentary-nfl-concussions/293974*.

Morgan, J. E., & Sweet, J. J. (2009). *Neuropsychology of malingering casebook*. New York: Psychological Press.

Moser, R. S., Schatz, P., Neidzwski, K., & Ott, S. D. (2011). Group versus individual administration affects baseline neurocognitive test performance. *American Journal of Sports Medicine, 39*(11), 2325–2330.

Omalu, B. I., Steven, D. T., Minster, R. L., Kamboh, M. I., Hamilton, R. L., & Wecht, C. H. (2005). Chronic traumatic encephalopathy in a National Football League player. *Neurosurgery, 57*(1), 128–134.

Pennington, B. (2014, May 5). Flubbing a baseline test on purpose is often futile. Retrieved May 8, 2014, from *www.nytimes.com/2013/05/06/sports/sandbagging-first-concussion-test-probably-wont-help-later.html?_r=0*.

Rahman-Filipiak, A. A. M., & Woodard, J. L. (2013). Administration and environment considerations in computer-based sports-concussion assessment. *Neuropsychology Review, 23*(4), 314–334.

Randolph, C. (2011). Baseline neuropsychological testing in managing sport-related concussion: Does it modify risk? *Current Sports Medicine Report, 10*(1), 21–26.

Randolph, C. (2014). Is chronic traumatic encephalopathy a real disease? *Current Sports Medicine Reports, 13*(1), 33–37.

Randolph, C., & Kirkwood, M. W. (2009). What are the real risks of sport-related concussion, and are they modifiable? *Journal of International Neuropsychological Society, 15*(4), 512–520.

Reeves, D. L., Winter, K. P., Bleiberg, J., & Kane, R. L. (2007). ANAM genogram: Historical perspectives, description and current endeavors. *Archives of Clinical Neuropsychology, 22*(1), S15–S37.

Resch, J. E., McCrea, M. A., & Cullum, C. M. (2013). Computerized neurocognitive testing in the management of sport-related concussion: An update. *Neuropsychology Review, 23*(4), 335–349.

Rohling, M. L. (2011, June). *The automation of cognitive assessment in clinical trials: Strategies and methods to produce a stable neurocognitive baseline at randomization: Perspectives from CNS Vital Signs*. Paper presented at the annual meeting of the American Society of Clinical Psychopharmacology, Boca Raton, FL.

Rohling, M. L., Binder, L. M., Demakis, G. J., Larrabee, G. J., Ploetz, D. M., & Langhinrichsen-Rohling, J. (2011). A meta-analysis of neuropsychological outcome after mild traumatic brain injury: Re-analyses and reconsiderations of Binder et al. (1997), Frencham et al. (2005), and Pertab et al. (2009). *The Clinical Neuropsychologist*, *25*(4), 608–623.

Rohling, M. L., Meyers, J. E., & Millis, S. (2003). Neuropsychological impairment following TBI: A dose–response analysis. *The Clinical Neuropsychologist, 17*, 289–302.

Schatz, P., & Glatts, C. (2013). "Sandbagging" baseline test performance on ImPACT, without detection, is more difficult than it appears. *Archives of Clinical Neuropsychology, 28*(3), 236–244.

Schatz, P., Moser, R. S., Solomon, G. S., Ott, S. D., & Karpf, R. (2012). Incidence of invalid computerized baseline neurocognitive test results in high school and college students. *Journal of Athletic Training, 47*(3), 289–296.

Smith, M. D. (2011, April 22). To avoid concussion rules some players sandbag their baseline tests. Retrieved May 8, 2014, from *http://profootballtalk.nbcsports.com/2011/04/22/to-avoid-concussion-rules-some-players-sandbag-their-baseline-tests*.

Szabo, A. J., Alosco, M. L., Fedor, A., & Gunstad, J. (2013). Invalid performance and the ImPACT in National Collegiate Athletic Association Division I football players. *Journal of Athletic Training, 48*(6), 851–855.

Tiedens, J. (2014, February). High school athletes sandbag concussion tests, keep playing. Retrieved April 8, 2014, from *www.studentreportinglabs.com/video/sports/high-school-athletes-sandbag-concussion-tests-keep-playing*.

Womble, M. N., Rohling, M. L., Hill, B. D., & Corley, E. R. (2012, November). *Group size effects for baseline and post-injury sports-related concussion assessment*. Poster presented at the annual meeting of the National Academy of Neuropsychology in Nashville, TN. *Archives of Clinical Neuropsychology, 27*(6), 575

Zuspan, S. J. (2013, April). Pediatric concussion: Do you really know pediatric sports concussion? Workshop presented at the annual meeting of the Society of Trauma Nurses, Las Vegas, NV. Retrieved May 10, 2104, from *www.slideshare.net/mdconferencefinder/stn-16th-annual-conference*.

13

Pediatric Forensic Neuropsychological Evaluations

JACOBUS DONDERS

This chapter reviews the unique nature, contingencies, and challenges that are inherent to pediatric forensic neuropsychological evaluations, defined here as those assessments of children that occur within a civic medicolegal context pertaining to alleged personal or medical injury. Evaluations pertaining to child custody or criminal responsibility are outside the scope of this chapter, and the reader is referred to other sources for those evaluations (Denney, 2012; King, 2013). Furthermore, issues that are specific to disability evaluations are discussed by Chafetz in this volume (Chapter 14). In this chapter, some of the basic aspects of medicolegal procedures and contingencies, and how these differ from a traditional clinical context, are reviewed first. Next, the specific role and the scientific foundation of performance validity tests (PVT) and symptom validity tests (SVT) are discussed, along with other interpretive issues that often arise in the context of pediatric forensic neuropsychological evaluations. Finally, a case example is provided to illustrate how PVTs and SVTs are used during a pediatric forensic neuropsychological evaluation, along with possible ways to present these findings in a report and to a jury.

DIFFERENCES BETWEEN CLINICAL AND FORENSIC EVALUATIONS

Prior to delving into forensic evaluations, the pediatric neuropsychologist (PN) should make sure to have a good understanding of how a forensic context and/or role differs from the traditional clinical ones. In addition, the PN needs to have a basic understanding of applicable state and federal laws as they pertain to forensic evaluations and of how the medicolegal system works.

In a clinical setting, a PN will typically receive a referral from a known physician, with a joint interest as both a treating doctor and/or an advocate for the patient. In contrast, in a forensic context, the referral is typically from an attorney who represents one of two sides in a legal dispute that is inherently adversarial in nature, and the PN is instead asked to function as an objective expert. As such, the client is typically the attorney and not necessarily the child and family; particularly not if the PN has been retained by the defense in a personal injury lawsuit. The PN's service is ultimately designed to help the trier of the facts in the case (i.e., a judge or jury) and not any individual party in it. The PN will most often not be allowed to discuss test results directly with the family and typically cannot initiate treatment on the basis of the evaluation.

A second very important difference between clinical and forensic pediatric neuropsychological evaluations pertains to the level and limits of confidentiality. Typically, in a clinical setting, PNs will go to great lengths to protect private health information about the child unless they are legally required otherwise. However, this privilege has essentially been waived in a forensic context when the issue of contention is the cognitive or behavioral health of the child and its alleged relationship to some external event such as a motor vehicle collision, exposure to environmental toxin, or alleged medical malpractice. Even if the PN has been retained by the plaintiff's attorney, all test results and the entire report will eventually be discoverable; that is, the opposing side will have access to all the documented history, observations, and neuropsychological data. In a clinical setting, most PNs will try to protect and maintain test security, but in forensic cases, it is not unusual to receive a subpoena for the raw data to be sent to the opposing attorney. Attix and colleagues (2007) have detailed a specific pathway of decisions on how to deal with such requests, which reflects the official joint position of various professional neuropsychological organizations.

If the PN is aware and accepting of these major differences between a clinical and a forensic referral, then the next step is to develop a basic understanding of the major procedures and requirements of the legal system. This is not optional. It is part and parcel of the American Psychological Association (APA; 2002) ethics code; specifically, Standard 2.01(f). The first part of the procedure to understand is that since the referral is medicolegal in nature and the attorney is the client, it is not appropriate to bill the child's medical insurance for a forensic neuropsychological evaluation. As a consequence, any fees and/or retainers for services should be negotiated with the referring attorney and documented explicitly in writing prior to the evaluation.

Another important consideration before even seeing the child is that, in a forensic evaluation, the PN is typically required to review more voluminous records than is typical in most clinical evaluations. It is imperative that the PN request access to any relevant premorbid medical records (e.g., developmental milestones during office visits with a pediatrician prior to the commencement of lead exposure), as well as premorbid school records (e.g., standardized test scores, disciplinary reports). Exclusive reliance on self or parent report about these matters is not acceptable in forensic evaluations. All records should typically be securely maintained by the PN until after the case has been settled or otherwise adjudicated.

Yet another aspect of the forensic arena that is important for the PN to understand is that there are state and federal standards pertaining to admissibility of evidence. A detailed discussion of these rules of evidence is beyond the scope of this chapter, and the reader is referred to Greiffenstein (2008) and Kaufmann (2012) for this purpose. At a minimum, though, the PN needs to ascertain whether the state or jurisdiction operates under the Frye or under the Daubert standard. The former is more liberal and only requires an opinion by the court that the methods used by the PN are generally accepted in the practice field. The Daubert standard is more specific and requires, in addition to general acceptance, that those methods (1) have been subjected to peer review, (2) can potentially be empirically evaluated and disproved, (3) have a known error rate in classification, and (4) are accompanied by a professional manual. Because of these requirements, the PN should think very carefully before considering any procedures that are not well standardized, that have low reliability, and/or that have not been validated for specific pediatric purposes.

The PN who is considering forensic work should also have a general understanding of the audience for the report or testimony, which means being prepared to document and explain the findings in a way that members of the public who constitute a jury can understand. PNs who feel uncomfortable speaking in public or who have difficulty with phrasing complex psychometric or neuropathological issues in laymen's terms should probably refrain from doing forensic work.

The PN is also commonly asked to provide some kind of sworn testimony about the case, most often in the form of a deposition but sometimes also at trial (discussed later). The PN who is considering forensic work should be prepared for the fact that opposing attorneys are expected by the nature and standards of their trade to argue a case for their client in a vigorous manner. That means that the PN may face assertive challenges to the choice of instruments, details of interpretation, or attribution of causation. This is to be expected and should most often not be taken personally. However, those individuals who tend to get flustered or defensive very easily may want to reconsider whether venturing into the forensic arena is the right thing for them.

PREPARING FOR A FORENSIC EVALUATION

Once the PN has a good understanding of the general contingencies of a medicolegal evaluation, as well as of the basics of the most relevant deposition and court proceedings, specific referrals can be considered. First, the PN needs to ascertain from the outset the nature of the case and associated questions that he or she will be expected to answer. When considering a forensic referral, PNs must only accept cases that are within their boundaries of professional competence, as required by Standard 2.01(a) of the APA ethics code (2002). Therefore, a PN who has only limited clinical experience and/or continuing education with regard to the condition of interest (be that hypoxia in a neonate, lead poisoning in a toddler, or something else) should decline a request for a forensic evaluation of a child with such an alleged history. Boundaries of competence also pertain to scope of testimony. It is very important that the PN offers only written and verbal opinions that are (1) based on his or her expertise, (2) evidence based, and (3) not speculative in nature. For example, in the case of an allegation of medical malpractice during a complicated delivery 3 years previous, the PN may prepare to comment on the degree to which the child's current neurobehavioral functioning is abnormal but should plan to refrain from comments about whether or not the obstetrician acted negligently during the perinatal process.

Once the PN is satisfied that the nature of the case is understood and that the topic is within an area of professional competence, assurance is needed from the retaining attorney that an unbiased and objective examination can be conducted, with access to all relevant documents, as well as to the child. If the child has already undergone a prior neuropsychological or educational evaluation with a different provider, it is standard of care to request that report, as well as the associated raw data. Any potential conflicts of interest should be ruled out prior to accepting the case. In particular, dual relationships should be avoided, as stated in Standard 3.05(a) of the APA ethics code. For example, a PN who has at one time completed a regular clinical evaluation at the request of the child's physician cannot later accept a request from the parents' attorney for another evaluation for medicolegal purposes.

In some cases, the PN will be asked only to do a records review. This is permissible as long as the limits this review places on specific diagnostic impressions are documented, consistent with standards 9.01(b) and 9.01(c) of the APA ethics code. More commonly, the attorney will ask for an in-person evaluation. Key requirements throughout the entire process of evaluation, report preparation, and sworn testimony are personal and empirical neutrality. It is distinctly not the role of the PN to try to "win the case" but to discern objective information and to communicate that information in an intelligible way so that it will ultimately help the judge and/or jury.

A potential barrier for a forensic evaluation can be a request from the opposing attorney for the presence of a third-party observer during the testing. Jurisdictions vary considerably in the degree to which they allow or discourage this, and this situation highlights the importance of familiarity with local legal procedures. In general, though, the PN should object to the presence of third-party observers during the actual testing, because of the deviation from standardized procedures and the likely confounding impact on the behavior of the examinee (American Academy of Clinical Neuropsychology, 2001; National Academy of Neuropsychology, 2000). Most attorneys will agree to limit the presence of a third person to the interview, but this is not universal. Howe and McCaffrey (2010) provide detailed suggestions about how to deal with third-party observer requests.

Another potential barrier may be that the opposing attorney may request, prior to the evaluation, a list of all the tests that the PN plans to administer. It is advisable to respond to such requests with an explanation that specific tests will not be chosen until after the completion of the interview and history but that a complete list can be provided of every possible test that is available in the PN's arsenal.

CONDUCTING A FORENSIC EVALUATION

The PN must explain to the parents or guardians of the child at the very beginning of the evaluation who had retained him or her, what the nature and purpose of the evaluation are, and that there is no traditional or confidential doctor–patient relationship in this case. Informed consent from the parent or guardian, as well as assent from the child, must be obtained in a language that they can understand, consistent with Standard 3.10(a–d) of the APA ethics code. Sample consent forms are available from the website of the National Academy of Neuropsychology (*www.nanonline.org*).

If the PN conducts the evaluation at the request of the defense, families may sometimes be reluctant to, or may have been advised by their attorney not to, answer questions about specific issues, such as family medical history or disagreements with their doctor and/or insurance. If that happens, the PN must clearly note in the report how this limits the scope of the conclusions that can be made. Similarly, if important records appear to be missing (e.g., the parents report that the child had a prior evaluation with a different provider), the PN must document that he or she reserves the right to change the conclusions after a review of such records.

Although the literature lacks consensus about the ideal assessment approach, the PN should use only procedures that are (1) consistent with the ethical and professional guidelines in his or her field, (2) widely accepted in the professional community, and (3) solidly grounded in psychometric and neurobehavioral science. For example, in a case involving pediatric traumatic brain injury (TBI), the PN would be on solid ground with use of a test such as the California Verbal Learning Test—Children's Version (CVLT–C; Delis, Kramer, Kaplan, & Ober, 1994), given that this instrument has been nationally standardized and norm referenced as documented in a comprehensive test manual, is widely used in the field, and has also been thoroughly researched with regard to its clinical utility in children with TBI (Donders & Nesbit-Greene, 2004; Miller & Donders, 2003; Mottram & Donders, 2005).

VALIDITY TESTS

The reasons to use PVTs and SVTs during pediatric forensic neuropsychological evaluations mirror, to a large degree, the ones that have been discussed in previous chapters with regard to specific clinical evaluations. Even when there are no financial contingencies, children and adolescents simply do not always do their best during neuropsychological evaluations. Their reasons for this may range from desire to avoid nonpreferred activities to complicating psychosocial stressors (Kirkwood, Kirk, Blaha, & Wilson, 2010). In a forensic context, in which financial incentives are typically prominent, it then becomes even more important to have reliable information about the validity of the examinee's performance, particularly given the large impact of atypical effort, as assessed by failed PVTs, on results from the rest of a pediatric test battery (Kirkwood, Yeates, Randolph, & Kirk, 2012). Several national professional organizations have also underscored the crucial importance of validity assessment during neuropsychological evaluations (Bush et al., 2003; Heilbronner et al., 2009). The reader is referred to the chapter by Chafetz (Chapter 14, this volume) for a discussion of the issue of malingering by proxy. This chapter instead focuses on how to use and report PVT and SVT findings, along with other interpretive issues that are crucial to consider during pediatric forensic neuropsychological evaluations. One important general guideline is that PVT and/or SVT failure does not automatically imply malingering; it just means that the rest of the data cannot be interpreted in the conventional manner and are likely confounded by factors other than acquired cerebral dysfunction.

Although there is no universal agreement about which PVT to use during pediatric forensic neuropsychological evaluations, it is advisable to use measures that have been cross-validated in various independent samples. Most of the research on testing measures has been done with TBI, but other conditions have also been included. This research has shown that both the Test of Memory Malingering (TOMM; Tombaugh, 1996) and the Medical Symptom Validity Test (MSVT; Green, 2004) can be used with considerable confidence with school-age children and adolescents (for a review, see Donders & Kirkwood, 2013). Both of these tests would likely meet a Daubert challenge. With adults, it is typically advisable to use more than one PVT during a forensic evaluation and also to use a combination of stand-alone and embedded measures (Heilbronner et al., 2009). However, the research is at this time inconsistent with regard to the clinical utility of embedded measures, such as Reliable Digit Span, in pediatric cases, even when traditional “adult” cutoffs are adjusted (Kirkwood, Hargrave, & Kirk, 2011; Welsh, Bender, Whitman, Vasserman, & MacAllister, 2012).

There is much less literature on the use of SVTs in pediatric neuropsychological evaluations. Traditional scales such the F family on the Minnesota Multiphasic Personality Inventory—Adolescent (MMPI–A; Butcher et al., 1992) have not fared well in terms of ability to detect feigned psychopathology (Rogers, Hinds, & Sewell, 1996). There have been some investigations (Wrobel et al., 1999) that have included healthy adolescents who were asked to simulate emotional impairment for monetary gain on the Personality Inventory for Youth (PIY; Lachar & Gruber, 1995) or that confirmed that parental response bias on the Personality Inventory for Children—Second Edition (PIC–2; Lachar & Gruber, 2001) was associated with relative overreporting of pathology in a preadolescent inpatient psychiatric sample (Stokes, Pogge, Wecksell, & Zaccario, 2011). However, there are currently no peer-reviewed published studies about how the validity scales on those instruments typically fare in pediatric samples with known, suspected, or disputed neurological compromise. In a recent study of parent- and self-reported levels of executive functioning after adolescent TBI on, respectively, the Behavior Rating Inventory of Executive Function (BRIEF; Gioia, Isquith, Guy, & Kenworthy, 2000) and the Behavior Rating Inventory of Executive Function—Self-Report (BRIEF-SR; Guy, Isquith, & Gioia, 2004), disagreements between parents and adolescents were common, but overt negative response bias was not (Wilson, Donders, & Nguyen, 2011). However, the true sensitivity of validity indices on those instruments to simulated or exaggerated acquired impairment is still unknown.

One recent study investigated the level of agreement between the validity scales from the Behavior Assessment System for Children—Second Edition (BASC–2; Reynolds & Kamphaus, 2004) and the MSVT in a pediatric sample with TBI (Kirk, Hutaff-Lee, Connery, Baker, & Kirkwood, 2014). This was the first study to systematically investigate the relationship between an SVT and a PVT in a neurologically involved pediatric sample. The authors found that failure of validity criteria on the BASC-2 was far less common than on the MSVT and that, in general, there was no substantive relationship between validities as assessed by either instrument. These findings stand in sharp contrast to the adult literature, in which considerable agreement between SVTs and PVTs has been reported (Martens, Donders, & Millis, 2001; Whitney, Davis, Shepard, & Herman, 2008), and they reinforce the fact that pediatric forensic neuropsychology is a distinct subspecialty. Although the Kirk et al. (2014) findings require independent replication and extension with different instruments, the provisional lesson to be learned is that a PN should not rely exclusively on SVTs during an evaluation to rule out invalid performance on the cognitive tests. In general, PVTs and SVTs may measure very different types of validity during pediatric neuropsychological evaluations, and the PN will most likely want to use both of them, but not consider them interchangeably.

When using PVTs or SVTs, it is advisable to always comment in the report on the findings, both when the child passed the test and when the child failed. Normal findings on the MSVT by an expert, retained by the defense in a lead poisoning lawsuit in a girl whose parents were originally opposed to the evaluation, are likely very informative with regard to the effort that the child nevertheless put into her performance. At the same time, an extremely elevated negativity index on the BRIEF–SR could raise concerns about the veracity of an adolescent's claim of amnesia about whether or not he was the alleged driver in a motor vehicle collision in which several other teens were killed.

It is important for the PN to avoid giving much information directly to the examinee about the PVT or SVT findings. In a regular clinical context, it is arguably permissible to discuss concerns about effort or task engagement during a feedback session (Carone, Iverson, & Bush, 2010) or even halfway through the evaluation to see whether this would alter the examinee's behavior (Donders, 2011). However, in a forensic context, feedback of any kind to the examinee is typically not permitted. It is advisable to emphasize at the beginning of the evaluation the importance of best effort and truthful responding, without going into detail about the use of PVTs or SVTs, and then let the chips fall where they may.

When reporting on PVTs or SVTs, it is also important to use language that is scientifically accurate, as well as intelligible to a nonpsychologist, while not giving away trade secrets. For example, instead of saying that the examinee "got only *X*% correct, whereas *Y* is the cutoff," the PN might use language such as "this person with an almost indistinguishable blood lead level of < 2 μg/dl performed way worse than the vast majority of patients who have significant neurological and developmental disabilities." In this context, it is also important to avoid pejorative language. Sticking to "just the facts" will typically serve the PN, as well as the public, better.

OTHER INTERPRETIVE ISSUES

Passing or failing on PVTs and/or SVTs is not the only interpretive issue that takes on even more prominence during forensic evaluations than during most clinical ones. One particularly important consideration is the issue of the effect of repeated exposure to the same test. In a forensic context, it is much more common for the examinee to have undergone at least one prior neuropsychological evaluation, and sometimes more than one, often within a relatively short time frame. In this context, the PN needs to consider the effects of practice, as well as regression toward the mean, along with more basic considerations such as the reliability of the test itself. From a purely psychometric point of view, it is advisable for the PN to consider formal actuarial methods to determine whether the observed change in performance is statistically reliable or relevant. Reliable change intervals (possibly with adjustment for practice effects), as well as regression methods, are available for this purpose (see Brooks, Strauss, Sherman, Iverson, & Slick, 2009, for a review). However, PNs should appreciate that even statistically significant change does not necessarily mean clinically meaningful change. In this context, consideration of the physiological or clinical plausibility of a particular pattern of change is important. For example, after a one-time, static, uncomplicated mild TBI in an adolescent, one would not expect deterioration in same-test performance over time. In contrast, a severe TBI that was sustained in early childhood can interfere with emerging and/or rapidly developing skills, leading to a widening of the gap between the child and his or her peers over the years (for a review, see Kirkwood, Yeates, & Bernstein, 2010).

Another factor that may affect interpretation of the test findings is that, in a forensic evaluation, a PN may be inclined to administer additional tests in order to be thorough. It needs to be realized that this will come at a price of increasing risk for spurious findings. The more tests one gives, the more likely it is that there will be at least one or a few poor scores, and overinterpretation of such scores may lead to erroneous conclusions. The PN should appreciate that test score scatter is the norm, not the exception, and that most neurologically healthy children will obtain some low scores when multiple tests are administered (Brooks, Iverson, Sherman, & Holdnack, 2009; Brooks, Sherman, & Iverson, 2010). In addition, the more comparisons one makes between scores on the same test, the more likely it is that one of them will be statistically significant, but that does not necessarily mean it is unusual or clinically significant (Donders, 2012). Furthermore, the number of low scores that can be expected will depend on characteristics of the child, as well as of the parents. For example, children of lower general ability level are relatively more likely to have low scores than highly intelligent children, whereas the prevalence of poor scores is relatively lower in children of college-educated parents than in children whose parents did not complete high school (Brooks & Iverson, 2012). None of this means that any low scores should just be dismissed. However, it is the task of the PN to consider them on the basis of their statistical probability and base rate, in concert with the known clinical characteristics of the case. For example, if a child sustained a left frontal cerebral hemorrhage as the result of perinatal complications, then a pattern at the age of 5 years of selectively low scores on several tests of fine motor speed with the right hand would make conceptual sense.

The definition of what constitutes a low score also needs careful consideration by the PN. Practitioners vary widely in what they consider to be a score indicative of "impairment," with some using a fairly liberal criterion of 1 standard deviation below the mean and others a more conservative cutoff of 1.5 or even 2 standard deviations. As Schoenberg, Scott, Rinehardt, and Mattingly (2014) have indicated, less strict cutoffs result in increased sensitivity to potential mild dysfunction but increase the risk of making false-positive errors, whereas more conservative standards result in increased specificity but increase the risk of making false-negative errors. Whether a false-positive or a false-negative error is more important to avoid is something the PN needs to consider in advance and be ready to defend later on. Regardless of which threshold for impairment is utilized, it is very important in a forensic context to be consistent in one's use of criteria for impairment (e.g., not using a more stringent standard for defense than for plaintiff cases). In addition, it is important to be able to explain to a lay person what that score or cutoff means. For example, a jury member or even an attorney might think that an IQ of 88 is way below average and therefore suggestive of a problem, but if the PN explains in the report or during testimony that 1 out of every 5 people in the general population has an IQ of that magnitude or lower, this might correct some potentially important misperceptions.

PREPARING A FORENSIC REPORT

A report on a pediatric neuropsychological evaluation should be clear as to who made the referral and what the PN knew about the case before even laying eyes on the child. It is advisable to document who provided available medical or school records. Did they come only from the referring attorney, or did the family bring in additional ones? If specific records were not available, what (if any) attempt was made to obtain them? Some attorneys or psychologists like to see in the report an itemized list of all the records that were reviewed; others prefer a summary of the most pertinent findings, with specific references when indicated (e.g., that the lead level as documented by the local health department on a particular date was 24 μg/dl). In either case, the records review section of the report should be clear as to what the PN was able to ascertain about the child and the alleged neurological injury from outside sources. If information becomes available at some later point that alters the PN's conclusions, this information should be documented in an addendum to the report.

The history as obtained from the child and family should then be detailed as well. If there were discrepancies between parental report and what was ascertained from objective medical records, it is typically a good idea to point those out in a factual manner. For example, rather than accusing the parents of being "untruthful," it is better to simply state that their report of their child being a "straight A student" prior to the alleged neurological injury in question was "inconsistent with" the year-to-date and cumulative grade point averages of 2.37–2.52 that were documented in the actual academic records.

The report should also document that informed consent and assent were obtained, with specific reference to the explanation of issues such as suspension of the usual rules of confidentiality and absence of establishment of a treating doctor–patient relationship. In addition, in a forensic report, the PN may need to include more than the usual amount of detail about which tests were selected and why. For example, if an adolescent had already performed within the low-average range on the Wisconsin Card Sorting Test less than 1 year ago with a different neuropsychologist, there would be little reason to repeat such a test, which is known to have considerable practice effects. This point may be obvious to another neuropsychologist, but it will likely require explanation to those for whom the report is mostly written—that is, lawyers and the trier of facts.

The results section of the report should document objectively and in a nonpartisan manner how the child's performance differed from that of his or her peers and—in the case of a repeat evaluation—whether it changed meaningfully over time. It is important to do so in a way that nonpsychologists can understand. For example, if it is known from independent empirical studies that test *X* has a specificity of 90% and test *Y* a specificity of 85%, then low scores below the cutoff for impairment on both tests in the same child would typically be expected less than 2% of the time ($[1 - 0.90] \times [1 - 0.85] = 0.015$). Describing this performance as something that would most likely occur in fewer than 2 out of every 100 neurologically healthy children would place the information in a context that is likely intelligible to attorneys, judges, and juries.

The section of the report in which the PN makes specific interpretations of the findings and conclusions about how they relate to the integrity of cerebral functioning is likely of most interest to the ultimate audience. It is in this section that the PN needs to be especially cognizant of the need to be objective, evidence-based, and impartial. When PVTs or SVTs were failed, it is advisable to use neutral language that still makes the weight and implication of that information very clear. For example, in the case of a child who failed the TOMM, as well as the MSVT, the PN might say:

> "Unfortunately, the validity of the current cognitive test results is very doubtful because there were several unmistakable indications that he was not consistently putting forth his best effort. That performance cannot be explained by any neurological condition."

Alternatively, in case of a failed SVT, the PN might report the following:

> "She answered the personality inventory in a very consistent manner. Therefore, lack of attention or misunderstanding cannot explain away the fact that she reported a degree of anxiety that would be extremely unusual, even in adolescents who have been exposed to much worse trauma (e.g., assault, rape) than the low-speed fender-bender in which she was involved."

The report must ultimately answer in an unequivocal manner any questions that were posed for the purpose of the evaluation, as long as the subject matter is within the expertise of the PN. Saying that a child does or does not have brain damage is likely not going to be sufficient. The issue that typically needs clarification is whether any cerebral dysfunction is causally related to the alleged neurological injury in question or to any other factors such as premorbid history, lack of effort, or a host of other variables. Although the PN may comment on a recommended course of action (e.g., to stop unnecessary treatment or to follow up on an issue that has not yet been addressed), a forensic report is typically not prepared for treatment planning purposes. In the end, the PN can typically send the report only to the retaining attorney, but it needs to be realized that this documentation will eventually be accessible to opposing counsel.

SWORN TESTIMONY

Sworn testimony about the case can be in the form of an affidavit, a deposition, or trial testimony. An affidavit is nothing more than a set of written statements by the PN that is sworn and signed in the presence of a notary and then submitted to the attorney. In a deposition, attorneys from both sides of the case plus a court recorder are present, but there is no judge or jury, as in an actual courtroom appearance. Some depositions may be videotaped for later presentation at trial in lieu of live testimony. Because many cases are settled out of court, it is more common for PNs to participate in depositions than to testify live at trial.

For any deposition or trial testimony, the PN must be intimately familiar with the facts of the case, the contents of the child's file, and the standard of care and literature consensus regarding the condition of interest. Particularly if the examination was done months or years ago, the PN should make sure to refamiliarize him- or herself with the case. Before entering a deposition or trial, the PN should confer with the retaining attorney about the procedures and the specific issues that may come up. During testimony, PNs should remember that they are primarily there to assist the jury or judge, during both direct examination by the retaining attorney and cross-examination by the opposing one. It is customary to bring a current curriculum vitae, as well as the case file, or at least a copy thereof, and when possible all relevant records that were reviewed. Particularly in federal cases, it is also advisable to provide a list of all the depositions or trial testimonies that the PN has given over the past 3 years, along with a breakdown of plaintiff versus defense, when possible.

Both depositions and court appearances typically start with the PN being sworn in to give only and completely truthful testimony. The direct examination is then done first by the attorney who retained the PN, after which the opposing attorney may cross-examine. This process typically starts with a review of the PN's training and credentials. It is advisable to be concise about this, focusing on the most important points, and to avoid either self-aggrandizement or disparagement of other experts on the case. If the PN has received training from APA-accredited institutions, is licensed to practice in the state in which the case is pending, and has experience with the condition of interest, there are rarely objections to the PN being allowed to offer expert testimony. Board certification can certainly help but is typically not required in order to be considered an expert.

If the PN had been called as a treating doctor, he or she would be a fact witness who could technically only speak directly to the actual assessment findings. In contrast, an independent expert witness has more leeway to offer professional opinions that may include attribution of causality and discussion of the consensus in the literature. In reality, though, PNs are typically treated in most legal settings as experts, regardless of whether they were treating doctors or retained examiners. Either way, one of the participating attorneys may still, at any point during the testimony, voice some kind of "objection," which could range from lack of sufficient foundation by the other attorney for a specific question to finding that the PN is not responsive or too speculative in answering. During a deposition, it is best to wait until the respective attorneys are done arguing with each other before answering. During a trial, the PN should wait for instruction from the judge.

During sworn testimony, the PN will be expected to answer any reasonable questions about the case at hand. It is important for the PN to appreciate that in the legal system there are different degrees of certainty that are required for an opinion. The issue of whether something is "more likely than not" only requires that it is ≥ 51% likely. In contrast, when asked about a "reasonable degree of scientific certainty," a likelihood of ≥ 90% is required. The most important thing for the PN to remember is that any opinions rendered should be soundly grounded in psychometric and neuropsychological science.

Several references offer specific suggestions about testifying during depositions or trials, including Greiffenstein and Kaufmann (2012), Tsushima and Anderson (1996) and the series of books by Brodsky (1991, 1999, 2004). Donders, Brooks, Sherman, and Kirkwood (in press) provide specific examples of various strategies or challenges that a pediatric PN may encounter during sworn testimony, along with considerations of how to deal with them. The PN must be specifically prepared to field questions regarding the nature and interpretation of PVTs and SVTs. It is important to try to maintain test security while still explaining the findings in a manner that is both empirically accurate and intelligible to a jury, illustrated in the following exchange between a PN and a cross-examining attorney.

Q: Doctor, you testified earlier that my client did not answer enough items on that TOMM test correctly and that you thought that this meant that he was not putting forth his best effort, did you not?

A: I did not think that. I know that to be a fact, based on numerous independent studies that have been published that have shown that kindergartners with much more severe brain injury than your 14-year-old client typically do way better than that on this test.

Q: Would you then please explain to the jury on which items of that TOMM test my client was not doing his best? After all, it looks to me that he got more than half of them right, so I really want to know on which ones you think he was doing something wrong.

A: I am going to respectfully refuse to do that because it would be scientifically indefensible.

Q: But why, doctor? I think it is only fair for the jury to know how you can be so sure of yourself.

A: It would be akin to arguing about which of the 24 cans in a case of beer somebody drank did or did not make him an alcoholic. Many adults have a drink or two, once in a while; they just don't drink an entire case of beer in one sitting. It is just the same with the TOMM; many kids get one or two of the items wrong, but they don't blow almost two dozen of them, like your client did.

When the sworn testimony is concluded, the PN should leave the proceedings in a professional manner. Shaking hands with the attorneys from both sides would show courtesy at the end of a deposition. In a courtroom, the PN should wait until excused by the judge and then simply walk out. Trying to find out later who got a favorable verdict or how much money was awarded is ill advised, as it might bias the PN with regard to future cases. However, PNs who are doing their first deposition or court appearance may benefit from constructive feedback from a senior colleague who has observed the procedures or read the transcript.

CASE EXAMPLE

A 13-year-old, right-handed young woman with a history of uncomplicated mild TBI from a motor vehicle collision 9 months previously was referred by the attorney for the defense for an independent neuropsychological evaluation to determine the presence or absence of any cognitive deficits and their causal relationship to the accident in question. A review of medical records revealed a Glasgow Coma Scale score of 15 and a negative head computerized tomography (CT) scan in the presence of several orthopedic fractures. Review of premorbid school records revealed a mild discrepancy between performance on standardized tests (typically > 75th percentile) and letter grades based on classroom performance and homework (mostly Cs). The parents reported an unremarkable premorbid developmental, medical, and psychosocial history. They were concerned about current concentration problems and increased irritability. The child denied any distress except mild pain in the pelvic area related to fractures. Any substance use was denied, and she was not on any routine prescription medications. There was one prior neuropsychological evaluation, completed 3 months prior to the current one, at the request of the plaintiff's attorney by a different provider. That individual had concluded that the young woman had an acquired ("secondary") form of attention-deficit/hyperactivity disorder (ADHD) as the result of her TBI and had recommended resource room support, which the school had declined to provide.

During the current formal testing, the young woman presented in casual attire with good grooming. Posture was slouched but gait and other motor behaviors were unremarkable. Speech was fluent and coherent. She expressed dissatisfaction with the need to go through several hours of testing but did agree to proceed, based on encouragement by the parents. Test results were as presented in Table 13.1. The BRIEF and BRIEF-SR had also been administered but were considered invalid (discussed later).

The plaintiff's neuropsychologist had apparently drawn her conclusion about acquired ADHD on the basis of a depressed Working Memory index on the Wechsler Intelligence Scale for Children (WISC–IV), which she indicated was statistically significantly below all other indices on the instrument. To further bolster the argument that this was most likely due to TBI, she had mentioned that the Children's Category Test composite T-score was more than 1 standard deviation below the mean and that, because this was considered to be a "general indicator of cerebral integrity," this was purportedly indicative of acquired cerebral dysfunction. The following excerpts provide examples of how the defense expert could address, respectively, the plaintiff's logic and the validity of the current test results and then integrate those issues into a diagnostic conclusion. These examples are offered to illustrate how disagreements with another neuropsychologist can

TABLE 13.1. Neuropsychological Test Results

Measure	Prior evaluation	Current evaluation
TOMM, Trial 1 (raw correct)	n/a	37
TOMM, Trial 2 (raw correct)	n/a	47
TOMM, Retention (raw correct)	n/a	42
WISC-IV, Verbal Comprehension (SS)	95	n/a
WISC-IV, Perceptual Reasoning (SS)	94	n/a
WISC-IV, Working Memory (SS)	80	n/a
WISC-IV, Processing Speed (SS)	97	n/a
KTEA-II, Reading Comprehension (SS)	84	n/a
KTEA-II, Numerical Operations (SS)	98	n/a
Trail Making Test, part A (seconds)	16	24
Trail Making Test, part B (seconds)	42	31
Grooved Pegboard, right hand (seconds)	68	66
Grooved Pegboard, left hand (seconds)	72	74
CVLT-C Trials 1–5 (T-score)	48	n/a
Children's Category Test, Level 2 (T-score)	39	n/a
WCST, Perseverative Errors (T-score)	n/a	58
WCST, Nonperseverative Errors (T-score)	n/a	35
CCPT-II, Omissions (T-score)	n/a	54
CCPT-II, Commissions (T-score)	n/a	58
CCPT-II, Reaction Time (T-score)	n/a	44
CCPT-II, Variability (T-score)	n/a	42
RCFT, Immediate Recall (T-score)	n/a	44
RCFT, Delayed Recall (T-score)	n/a	48
RCFT, Recognition (T-score)	n/a	32

Note. SS, standard score; WISC-IV, Wechsler Intelligence Scale for Children—Fourth Edition; KTEA-II, Kaufman Test of Educational Achievement—Second Edition; CVLT-C, California Verbal Learning Test—Children's Version; WCST, Wisconsin Card Sorting Test; CCPT-II, Connors' Continuous Performance Test—Second Edition; RCFT, Rey Complex Figure Test.

be reported in a factual and nonpejorative manner and how psychometric issues can be clarified in layman's terms.

Discussion of Prior Neuropsychological Findings

Dr. Jones was correct that the WISC–IV Working Memory index was statistically significantly lower than all of the other three indices. However, it was only a marginally below-average result. Furthermore, Dr. Jones did not demonstrate appreciation of the fact that each of those three discrepancies actually has a base rate of at least 14% in the WISC–IV standardization sample. In other words, about one out of every seven neurologically healthy children has at least one such difference. Furthermore, if one takes into account that three different comparisons were made, the odds that at least one of them of that magnitude would be found here are even higher: about one out of every three children. Therefore, I am not convinced that the relatively low Working Memory index is all that unusual. In addition, I am a bit puzzled by the fact that Dr. Jones did not comment that, if anything, Processing Speed was this child's highest factor index score on the WISC–IV, and that is actually the only one of the four indices that is known to be sensitive to TBI.[1]

To experience "secondary" ADHD as the result of a severe TBI is certainly possible, but this young woman only sustained an uncomplicated mild injury. I should note that Dr. Jones did not include formal tests of response or symptom validity. In addition, Dr. Jones did not administer other tests of complex attention. I should add that, if this young woman truly had significant problems with attention, then I would find it hard to understand how she did so well on the CVLT–C. That is a list-learning task where the examinee has to concentrate on what the examiner is saying, in order to be able to learn and remember the information. In this context, it should also be noted that the CVLT–C is known to be much more sensitive to TBI than the Children's Category Test.[2] Furthermore, in this young woman's case, more than half of her total errors on the latter test came from the third subtest, which is notoriously insensitive to TBI. Under such circumstances, interpretation of the composite T-score is ill advised.[3] For these reasons, I am not convinced that the *relatively* low score on the Children's Category Test reflects true, acquired impairment that could be associated with this young woman's TBI.

Discussion of Current Test Results

Due to the recency of Dr. Jones's evaluation, I did not want to repeat all of the same tests because (a) this young woman had done well on some of them already and (b) several of the other ones are known to have significant practice effects. I did repeat a few of them to see if

---

[1]Donders, J., & Janke, K. (2008). Criterion validity of the Wechsler Intelligence Scale for Children—Fourth Edition after pediatric traumatic brain injury. *Journal of the International Neuropsychological Society, 14,* 651–655.

[2]Donders, J., & Giroux, A. (2005). Discrepancies between the California Verbal Learning Test—Children's Version and the Children's Category Test after pediatric traumatic brain injury. *Journal of the International Neuropsychological Society, 11*, 386–391.

[3]Donders, J. (1999). Latent structure of the Children's Category Test at two age levels in the standardization sample. *Journal of Clinical and Experimental Neuropsychology*, *21*, 279–282.

there was a significant change in level or pattern of performance. I also added a few new ones, to get a broader understanding of this young woman's current functioning and to ensure that the validity of the data was objectively measured.

There were several test findings that were unusual and that raised concern about whether this young woman was consistently giving her best effort. First of all, she did not meet empirically established and cross-validated criteria for genuine effort on the TOMM, a test in which she had to pick out which one of two pictures she had been shown before. Adolescents with even severe TBI are known to typically do much better on that instrument than she was performing. Second, there were unusual inconsistencies between the original evaluation by Dr. Jones and the current one, involving the same test. Specifically, on a test of visual attention, scanning, and speed of processing (Trail Making), this young woman needed about 50% more time to complete the relatively easy part A, whereas she needed about 25% less time to complete the cognitively more demanding part B. A third factor of concern was the unusual pattern of performance on some of the current tests that she had not done previously. For example, on a visual memory test (Rey Complex Figure Test; RCFT), she actually missed some items on the multiple-choice recognition trial that she had just drawn correctly by freehand, about 30 minutes after she had first learned the information. Yet, at the same time, it did not appear that this young woman was simply doing poorly or haphazardly across the board. In fact, on a fairly challenging test of sustained attention and impulse control (Conners' Continuous Performance Test; CCPT–II), her performance was entirely within normal limits.

An attempt was also made to obtain quantitative ratings from both this young woman and her parents about their respective impressions of her day-to-day functioning and symptoms. Unfortunately, those findings were not interpretable because of inconsistency in responding in the parent version (making the findings unreliable) and an overly negative response style in self-report (raising doubt about the veracity of the description).

Integration of Findings

Any head trauma that this young woman sustained in the accident in question was uncomplicated, mild in nature. It is unusual to have persistent psychological sequelae for more than a few months after such a minor head injury. The pattern of neuropsychological test findings, both currently and in comparison with the previous evaluation by Dr. Jones, is marred by inconsistencies, atypical findings, and fluctuating effort. On the whole, the pattern is not consistent with any genuine cognitive deficits that can be attributed to TBI. There is specifically no evidence for "secondary" ADHD.

CONCLUSION

The pediatric forensic neuropsychological evaluation differs from a clinical one in a number of important ways, but with proper preparation, PNs who adhere to an ethical and scientifically defensible approach, without personal zeal or bias, can play an important role in legal proceedings concerning alleged neurological injuries in children and adolescents. This preparation requires understanding of (1) the current state of the literature on the condition of interest, (2) performance/symptom validity and other psychometric interpretive issues, and (3) basic legal contingencies and proceedings. The PN who offers clear, intelligible, and evidence-based reports and sworn testimony provides a service to both the profession of clinical neuropsychology and the public. A specific goal for future research is the development and cross-validation of indicators that are embedded in cognitive tests for children to inform about performance validity and that are not ‘just" downward extensions of criteria that were originally developed for adults. In addition, more research is needed with regard to the positive and negative predictive values of symptom validity indicators on widely used inventories of behavior, emotional adjustment, and day-to-day functioning, specifically in samples of children and adolescents with known, suspected, or disputed neurological impairment.

REFERENCES

American Academy of Clinical Neuropsychology. (2001). Policy statement on the presence of third party observers in neuropsychological assessment. *The Clinical Neuropsychologist, 15*, 433–439.

American Psychological Association. (2002). Ethical principles of psychologists and code of conduct. *American Psychologist, 57*, 1060–1073.

Attix, D. K., Donders, J., Johnson-Greene, D., Grote, C. L., Harris, J. G., & Bauer, R. M. (2007). Disclosure of neuropsychological test data: Official position of Division 40 (Clinical Neuropsychology) of the American Psychological Association, Association of Postdoctoral Programs in Clinical Neuropsychology, and American Academy of Clinical Neuropsychology. *The Clinical Neuropsychologist, 21*, 232–238.

Brodsky, S. L. (1991). *Testifying in court: Guidelines and maxims for the expert witness*. Washington, DC: American Psychological Association.

Brodsky, S. L. (1999). *The expert witness: More maxims and guidelines for testifying in court*. Washington, DC: American Psychological Association.

Brodsky, S. L. (2004). *Coping with cross-examination and other pathways to effective testimony*. Washington, DC: American Psychological Association.

Brooks, B. L., & Iverson, G. L. (2012). Improving accuracy when identifying cognitive impairment in pediatric neuropsychological assessments. In E. M. S. Sherman & B. L. Brooks (Eds.), *Pediatric forensic neuropsychology* (pp. 6688). New York: Oxford University Press.

Brooks, B. L., Iverson, G. L., Sherman, E. M. S., & Holdnack, J. A. (2009). Healthy children and adolescents obtain some low scores on a battery of memory tests. *Journal of the International Neuropsychological Society, 15*, 613–617.

Brooks, B. L., Sherman, E. M. S., & Iverson, G. L. (2010). Healthy children get some low scores too: Prevalence of low scores on the NEPSY–II in preschoolers, children, and adolescents. *Archives of Clinical Neuropsychology, 25*, 182–190.

Brooks, B. L., Strauss, E., Sherman, E. M. S., Iverson, G. L., & Slick, D. J. (2009). Developments in neuropsychological assessment: Refining psychometric and clinical interpretive methods. *Canadian Psychology, 50*, 196–209.

Bush, S. S., Ruff, R. M., Tröster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., et al. (2005). Symptom validity assessment: Practice issues and medical necessity: NAN Policy and Planning Committee. *Archives of Clinical Neuropsychology, 20*, 419-426.

Butcher, J. N., Williams, C. L., Graham, J. R., Archer, R. P., Tellegen, A., Ben-Porath, Y. S., et al. (1992). *Minnesota Multiphasic Personality Inventory—Adolescent*. Minneapolis: University of Minnesota Press.

Carone, D. A., Iverson, G. L., & Bush, S. S. (2010). A model to approaching and providing feedback to patients regarding invalid test performance in clinical neuropsychological evaluations. *The Clinical Neuropsychologist, 24*, 759–778.

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (1994). *California Verbal Learning Test—Children's Version*. Austin, TX: Psychological Corporation.

Denney, R. L. (2012). Criminal responsibility and other criminal forensic issues. In G. J. Larrabee (Ed), *Forensic neuropsychology: A scientific approach* (2nd ed., pp. 473–500). New York: Oxford University Press.

Donders, J. (2011). Pediatric traumatic brain injury: Effects of questionable effort. In J. E. Morgan, I. S. Baron, & J. H. Ricker (Eds.), *Casebook of clinical neuropsychology* (pp. 183–190). New York: Oxford University Press.

Donders, J. (2012). Interpretive confounds in the independent pediatric neuropsychological evaluation. In. E. M. S. Sherman & B. L. Brooks (Eds.), *Pediatric forensic neuropsychology* (pp. 182–201). New York: Oxford University Press.

Donders, J., Brooks, B. L., Sherman, E. M. S., & Kirkwood, M. W. (in press). Pediatric forensic neuropsychology. In J. E. Morgan & J. H. Ricker (Eds.), *Textbook of clinical neuropsychology* (2nd. ed.). New York: Taylor & Francis.

Donders, J., & Kirkwood, M. W. (2013). Symptom validity assessment with special populations. In D. A. Carone & S. S. Bush (Eds.), *Mild traumatic brain injury: Symptom validity assessment and malingering* (pp. 399–410). New York: Springer.

Donders, J., & Nesbit-Greene, K. (2004). Predictors of neuropsychological test performance after pediatric traumatic brain injury. *Assessment, 11*, 275–284.

Gioia, G. A., Isquith, P. K., Guy, S. C., & Kenworthy, L. (2000). *Behavior Rating Inventory of Executive Function*. Odessa, FL: Psychological Assessment Resources.

Green, P. (2004). *Medical Symptom Validity Test*. Edmonton, Alberta, Canada: Green's Publishing.

Greiffenstein, M. F. (2008). Basics of forensic neuropsychology. In J. E. Morgan & J. H. Ricker (Eds.), *Textbook of clinical neuropsychology* (pp. 905–941). New York: Taylor & Francis.

Greiffenstein, M. F., & Kaufmann, P. M. (2012). Neuropsychology and the law: Principles of productive attorney–neuropsychologist relations. In G. L. Larrabee (Ed.), *Forensic neuropsychology: A scientific approach* (2nd ed., pp. 23–69). New York: Oxford University Press.

Guy, S. C., Isquith, P. K., & Gioia, G. A. (2004). *Behavior Rating Inventory of Executive Function —Self-Report Version: Professional manual*. Lutz, FL: Psychological Assessment Resources.

Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., Millis, S. R., & Conference Participants. (2009). American Academy of Clinical Neuropsychology Consensus Conference Statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist, 23*, 1093–1129.

Howe, L. L. S., & McCaffrey, R. J. (2010). Third-party observation during neuropsychological evaluation: An update on the literature, practical advice for practitioners, and future directions. *The Clinical Neuropsychologist*, *24*, 518–537.

Kaufmann, P. M. (2012). Admissibility of expert opinions based on neuropsychological evidence. In G. L. Larrabee (Ed.), *Forensic neuropsychology: A scientific approach* (2nd ed., pp. 70–100). New York: Oxford University Press.

King, H. E. (2013). Assessment in custody hearings: Child custody evaluations. In K. F. Kiesinger et al. (Eds.), *APA handbook of testing and assessment in psychology. Vol. 2: Testing and assessment in clinical and counseling psychology* (pp. 587–605). Washington, DC: American Psychological Association.

Kirk, J. W., Hutaff-Lee, C., Connery, A., Baker, D. A., & Kirkwood, M. W. (2014). The relationship between the self-report BASC-2 validity indicators and performance validity test failure after pediatric mild traumatic brain injury. *Assessment, 21*(5), 562–569.

Kirkwood, M. W., Hargrave, D. D., & Kirk, J. W. (2011). The value of the WISC–IV Digit Span subtest in detecting noncredible performance during pediatric neuropsychological examinations. *Archives of Clinical Neuropsychology, 26*, 377–384.

Kirkwood, M. W., Kirk, J. W., Blaha, R. Z., & Wilson, P. E. (2010). Noncredible effort during pediatric neuropsychological exam: A case series and literature review. *Child Neuropsychology, 16,* 604–618.

Kirkwood, M. W., Yeates, K. O., & Bernstein, J. H. (2010). Traumatic brain injury in childhood. In J. Donders & S. J. Hunter (Eds.), *Principles and practice of lifespan developmental neuropsychology* (pp. 299–313). Cambridge, UK: Cambridge University Press.

Kirkwood, M. W., Yeates, K. O., Randolph, C., & Kirk, J. W. (2012). The implications of symptom validity test failure for ability-based test performance in a pediatric sample. *Psychological Assessment, 24,* 36–45.

Lachar, D., & Gruber, C. P. (1995). *Personality Inventory for Youth.* Los Angeles: Western Psychological Services.

Lachar, D., & Gruber, C. P. (2001). *Personality Inventory for Children—Second Edition.* Los Angeles: Western Psychological Services.

Martens, M., Donders, J., & Millis, S. R. (2001). Evaluation of invalid response sets after traumatic head injury. *Journal of Forensic Neuropsychology, 2*, 1–18.

Miller, L. J., & Donders, J. (2003). Prediction of educational outcome after pediatric traumatic brain injury. *Rehabilitation Psychology, 48*, 237–241.

Mottram, L., & Donders, J. (2005). Construct validity of the California Verbal Learning Test—Children's Version (CVLT–C) after pediatric traumatic brain injury. *Psychological Assessment, 17*, 212–217.

National Academy of Clinical Neuropsychology. (2000). Presence of third-party observers during neuropsychological testing: Official statement of the National Academy of Neuropsychology. *Archives of Clinical Neuropsychology, 15*, 379–380.

Reynolds, C. R., & Kamphaus, R. W. (2004). *Behavior Assessment System for Children—Second Edition.* Circle Pines, MN: American Guidance Service.

Rogers, R., Hinds, J. D., & Sewell, K. W. (1996). Feigning psychopathology among adolescent offenders: Validation of the SIRS, MMPI–A, and SIMS. *Journal of Personality Assessment, 67*, 244–257.

Schoenberg, M. R., Scott, J. G., Rinehardt, E., & Mattingly, M. (2014). Test administration, interpretation, and issues in assessment. In K. J. Stucky, M. W. Kirkwood, & J. Donders (Eds.), *Clinical neuropsychology study guide and board review* (pp.97–114). New York: Oxford University Press.

Stokes, J., Pogge, D., Wecksell, B., & Zaccario, M. (2011). Parent–child discrepancies in report of psychopathology: The contributions of response bias and parenting stress. *Journal of Personality Assessment, 93*, 527–536.

Tombaugh, T. N. (1996). *Test of Memory Malingering.* Toronto, Ontario, Canada: Multi-Health Systems.
Tsushima, W. T., & Anderson, R. M. (1996). *Mastering expert testimony: A courtroom handbook for mental health professionals*. Mahwah, NJ: Erlbaum.
Welsh, A. J., Bender, H. A., Whitman, L. A., Vasserman, M., & MacAllister, W. S. (2012). Clinical utility of Reliable Digit Span in assessing effort in children and adolescents with epilepsy. *Archives of Clinical Neuropsychology, 27*, 735–741.
Whitney, K., Davis, J., Shepard, P., & Herman, S. (2008). Utility of the Response Bias Scale (RBS) and other MMPI-2 validity scales in predicting TOMM performance. *Archives of Clinical Neuropsychology, 23*, 777–786.
Wilson, K. R., Donders, J., & Nguyen, L. (2011). Self- and parent ratings of executive functioning after traumatic brain injury. *Rehabilitation Psychology, 56,* 100–106.
Wrobel, T. A., Lachar, D., Wrobel, N. H., Morgan, S. T., Gruber, C. P., & Nether, J. A. (1999). Performance of the Personality Inventory for Youth validity scales. *Assessment, 6*, 367–379.

14

Disability

Social Security Supplemental Security Income Exams for Children

MICHAEL D. CHAFETZ

In his delightfully titled review of validity testing in children ("Who Do They Think They're Kidding?"), Rohling (2004) addresses several of the reasons that children might perform poorly on validity tests. The motivations behind poor validity performance may be traced to the financial stake parents and attorneys have in the outcomes of litigation, often in cases of traumatic brain injury (TBI). There is also pressure (typically from parents) for students to seek accommodations for "disabilities" within the school setting. Medications to help a student focus, or that have abuse and sale potential, might be more reliably secured if a student does poorly on cognitive testing. Rohling warned that children might comply with their parents' wishes to gain more attention and affection, and so the model of malingering for compensation in adults may not be entirely applicable. Rohling thus set the stage for the ways in which children might perform under disability examinations that are the concern of this chapter.

This chapter examines the cognitive and validity performance of children whose parents are seeking Social Security Disability (SSD) on their behalf. Over a period from 2002 to 2007, I (with my students) collected and studied archived data on children (ages 6–16) sent from the local Disability Determination Services (DDSs) for examination for Supplemental Security Income (SSI). Of concern is the nature of the SSI disability program, the relationship between validity test scores and performance on IQ measures, and the examining psychologists' role boundaries within this system.

Donders (Chapter 13, this volume) provides an extensive review of validity testing within the pediatric forensic neuropsychological examination.

DISABILITY HISTORY FOR CHILDREN

President Franklin Delano Roosevelt signed the Social Security Act into law on August 14, 1935, providing an old-age retirement pension for those who were no longer working. Amendments to the Social Security Act in 1954 led to the Disability Insurance program, and on August 1, 1956, President Dwight Eisenhower signed the new disability legislation, which provided monetary benefits to older disabled workers after a waiting period. For several years thereafter, legislation broadened the protections for workers, protecting people with disabilities. Eventually, on October 30, 1972, President Richard M. Nixon signed into law the needs-based SSI program, which provided a social safety net for people who were never able to work and for children (Social Security Administration [SSA], 2012). Under the new law, SSI provided a uniform federal income floor. There was also coordination with the Food Stamp program and medical assistance programs (e.g., Medicaid).

SOCIAL SECURITY DEFINITION OF DISABILITY

Under Social Security administrative law, for an individual to be considered disabled, this individual "must have a medically determinable physical or mental impairment that is expected to last (or has lasted) at least 12 continuous months or to result in death and (1) if 18 or older, prevents him or her from doing any substantial gainful activity, or (2) if under 18, results in marked and severe functional limitations" (SSA, 2012). If addiction to drugs or alcohol contributes materially to the determination of the disability, the individual is not eligible for benefits.

The Listings

For the disability determination, the claimant must meet the listing requirements for the various mental impairments thought to be causing the disabling condition (Social Security Administration, 2014). A mere diagnosis of the child does not automatically invoke the disability determination, as the professionals at the DDSs must still figure out whether there are marked impairments in age-appropriate cognitive–communicative, social, or personal functioning or marked difficulties in maintaining concentration, persistence, or pace. Thus severity is indicated by the functional limitations that are produced. If an impairment is not listed, the DDSs will determine whether it is severe enough to medically or functionally equal the listings.

Social Security Programs

Social Security Disability Insurance (SSDI) is paid to workers and the dependents of workers who have paid into the Social Security Trust Fund through the Federal Insurance Contributions Act tax and the Self-Employment Tax for individuals who work for themselves. The SSI program covers disabled individuals and dependents with limited income and assets. General tax revenues fund SSI, and it is "means-tested," so that people with substantial gainful activity, in which a little more than $1,000 per month is earned, are not qualified. Individuals with this amount of income are considered to be earning enough gainful income to offset any revenues from the SSI. For children with disabilities, the SSI program ensures that parents are getting extra income to offset the costs of their children's special needs.

THE PSYCHOLOGICAL CONSULTATIVE EXAMINATION FOR SOCIAL SECURITY DISABILITY

Puente (1987) and Chafetz (2011a) have discussed the particulars of the psychological consultative examination for Social Security Disability examinations, primarily focusing on adults. Nevertheless, many of the interactions a pediatric psychologist has with the DDS are similar. The liaison at the local DDS must determine whether the psychologist has the appropriate credentialing and is experienced in interpreting and reporting the results of psychological testing. A scheduler from the local DDS will call the credentialed psychologist and arrange the dates and times of appointments for the children to be examined. Before the appointments, the psychologist will receive information that may or may not include previous reports. The referral sheet typically contains only a few words about the referral questions, such as "slow learner" or "ADHD."

THE ROLE OF THE PSYCHOLOGIST IN THESE EXAMINATIONS

Evaluators who perform consultative examinations for SSA should understand that they are working in the medicolegal arena. It is helpful to understand the differences between the medicolegal and clinical examinations. In a clinical consultation, the psychologist's roles might include assessment, diagnosis, recommendations, or forms of intervention and treatment. The client in this arena is the patient, whether self-referred or referred by another provider or insurance carrier.

In the clinical arena, payment to the health care provider is usually expected from the patient's parents or guardians or from an insurance company, except when the psychologist takes the case pro bono, which is work undertaken for the public good without charge. However, no matter what the form or source of payment, the psychologist's goal is to help the patient. The psychologist's duty is to the patient (and/or parents or guardians).

Moreover, the patient has privacy rights under the relevant Health Insurance Portability and Accountability Act of 1996 provisions (*http://go.cms.gov/N2HW0W*). The psychologist attempts to form an alliance with the patient to facilitate assessment and treatment, and a doctor–patient relationship exists as part of this alliance. Within this relationship, appropriate clinical boundaries exist to avoid harm and exploitation (American Psychological Association [APA], 2002).

Medicolegal evaluations like those performed for the DDS are quite different. The evaluator should understand that motivations of patients and claimants differ (Greenberg & Shuman, 1997, 2007; Strasburger, Gutheil, & Brodsky, 1997). The parent of the child in a clinical consultation is seeking to have the child diagnosed or treated to determine what is wrong and to help improve the child's condition. The parent in a disability evaluation is seeking a monetary benefit that is awarded when a child has a marked limitation in functioning.

These different motivations may lead to vastly different behaviors on the part of the parents and their children. Claimants are frequently uncomfortable about anything that might lead to denial of their claims, and it is sometimes difficult to get information about a child's strengths or other normal activities. In a typical clinical consultation, parents frequently provide all kinds of information that helps the examiner with differential diagnosis.

The client in a disability examination for the DDS is the DDS, unless a private Social Security attorney has contracted with the examiner for an evaluation of his client. In this example, the claimant, through his or her attorney, is the client. The motivation of this client can be regarded as the same as that of the claimant, that is, to prove to the DDS that the child has enough impairment to meet the listing requirements for disability benefits.

FORMS OF NONCREDIBLE BEHAVIOR

We are all familiar with the math phobic child who, on the morning of his big math test, develops a stomachache, complaining he is sick. The first time this happens, it is likely an example of somatoform illness in which a stressor conflicting with the child's interests is converted into somatic illness.

It is easy enough for the unwitting parent to shape this child's behavior, facilitating somatoform complaints, and the astute child may eventually start using the sick role in a more deliberate way. The noncredible behavior may have to become more exaggerated and yet at the same time more authentic if the child is to achieve his goals.

In cases of both children and adults, neuropsychologists are confronted with examinees who present with symptoms in excess of explainable medical or psychological conditions. Noncredible behaviors may be somatoform, in which the presentation involves bodily pain, illness, or neurological problems, or they may be cogniform, with a presentation concerning cognitive or memory problems (Delis & Wetter, 2007).

Factitious disorders (Munchausen disorder at the extreme) occur when people deliberately produce or feign physical or psychological illness. The goal is not for compensation but to assume the sick role and gain attention from doctors (American Psychiatric Association, 2000, 2013). Munchausen disorder is particularly severe and chronic, including self-injecting or ingestion of various substances that would necessitate urgent medical treatment and care.

Malingering is also deliberate and intentional (Heilbronner et al., 2009; Slick, Sherman, & Iverson, 1999), with goal seeking that may involve some form of compensation or avoidance of duty (Heilbronner et al., 2009). Sweet and colleagues (2000, p. 106) defined malingering as "purposeful insufficient effort to obtain external gain." In Larrabee's (2007, p. vii) words, malingering is "the exaggeration and/or fabrication of deficits in the pursuit of some external incentive." Malingering has been discussed extensively in the crucible of the disability examination for Social Security (Chafetz, 2010, 2011a).

As traditionally discussed (Delis & Wetter, 2007), both malingering and factitious disorder involve symptoms that are produced deliberately or intentionally, whereas somatoform and cogniform disorders involve behaviors that are unintentional and possibly involuntary. In both malingering and factitious disorders, the behaviors are "put on" only when being observed, though they may require some preparation beforehand. The behaviors are useful only when the goals can be obtained from interacting with others. Boone's (2007, 2009) differentiation between self-deception (somatoform disorders) and other-deception (malingering; factitious disorder) is helpful. In malingering, the behaviors are "put on" only when being observed, but with self-deception they might occur continuously. Such bright-line distinctions are frequently more complicated in real life, and Lamberty (2008) has made the point that in more complex cases, comorbidity is likely.

Except in cases of child abuse in which examinations may be court- or agency-ordered, the clinician is more likely to see somatoform or factitious disorders in a clinical context rather than a medicolegal context. As Chafetz (2011a) has suggested, factitious and somatoform disorders are not likely to be the forms of noncredible presentation in the psychological consultative examination for SSD. Social Security claimants are aware of the burdens of proving their disabilities to the DDS. This proof requires much deliberation with many hurdles, including undergoing an examination in which cognitive or psychological problems must be "marked" before the DDS will consider the impairments severe enough to meet various listing requirements. Frequently, the help of specialized disability attorneys is sought. Coupled with noncredible findings, this deliberate behavior is more aptly placed under the rubric of malingering rather than in any other noncredible behavior category, as the goal orientation makes the nature of the behavior clear.

MALINGERING BY PROXY

Munchausen syndrome by proxy occurs when a parent creates illness in a child, causing the child to produce symptomatology that serves the parent's interests (Chafetz & Prentkowski, 2011). The focus is on the parent suffering the sick role by proxy, although in custody cases the parent is frequently seen attempting to prove herself as an especially attuned caregiver.

The term "malingered neurocognitive dysfunction by proxy" was coined by Slick et al. (1999) to describe an individual in an evaluation responding noncredibly from direction or pressure by others for compensatory gain. The nature of the gain is what distinguishes the behavior from Munchausen by proxy behaviors, which serve psychological needs. Moreover, although this chapter focuses on children, there is no age limit to malingering by proxy, as spouses in litigation have been observed to engage in the same behaviors.

Malingering by proxy is not a new concept. Kompanje (2007) presented such a case, reported in 1593 by the renowned surgeon Guilhelmius Fabricius Hildanus. Parents had drilled an opening through the top of their child's head and inserted a small pipe through which they inflated air, enlarging their child's head. The parents concealed the opening with wax. For considerable income, the parents presented their child from town to town as a monster. Because of the secondary gain, this case is considered as malingering by proxy rather than Munchausen syndrome by proxy, even though a physical deformity was actually produced. The parents were sentenced to pay with their lives.

Lu and Boone (2002) presented a litigation case involving a 9-year-old who failed four specialized validity tests designed to detect noncredible performance. This child also showed atypical performance patterns on many standard cognitive measures. McCaffrey and Lynch (2009) discussed the case of a 13-year-old female who had documented brain injury but who also showed noncredible behavioral changes.

In a case presented by Chafetz and Prentkowski (2011), a 9-year-old boy referred by the local DDS failed the Test of Memory Malingering

(TOMM; Tombaugh, 1996), obtaining scores that were significantly below chance ($p < .01$) as determined on the binomial distribution. He also obtained a scaled score of 1 on all Wechsler Intelligence Scale for Children—Third Edition (WISC-III) subtests, giving him a Full-Scale IQ of 40 (< 0.1 percentile; 95% confidence interval = 37–48). If this IQ score were an accurate assessment of his true abilities, it is not likely that he would have been able to disobey his mother as described, to appreciate directions on assessment, or to have the social interactions noted in the evaluation. There were numerous other inconsistencies noted in this examination, which was clearly a case of intentional deception. Considering the development of intentional deception, Oldershaw and Bagby (1997) noted that children as young as 2½ years are capable of lying but that they have to be at least 3 before they understand the concept of a false belief and about 9 before they can successfully fool others. Although the child in this case example employed deceptive strategies, we did not suggest that he did this on his own; rather, it was more likely done at the urging and coaching of his parents.

MALINGERING BY PROXY AS CHILD ABUSE

What are the consequences of these coaching strategies by parents? Although these behaviors do not involve drilling holes in the child's head as in the Kompanje (2007) example from the late 1500s, the consequences may still be severe (Chafetz & Prentkowski, 2011).

Assuming that the deceptive strategies were successful, which has been made more likely by SSA's stance against the use of invalidity detection methods (Chafetz, 2010, 2011a), the children involved then become part of the engine of their parents' economies. Thus this behavior is not a one-time event. There becomes a need to play a role for the continuation of benefits, as the DDSs will reevaluate periodically to determine whether there has been any improvement in the child's condition and progress in school.

In a poster at the 2011 American Academy of Clinical Neuropsychology (AACN) scientific session, Chafetz and Binder (2011) discussed a case of malingering by proxy in which a child had clearly suffered and failed to progress in the 6 years since she was evaluated for the DDS at age 11. An argument can be made that using children in this way is a form of child abuse (Chafetz & Dufrene, 2014).

The idea of fabricating illness as abuse is clearly recognized in cases of Munchausen syndrome by proxy (Stirling & the Committee on Child Abuse and Neglect, 2007). No matter what the motivation of the parents, the most important issue discussed by child abuse experts is the harm done to the children (Flaherty & MacMillan, 2013), and thus both Munchausen and malingering by proxy as child abuse would be covered under the new rubric of medical child abuse (Stirling & the Committee on Child Abuse and Neglect, 2007).

The Chafetz and Binder (2011) case was an 11-year-old child seen for a consultative examination for SSI when she was 11 and again for the state rehabilitative service when she was 17 years old. Her malingering was so egregious both times that she failed validity testing at significantly below-chance levels. When seen for vocational help, the mother was resistant to talking about her disability compensation, which is unusual. In the 6 intervening years since the consultative examination for the DDS, the child had not progressed in school and was earning Fs even while in special education. She was depressed and behavior-disordered, and she did no chores at home. Although the reporting may be suspect in a malingering by proxy case (Chafetz & Binder, 2011), the child clearly had not developed resources and could not read or perform simple arithmetic, even though developmentally she had walked and talked on time. Chafetz and Dufrene (2014) discuss these cases as prompting a need for reporting by alert evaluators to protect children from the attendant educational neglect.

WHAT ARE THE BASE RATES?

In the article that initially discussed the use of the Symptom Validity Scale (SVS) for Low-Functioning Individuals in adults and children (Chafetz, Abrahams, & Kohlmaier, 2007), 96 children whose parents were seeking disability on their behalf were included in the TOMM study, and 27 children were included in the study involving the Medical Symptom Validity Test (MSVT; Green, 2004). In the TOMM study, the children were administered the WISC-III as part of the referral. When the DDS shifted to the use of the WISC-IV, the MSVT was then used for validity testing.

Chafetz (2008), in an article on base rates, systematically showed rates of validity test failure in adults and children in these disability samples, along with corresponding levels of test performance. Various "doses" of effort were considered as a means of dividing the sample. The Definite Malingering group was made up of children (of the ages for WISC-III and IV) who failed the TOMM or the MSVT at significantly below-chance levels, according to the binomial distribution. Children were classified as chance-level performers when they failed the TOMM or MSVT within chance levels, again delineated by using the binomial distribution. To achieve a chance-level performance, the children's abilities must be so impoverished that they are unable to act with any recognition memory on the stimuli, performing as if blindfolded. Both the TOMM and MSVT manuals make clear that even with significant impairment, an individual is able to marshal some recognition memory to perform better than chance levels, and thus chance-level performance typically shows behavior inconsistent with abilities. The Fail-2 category in adults (failure of the SVS plus TOMM or MSVT, depending on the study) was a marker category for the Slick et al. (1999) criteria for probable malingering. The Fail-1 category indicated some suspicion of noncredible performance, though at a level at which malingering was not necessarily probable. In children, Fail-1 and Fail-2 were combined due to low sample size. There was also a category of Fail Indicators, in which the failure of a validity indicator or two was not sufficient to gain entry into a more serious category (i.e., Fail 1 validity test), and a Not Fail category in which the claimants passed all validity tests and indicators. Unlike adults, in children the sample size was not sufficient to retain all categories, and Table 14.1 illustrates which ones were combined. Table 14.1 also illustrates the results for children in the base-rate study (Chafetz, 2008).

Table 14.1 shows that 10% of children in the larger TOMM study failed the TOMM at below-chance levels, considered the "smoking gun of intent" by Pankratz and Erickson (1990). This rate is only somewhat below that of adults (12–13%) who show definite malingering in the same study. About 16% of children in the TOMM study obtained chance-level performance on the TOMM, and about 20% of children in the MSVT study obtained chance-level or below-chance performance on the MSVT. About 28–34% of children failed one or two PVTs. As these rates are distinct for the categories, we see that children in these disability samples show cumulatively 48–60% rates of some evidence for malingering. About 25–32%

TABLE 14.1. Failure Rate and IQ Data for Children in the Chafetz (2008) Base-Rate Study

	TOMM study		MSVT study	
Group	% (n)	Mean IQ ± SD	% (n)	Mean IQ ± SD
Definite malingering	10.0 (8)	43.9 ± 5.3	—	—
Chance–below-chance level	16.3 (13)	51.8 ± 9.4	20.0 (5)	48.4 ± 7.7
Fail 1 or 2 PVTs	33.8 (27)	57.9 ± 13.4	28.0 (7)	65.7 ± 6.3
Fail indicators	25.0 (20)	71.4 ± 10.1	32.0 (8)	72.1 ± 8.1
No validity problems	15.0 (12)	75.3 ± 9.6	20.0 (5)	75.2 ± 11.7

Note. TOMM, Test of Memory Malingering; MSVT, Medical Symptom Validity Test. In both the TOMM and MSVT studies, due to small sample size, the Fail 1 and Fail 2 PVT categories were combined: Fail 1 or 2 PVTs. In the TOMM study, Definite Malingering refers to significantly below-chance performance on the TOMM and is distinct from the Chance-level group. In the MSVT study, due to low sample size, Definite and Chance-level groups were combined. There are 11 indicators in the SVS. If claimants failed one or two of these but did not fail the SVS, they would be added to the Fail Indicators group. If they had failed the SVS or the PVT (TOMM or MSVT, depending on the study), they would be added to the Fail 1 group. If they failed the SVS and the PVT (TOMM or MSVT), they would be in the Fail 2 group.

of children seeking disability fail indicators at a level that does not meet criteria for a determination of malingering, and 15–20% of children do not fail any PVTs or indicators. Cumulatively, 40–52% of children seeking disability do not meet criteria for malingering. The percentage of child disability claimants not meeting criteria for malingering may actually be somewhat higher if it is considered that failing one PVT is not sufficient to meet criteria. Thus some unknown proportion of the groups failing one or two PVTs would be added to the 40–52% figures.

COSTS OF MALINGERING IN CHILDREN

For adult SSI claimants, based on the 2011 SSI Annual Statistical Report (Social Security Administration, 2012) and using the most widely accepted base rate of malingering, Chafetz and Underhill (2013) calculated that $20.02 billion was spent on disability benefits for malingered mental disorders. Although this amount is staggering, it pales when one considers the costs of malingering among all disability programs (including those for children) and for all kinds of costs (e.g., Medicare and Medicaid benefits), which were calculated by Chafetz (2011a) at approximately $180 billion in a single year. Using the same 2011 SSI Annual Statistical Report and considering children under the age of 18 but subtracting all children under the age of 6, I now report that approximately $2.13 billion is spent in the SSI program on children who (with their parents) are probably malingering the evaluations. This figure takes into account the $5.32 billion spent on SSI benefits and the most widely held base rate of probable malingering of 40%. If one restricts the calculation to only those children (10%) who show the most egregious levels of malingering (Chafetz, 2008), the costs are about $532 million in that single year. These figures are restricted only to the mental health benefits.

LOW IQ AND MALINGERING

Both child and adult claimants for Social Security Disability benefits typically have low IQ (Chafetz et al., 2007). As can be seen in Table 14.1, Wechsler IQ levels closely follow the graded effort groupings. Indeed, Chafetz et al. (2007) reported that the correlation between the SVS total (composite) score and Full-Scale IQ in adults was $r = -.83$ ($p < .001$) in the TOMM study and $r = -.83$ ($p < .001$) in the MSVT study. In children, the correlation was $r = -.75$ ($p < .001$) in both the TOMM study and the MSVT study.

As correlations are bidirectional, they raise the question of whether IQ drops in the child samples because of poor quality of effort or whether the scores on validity tests (in this case, the SVS composite) change because of low cognitive abilities.

To answer this question, it is helpful to note (as shown in Table 14.1) that the first and last effort groups are anchored by the labels Definite Malingering (significantly below chance) and No Validity Problems. Using the TOMM study in which the sample size is larger, the children who showed below-chance performance on a PVT obtained a mean IQ of 43.9 ± 5.3. Although this performance appears to be in the moderate range of intellectual disability, these children clearly attempted to lower their performance, obtaining significantly below-chance scores on a PVT.

In contrast, children who failed no PVT or indicator obtained a mean IQ of 75.3 ± 9.6. In the regression equation that characterizes the relationship in the child TOMM study between IQ and the SVS composite score, the intercept is close to this mean level: Full-Scale IQ predicted = –1.62 × SVS Total + 74.6. Thus, when the SVS total = 0, indicating no evidence for effort problems, the predicted IQ is at the intercept of 74.6 (~5th percentile). With each point of increasing effort problems on the SVS, the Full-Scale IQ drops 1.62 points. When the effort problems are so egregious that the child is in the Definite Malingering range, the IQ is in the moderate range of intellectual disability.

For example, the 9-year-old child in the Chafetz and Prentkowski (2011) study obtained a Full-Scale IQ score of 40 on the WISC-III (all 1s on the subtest scaled scores), with a composite score of 23 on the SVS. He also failed the TOMM at significantly below-chance levels. To obtain a score this high on the SVS, several indicators had to be failed, which gives a high posterior probability of 99+% (Chafetz, 2011b). The child missed 10 out of 11 total items on the SVS, including his choice of a 19th-century Louisiana author, Lafcadio Hearn, as president of the United States, his inability to do a two-digit span on the WISC-III subtest, and not knowing his birthday. This example illustrates the effect on IQ of poor quality of effort at the extreme.

Other findings from the base-rate study (Chafetz, 2008) also suggest the directionality of the relationship between IQ and effort, with poor effort lowering IQ. Table 14.2 shows other performance effort variables in the TOMM part of the base-rate study, including the TOMM variables, the SVS total, A-Test errors (Strub & Black, 1993), and Reliable Digit Span (RDS; Greiffenstein, Baker, & Gola, 1994).

To develop evidence for the base rates, the TOMM had to be used in large part to define the groups, and thus it is no surprise to find the large separation among TOMM scores between the various effort doses. The SVS, on the other hand, was used to define the Fail 1 or 2 PVTs group. Although its scores did not participate in the poorer effort dosage groups, its components did participate in the Fail Indicators and Not Fail groups.

TABLE 14.2. Performance Validity Test Means and Standard Deviations from the TOMM Part of the Chafetz (2008) Base-Rate Study

Group	TOMM-I	TOMM-II	SVS	A-Test	RDS
Definite Malingering	16.1 (8.1)	7.5 (3.5)	17.9 (5.7)	11.1 (6.9)	1.8 (2.4)
Chance	26.0 (2.8)	24.1 (4.0)	12.2 (7.4)	6.0 (3.8)	5.8 (2.9)
Fail 1 or 2 PVTs	41.2 (7.2)	47.7 (3.1)	10.2 (4.9)	3.5 (7.9)	5.3 (2.0)
Fail indicators	46.5 (3.7)	49.6 (0.9)	3.2 (1.4)	1.0 (1.2)	6.2 (1.0)
No validity problems	45.3 (5.7)	49.6 (0.8)	0.8 (0.8)	0.6 (0.5)	7.8 (1.2)

Note. TOMM, Test of Memory Malingering; SVS, Symptom Validity Scale for Low Functioning Individuals; A-Test, "A" Random Letter Test of Auditory Vigilance; RDS, Reliable Digit Span.

Thus it is relevant and interesting to see a mean composite SVS score of almost 18 in the Definite Malingering group. A score this high indicates failure of multiple indicators on the SVS. In the chance-level group, the mean failure is about 12, also indicating failure of multiple validity indicators, though to a lesser extent. Again, this finding shows multiple validity indicator failure in the two most egregious groups.

The A-Test failure levels are also illustrative. The A-Test is independent of all the egregious failure groups and the Fail 1 or 2 PVTs group, though its scores participated in the Fail Indicators and the No Validity Problems groups. As one can see in Table 14.2, children in the Definite Malingering group obtain a mean of about 11 errors on this simple test. In the chance-level group, the error rate is about 6 errors. When one or both of the main validity tests (TOMM or SVS) is failed, the A-Test error rate is 3.5 errors. When only a couple of indicators are failed at a nonsignificant level for the failure of the SVS, the children produce only 1 error on the A-Test, which is below the adult cutoff of > 2 errors. In the No Validity Problems group, when no other validity test or indicator is failed, the failure level mean on the A-Test is less than 1 error.

RDS is a component of the SVS, albeit not directly, as its score is converted into a scale on which a score of < 7 adds 1 point to the SVS and a score of < 6 adds 2 points. Thus RDS does not participate in the two most egregious validity groups. As Table 14.2 shows, children who obtain a below-chance score on the TOMM obtain a mean RDS of 1.8. However, the chance level and the Fail 1 or 2 levels are not well discriminated by the RDS, with respective means of 5.8 and 5.3. When children fail some indicators (though not at a meaningful level), the mean RDS is 6.2. When children do not fail validity tests or indicators, the mean RDS is 7.8. Incidentally, the mean IQ of these children in this No Validity Problems group, as reported earlier, is 75.3. Similarly (not shown), children in the MSVT study who fail at chance or below-chance levels (groups combined) obtained a mean RDS of 3.6, but in the Fail 1 or 2 and the Fail Indicators group the RDS means are the same at 6.3. In the No Validity Problems group, the mean is 8.2.

In this context, it is interesting that Blaskewitz, Merten, and Kathmann (2008) reported that second, third, and fourth graders in their control group obtained mean RDS scores of 6.9, 7.3, and 8.0, with low percentages of these full-effort children passing the adult cutoff of < 8. Thus, in the base-rate study (Chafetz, 2008), the lack of separation in the validity dosage groups that are less egregious than the below-chance group may be due to some confounding with lower scores in impaired younger children. After all, RDS does have both validity and ability components. Nevertheless, a cutoff of < 6 appears to separate children who are showing significant evidence for malingering from children who are not, though classification accuracy statistics were not reported in this base-rate study.

These findings suggest that the directionality of the relationship is that quality of effort has a meaningful effect on neurocognitive levels ($r^2 = .56$), consistent with studies in adults showing that effort has a larger effect on neurocognitive variables than moderate to severe traumatic brain injury (Green, Rohling, Lees-Haley, & Allen, 2001; Ord, Greve, Bianchini, & Aguerrevere, 2010).

CLASSIFICATION ACCURACY

Although the base rates and directionality of these relationships are important to understand, decision making in individual cases requires knowledge of classification accuracy (Chafetz, 2008). The dataset on children (Chafetz, 2008) examined in this chapter permits a classification accuracy criterion-groups study, presented here.

Probable malingerers were selected with a criterion of the SVS composite score > 8. In the original study showing the development of the SVS (Chafetz et al., 2007), an SVS total score > 8 showed 100% specificity and 100% sensitivity when using the MSVT to define chance-level performance or below (immediate recognition [IR] or delayed recognition [DR] ≤ 70%) in children. As described, scores in the chance range are as easily obtained if the claimant had been blindfolded, and below-chance scores betray true knowledge of the correct choices. In a later study in adults, it was recognized that SVS scores > 8 indicate failure of at least three validity indicators, which has a high posterior probability in the aggregate for identifying malingering (Chafetz, 2011b).

As described, the A-Test is an auditory continuous performance test easily administered during a mental/cognitive status examination (Strub & Black, 1993). An A-Test error total > 2 errors was shown in adult disability claimants to have specificity of > 90% with sensitivity of about 72% in a classification accuracy study of malingering (Chafetz, 2012). In adults with IQs of 56–75 who have clearly defined motivation to perform well, the A-Test is not failed (Chafetz & Biondolillo, 2012). In this chapter, the base-rate studies in children (Chafetz, 2008) were reviewed, showing a close association between the number of A-test errors and the "dosage" levels of malingering, with about 11 A-Test errors in Definite malingerers. As suggested by the cutoff of 2 errors in adults, the A-Test is a simple test in which those who are showing good effort may have an occasional lapse. In this chapter, it is seen that > 2 errors on the A-Test in children, judged against a malingering criterion of SVS > 8, has a sensitivity of 67.7% and a specificity of 91.8%. Thus, in child disability claimants as well as in adults, the A-Test with a cutoff of > 2 errors (of either kind) is seen as an adequate instrument for the detection of malingering.

Now considering the TOMM, as judged against a criterion of > 8 cutoff on the SVS, a classification accuracy examination was performed to determine the highest cutoff that has adequate specificity (≥ 90%). In these child disability claimants, a TOMM Trial II score of < 42 has the minimum specificity (89.6%; effectively 90%), with a sensitivity of 50%. Restraining the cutoff to 41 or 40 changes neither the specificity nor the sensitivity in this sample of children claiming disability. When using the TOMM I plus TOMM II scores, a cutoff of < 74 has the minimum adequate specificity (89.6%) with 50% sensitivity.

In a review by DeRight and Carone (2013), it was acknowledged that freestanding recognition memory validity tasks have shown adequate specificity at adult cutoffs in children, but embedded validity tasks (such as RDS) tend to be less specific at adult cutoffs for many reasons, frequently having to do with an ability component. DeRight and Carone (2013) acknowledged the collection of embedded measures in the SVS as an exception in children and acknowledged the utility of the A-Test in children.

DECISION MAKING CONCERNING CHILD CLAIMANTS

The evidence concerning classification accuracy of validity tests useful in a child disability sample can be used directly for decision making about whether the child has indeed malingered the examination. Before the child is tested, the examiner knows that there is a probability of malingering. This knowledge is obtained from base rates. In child disability claimants, Chafetz (2008) found that children fail validity testing at chance or below-chance levels at rates of 20–26%. Child disability claimants additionally fail one or two PVTs at rates of 28–34%. Unfortunately, due to low sample sizes, the Fail 1 and Fail 2 categories had to be combined, whereas in adults the categories of Fail 1 or Fail 2 PVTs were distinct. If we assume that the rate of Fail 2 PVTs in children would be somewhere in between the range of the Fail 1 or Fail 2 combined category, the total rates of below-chance plus chance levels plus Fail 2 PVTs would be approximately 20–48% in one of the child disability studies and 26–60% in the other child disability study.

Larrabee, Millis, and Meyers (2009), reviewing numerous medicolegal studies, derived a rate of malingering in adult medicolegal samples of 40% ± 10%. The child disability data cited in this chapter certainly fall within this range, and thus one can suggest a pretest base rate (BR) of malingering of 40%. When a child claimant arrives, before any testing is performed, there is a likelihood of 40% that the child will be malingering in the disability examination. It is noted that this probability is below the lowest legal standard, a preponderance of the evidence (at least 51%), which is sometimes termed "more likely than not."

What if the child then fails the A-Test by producing more than two errors? The likelihood ratio (LR) is formed by dividing a test's sensitivity by the false-positive rate. In the case of the A-Test, the LR is .677/(1 – .918) = .677/.082 = 8.256. Multiplying the LR (8.256) by the pretest odds—BR/(1 – BR) = (.4/.6) = .667—gives a value of the posttest odds of 5.507. The posterior probability is then derived by odds/(odds + 1) = (5.507/6.507) = .846. Thus, after failing the A-Test, the child now has a probability of malingering of about 85%. The rationale for these calculations can be reviewed in Chafetz (2011b).

If one accepted that the base rate of child disability malingering was only 30%, the posterior probability after failing the A-Test would be (8.256)(.3/.7) = 3.53 (posttest odds), and (3.53/4.53)(100) = 78%.

The aggregation (by chaining of likelihood ratios) of validity test results requires independence of validity tests in nonmalingering samples (Chafetz, 2011b; Larrabee, 2008). Although the independence of these tests was not calculated in this child sample, it is noted that in adult disability claimants who are not malingering, these validity tests are independent (Chafetz, 2011b). Due to high floors in nonmalingerers (e.g., A-Test nonfailures having 0, 1, or 2 errors, with most children having no errors), it is likely that these tests are independent in nonmalingerers. Thus the chaining method is shown.

With the chaining method, the posterior probability after failure of one test becomes the pretest probability before entering the result of the second test. Thus, if the child fails the A-Test, the prior probability (before considering other tests) is now 85% (with a 40% base rate). Now, if the child claimant also fails the TOMM Trial II, the new chained calculation is pretest odds = (.85/.15) = 5.667. The LR for the TOMM in this child disability sample is .5/(1 – .9) = 5.0. Multiplying the pretest odds (5.667) by the LR (5.0) gives posttest (chained) odds of 28.335. The subsequent posterior probability is then (28.335/29.335)(100) = 97%. Thus, after failure of both the A-Test and the TOMM, the child disability claimant would be determined to have a 97% probability of malingering. After a failure of two established validity tests, the examiner is much more certain of the child's validity status and can so inform the DDS.

SUMMARY

A brief history of Social Security Disability legislation is provided, with President Franklin Delano Roosevelt signing the Social Security Act into law on August 14, 1935, President Dwight Eisenhower signing the new disability insurance program into law on August 1, 1956, and President Richard Nixon signing into law the needs-based SSI program on October 30, 1972.

To be "disabled" according to the SSI program, a child must meet the listing requirements for the various mental impairments thought to be causing the disabling condition. Mere diagnosis is not enough, and the disorder must cause impairment to a degree that meets the requirements for marked limitations in functioning. The psychological consultative examination is one way that the DDS gets evidence about a child's impairments and limitations. Psychologists should realize the boundary conditions concerning these claimants and that their motivations and concerns might differ greatly from those of patients who usually come into the office. In this scenario, noncredible behavior is thought to be a product of other-deception, rather than self-deception.

The problem of malingering by proxy is explored, and examiners are encouraged to understand the devastating effects on a child that can be produced by co-opting the child into a parent's financial gain system. Guidance for reporting this form of abuse can be found in Chafetz and Dufrene (2014).

The special problem of low IQ and failure on performance validity testing is discussed, with evidence presented that sufficient specificity can be found on some performance validity tests when children with low IQ are examined. Classification accuracy statistics for two performance validity tests on child SSI claimants are presented, along with calculations that can be used to guide decision making concerning the validity of these examinations.

REFERENCES

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: Author.

American Psychological Association. (2002). Ethical principles of psychologists and code of conduct. *American Psychologist, 57*, 1060–1073.

Blaskewitz, N., Merten, T., & Kathmann, N. (2008). Performance of children on symptom validity tests: TOMM, MSVT, and FIT. *Archives of Clinical Neuropsychology, 23*, 379–391.

Boone, K. B. (2007). A reconsideration of the Slick et al. (1999) criteria for malingered neurocognitive dysfunction. In K. B. Boone (Ed.), *Assessment of feigned cognitive impairment* (pp. 29–49). New York: Guilford Press.

Boone, K. B. (2009). Fixed belief in cognitive dysfunction despite normal neuropsychological scores: Neurocognitive hypochondriasis? *The Clinical Neuropsychologist, 23*, 1016–1036.

Chafetz, M. D. (2008). Malingering on the Social Security Disability consultative exam: Predictors and base rates. *The Clinical Neuropsychologist, 22*(3), 529–546.

Chafetz, M. D. (2010). Symptom validity issues in the psychological consultative examination for Social Security Disability. *The Clinical Neuropsychologist, 24*, 1045–1063.

Chafetz, M. D. (2011a). The psychological consultative examination for Social Security Disability. *Psychological Injury and Law*, *4*, 235–244.

Chafetz, M. D. (2011b). Reducing the probability of false positives in malingering detection of Social Security Disability claimants. *The Clinical Neuropsychologist*, *25*, 1239–1252.

Chafetz, M. D. (2012). The A-Test: A symptom validity indicator embedded within a mental status examination for Social Security Disability. *Applied Neuropsychology*, *19*, 121–126.

Chafetz, M. D., Abrahams, J. P., & Kohlmaier, J. (2007). Malingering on the Social Security Disability consultative examination: A new rating scale. *Archives of Clinical Neuropsychology, 22*, 1–14.

Chafetz, M. D., & Binder, C. (2011, June). Malingering-by-proxy: A case suggesting the need for policy regarding child neglect. *American Academy of Clinical Neuropsychology (AACN)*, Scientific Program, Washington, DC.

Chafetz, M. D., & Biondolillo, A. (2012). Validity issues in Atkins death cases. *The Clinical Neuropsychologist, 26*, 1358–1376.

Chafetz, M. D., & Dufrene, M. (2014). Malingering-by-Proxy: Need for child protection and guidance for reporting. *Child Abuse and Neglect, 38*(11), 1755–1765.

Chafetz, M. D., & Prentkowski, E. (2011). A case of malingering by proxy in a Social Security Disability psychological consultative examination. *Applied Neuropsychology, 18*, 143–149.

Chafetz, M. D., & Underhill, J. (2013). Estimated costs of malingered disability. *Archives of Clinical Neuropsychology, 28,* 633–639.

Delis, D., & Wetter, S. R. (2007). Cogniform disorder and cogniform condition: Proposed diagnoses for excessive cognitive symptoms. *Archives of Clinical Neuropsychology, 22,* 589–604.

DeRight, J., & Carone, D. A. (2013). Assessment of effort in children: A systematic review. *Child Neuropsychology, 21,* 1–24.

Flaherty, E. G., MacMillan, H. L., & Committee on Child Abuse and Neglect. (2013). Caregiver-fabricated illness in a child: A manifestation of child maltreatment. *Pediatrics, 132*, 590–597.

Green, P. (2004). *Medical Symptom Validity Test*. Edmonton, Alberta, Canada: Green's Publishing.

Green, P., Rohling, M. L., Lees-Haley, P. R., & Allen, L. A. (2001). Effort has a greater effect on test scores than severe brain injury in compensation claimants. *Brain Injury, 15*, 1045–1060.

Greenberg, S. A., & Shuman, D. W. (1997). Irreconcilable conflict between therapeutic and forensic roles. *Professional Psychology: Research and Practice, 28*, 50–57.

Greenberg, S. A., & Shuman, D. W. (2007). When worlds collide: Therapeutic and forensic roles. *Professional Psychology: Research and Practice, 38*, 129–132.

Greiffenstein, M. F., Baker, W. J., & Gola, T. (1994). Validation of malingered amnesia measures with a large clinical sample. *Psychological Assessment, 6*, 218–224.

Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., Millis, S. R., & Conference Participants. (2009). American Academy of Clinical Neuropsychology Consensus Conference Statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist, 23*, 1093–1129.

Kompanje, E. (2007). A case of malingering by proxy described in 1593. *Child Abuse and Neglect, 31*, 1013–1017.

Lamberty, G. (2008). *Understanding somatization in the practice of clinical neuropsychology*. New York: Oxford University Press.

Larrabee, G. J. (2007). *Assessment of malingered neuropsychological deficits*. New York: Oxford University Press.

Larrabee, G. J. (2008). Aggregation across multiple indicators improves the detection of malingering: Relationship to likelihood ratios. *The Clinical Neuropsychologist, 22*(4), 666–679.

Larrabee, G. J., Millis, S. R., & Meyers, J. E. (2009). 40 plus or minus 10, a new magical number: Reply to Russell. *The Clinical Neuropsychologist, 23*, 841–849.

Lu, P. H., & Boone, K. B. (2002). Suspect cognitive symptoms in a 9-year-old child: Malingering by proxy. *The Clinical Neuropsychologist, 16*, 90–96.

McCaffrey, R. J., & Lynch, J. K. (2009). Malingering following documented brain injury: Neuropsychological evaluation of children in a forensic setting. In J. E. Morgan & J. J. Sweet (Eds.), *Neuropsychology of malingering casebook* (pp. 377–385). New York: Psychology Press.

Oldershaw, L., & Bagby, R. M. (1997). Children and deception. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (2nd ed., pp. 153–166). New York: Guilford Press.

Ord, J. S., Greve, K. W., Bianchini, K. J., & Aguerrevere, L. (2010). Executive dysfunction in traumatic brain injury: The effects of injury severity and effort on the Wisconsin Card Sorting Test. *Journal of Clinical and Experimental Neuropsychology. 32*, 123–140.

Pankratz, L., & Erickson, R. D. (1990). Two views of malingering. *The Clinical Neuropsychologist, 4*, 379–389.

Puente, A. E. (1987). Social Security Disability and clinical neuropsychological assessment. *The Clinical Neuropsychologist, 1*, 353–363.

Rohling, M. L. (2004). Who do they think they're kidding?: A review of the use of symptom validity tests with children. *APA Division of Clinical Neuropsychology Newsletter, 22*(4), 20–26.

Slick, D. J., Sherman, E. M. S., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist, 13*(4), 545–561.

Social Security Administration. (2012). *SSI Annual Statistical Report, 2011* (SSA Publication No. 13-11827). Washington, DC: Author.

Social Security Administration. (2014). Disability evaluation under social security (SSA Publication No. 64–039). Retrieved from *www.ssa.gov/disability/professionals/bluebook/112.00-MentalDisorders-Childhood.htm#Top.*

Stirling, J., Jr., & the Committee on Child Abuse and Neglect. (2007). Beyond Munchausen syndrome by proxy: Identification and treatment of child abuse in a medical setting. *Pediatrics, 119*, 1026–1030.

Strasburger, L. H., Gutheil, T. G., & Brodsky, A. (1997). On wearing two hats: Role conflict in serving as both psychotherapist and expert witness. *American Journal of Psychiatry, 154*, 448–456.

Strub, R. L., & Black, F. W. (1993). *The Mental Status Examination in Neurology* (3rd ed.). Philadelphia: Davis.

Sweet, J. J., Wolfe, P., Sattleberger, E., Numan, B., Rosenfeld, J. P., Clingerman, S., et al. (2000). Further investigation of traumatic brain injury versus insufficient effort with the California Verbal Learning Test. *Archives of Clinical Neuropsychology, 15*, 105–113.

Tombaugh, T. N. (1996). *TOMM: Test of Memory Malingering.* North Tonawanda, NY: Multi-Health Systems.

Index

Note: f or t following a page number indicates a figure or a table.

N

R

T