Psychological Testing in the Service of Disability Determination

ISBN
978-0-309-37090-5

246 pages
6 x 9
PAPERBACK (2015)

Committee on Psychological Testing, Including Validity Testing, for Social Security Administration Disability Determinations; Board on the Health of Select Populations; Institute of Medicine

Psychological Testing in the Service of Disability Determination

Committee on Psychological Testing, Including Validity Testing,
for Social Security Administration Disability Determinations

Board on the Health of Select Populations

INSTITUTE OF MEDICINE
OF THE NATIONAL ACADEMIES

THE NATIONAL ACADEMIES PRESS
Washington, D.C.
www.nap.edu

THE NATIONAL ACADEMIES PRESS 500 Fifth Street, NW Washington, DC 20001

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competences and with regard for appropriate balance.

This study was supported by Contract/Task Order No. SS00-13-60048/0003 between the National Academy of Sciences and the U.S. Social Security Administration. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the organizations or agencies that provided support for the project.

International Standard Book Number-13: 978-0-309-37090-5
International Standard Book Number-10: 0-309-37090-6

Additional copies of this report are available for sale from the National Academies Press, 500 Fifth Street, NW, Keck 360, Washington, DC 20001; (800) 624-6242 or (202) 334-3313; http://www.nap.edu.

For more information about the Institute of Medicine, visit the IOM home page at: **www.iom.edu.**

Printed in the United States of America

The serpent has been a symbol of long life, healing, and knowledge among almost all cultures and religions since the beginning of recorded history. The serpent adopted as a logotype by the Institute of Medicine is a relief carving from ancient Greece, now held by the Staatliche Museen in Berlin.

Suggested citation: IOM (Institute of Medicine). 2015. *Psychological testing in the service of disability determination.* Washington, DC: The National Academies Press.

"Knowing is not enough; we must apply.
Willing is not enough; we must do."
—Goethe

INSTITUTE OF MEDICINE
OF THE NATIONAL ACADEMIES

Advising the Nation. Improving Health.

THE NATIONAL ACADEMIES

Advisers to the Nation on Science, Engineering, and Medicine

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Ralph J. Cicerone is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. C. D. Mote, Jr., is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Victor J. Dzau is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Ralph J. Cicerone and Dr. C. D. Mote, Jr., are chair and vice chair, respectively, of the National Research Council.

www.national-academies.org

COMMITTEE ON PSYCHOLOGICAL TESTING, INCLUDING VALIDITY TESTING, FOR SOCIAL SECURITY ADMINISTRATION DISABILITY DETERMINATIONS

HERBERT PARDES (*Chair*), Executive Vice Chairman of the Board, New York-Presbyterian Hospital, New York, New York

ARTHUR J. BARSKY III, Professor of Psychiatry, Harvard Medical School, and Vice Chair for Psychiatric Research, Brigham and Women's Hospital, Boston, Massachusetts

MARY C. DALY, Senior Vice President and Associate Director of Economic Research, Federal Reserve Bank of San Francisco, California

KURT F. GEISINGER, W. C. Meierhenry Distinguished University Professor of Educational Psychology and Director, Buros Center for Testing, University of Nebraska–Lincoln

NAOMI LYNN GERBER, University Professor, Center for the Study of Chronic Illness and Disability, George Mason University, Fairfax, Virginia

ALAN M. JETTE, Professor of Health Policy & Management, Boston University School of Public Health

JENNIFER I. KOOP, Associate Professor, Department of Neurology, Medical College of Wisconsin, Milwaukee

LISA A. SUZUKI, Associate Professor of Applied Psychology, New York University Steinhardt School of Culture, Education, and Human Development, New York, New York

ELIZABETH W. TWAMLEY, Associate Professor of Psychiatry, University of California, San Diego

PETER A. UBEL, Madge and Dennis T. McLawhorn University Professor of Business, Fuqua School of Business, and Professor of Public Policy, Sanford School of Public Policy, Duke University, Durham, North Carolina

JACQUELINE REMONDET WALL, Professor, School of Psychological Sciences, University of Indianapolis, Indiana, and Director, Office of Program Consultation and Accreditation, American Psychological Association, Washington, DC

Liaison to IOM Standing Committee of Medical Experts to Assist Social Security on Disability Issues

HOWARD H. GOLDMAN, Professor Psychiatry, University of Maryland School of Medicine, Baltimore

IOM Project Staff

CAROL MASON SPICER, Study Director
FRANK R. VALLIERE, Associate Program Officer
ALEJANDRA MARTÍN, Research Associate (since January 2015)
NICOLE GORMLEY, Senior Program Assistant (since December 2014)
JONATHAN PHILLIPS, Senior Program Assistant (April to November 2014)
JON SANDERS, Program Coordinator (through January 2015)
PAMELA RAMEY-McCRAY, Administrative Assistant
FREDERICK ERDTMANN, Director, Board on the Health of Select Populations

Reviewers

This report has been reviewed in draft form by individuals chosen for their diverse perspectives and technical expertise, in accordance with procedures approved by the National Research Council's Report Review Committee. The purpose of this independent review is to provide candid and critical comments that will assist the institution in making its published report as sound as possible and to ensure that the report meets institutional standards for objectivity, evidence, and responsiveness to the study charge. The review comments and draft manuscript remain confidential to protect the integrity of the deliberative process. We wish to thank the following individuals for their review of this report:

David Autor, Massachusetts Institute of Technology Economics
Leighton Chan, National Institutes of Health
Allen W. Heinemann, Northwestern University
Anita Hubley, University of British Columbia, Vancouver Campus
Michael Kirkwood, Children's Hospital Colorado-Aurora
Glenn J. Larrabee, Clinical Neuropsychology
Brian Levitt, Kaplan Psychologists
Patricia Owens, Health and Disability Policy Programs
Stephen M. Raffle, Forensic and Clinical Psychiatry
Jerry Sweet, University of Chicago, Pritzker School of Medicine

Although the reviewers listed above have provided many constructive comments and suggestions, they were not asked to endorse the conclusions

or recommendations nor did they see the final draft of the report before its release. The review of this report was overseen by **Nancy Adler,** University of California, San Francisco, and **Randy Gallistel,** Rutgers University. Appointed by the National Research Council and the Institute of Medicine, they were responsible for making certain that an independent examination of this report was carried out in accordance with institutional procedures and that all review comments were carefully considered. Responsibility for the final content of this report rests entirely with the authoring committee and the institution.

Preface

The U.S. Social Security Administration (SSA) disability programs provide important, sometimes vital, benefits to millions of adults and children annually in the United States. The programs are an expression of the nation's principle of caring for individuals who need support from the larger community. Within the confines of SSA policy, the state Disability Determination Services (DDS) agencies, which implement the policy, have the latitude to do so in whatever way they deem fit. It is not surprising that in a country as diverse as the United States we would find geographic variations in the style and methods with which that process is undertaken.

One element of such variation is the use or not of standardized psychological tests during the disability determination process, other than the use of intelligence tests in determinations of intellectual disability in children and adults. SSA asked the Institute of Medicine (IOM) to review selected psychological tests and to provide guidance on the use of psychological testing in SSA disability determinations.

SSA and the DDS agencies have the critical task of determining which applicants qualify for disability benefits, a task complicated by the lack of direct correlation between the presence of an impairment and disability, which SSA defines as the inability to work. DDS examiners undertake the very complex task of reviewing and developing applicants' files to determine which requests for disability benefits are justified. As described in the report, the committee felt that it was worth considering whether increased systematic use of standardized psychological testing in specific circumstances would strengthen the current process for disability determination.

The committee thanks colleagues, organizations, and agencies that were willing to share their expertise, time, and information during the committee's information-gathering meetings. The names of the speakers are included in the meeting agendas provided in Appendix A. The committee is grateful to the authors of the two commissioned papers, Erin Bigler, David Freedman, and Jennifer Manly, for the in-depth analyses they provided. The study sponsor, SSA, gladly provided information and data and responded to questions. We also thank Howard Goldman, chair of the IOM Standing Committee of Medical Experts to Assist Social Security on Disability Issues, who served as a consultant to the committee and provided valuable insight. The contributions from all of these sources informed the committee deliberations and enhanced the quality of this report.

I want also to pay tribute to and thank the expert members of our committee. A diversity of views, at times a difference of views, all contributed to generating a consensus about issues important to SSA and to the country. Throughout the project, they put in an enormous amount of time and effort; contributed their experience, knowledge, and perspective; listened to contending arguments; and ultimately generated the recommendations in this report. It is heartening to me and the other committee members to experience the excellence and the commitment of so many good colleagues. I trust this report will be helpful to and well received by SSA.

Finally, the committee thanks the IOM staff members who contributed to the production of this report, including Frederick "Rick" Erdtmann (board director), Carol Mason Spicer (study director), Frank Valliere (associate program officer), Alejandra Martín (research associate), Nicole Gormley (senior program assistant), Jonathan Phillips (senior program assistant), Jon Sanders (program coordinator), Julie Wiltshire (financial associate), and other staff of the Board on the Health of Select Populations and the IOM, who provided support. Research assistance was provided by Daniel Bearss, Rebecca Morgan, and Catherine van der List.

Herbert Pardes, *Chair*
Committee on Psychological Testing, Including
Validity Testing, for Social Security Administration
Disability Determinations

Contents

Boxes, Figures, and Tables

BOXES

FIGURES

TABLES

Acronyms and Abbreviations

AACN	American Academy of Clinical Neuropsychology
AADEP	American Academy of Disability Evaluating Physicians
ABCN	American Board of Clinical Neuropsychology
ABIME	American Boards of Independent Medical Examiners
ADL	activity of daily living
AFB	Ability-Focused Neuropsychological Test Battery
ALJ	administrative law judge
AMA	American Medical Association
APA	American Psychological Association
ASAPIL	Association for Scientific Advancement in Psychological Injury and Law
BDI	Beck Depression Inventory
BLS	Bureau of Labor Statistics
BPRS	Brief Psychiatric Rating Scale
BSI	Brief Symptom Inventory
BVMT-R	Brief Visuospatial Memory Test-Revised
C&P	compensation and pension
CASL	Comprehensive Assessment of Spoken Language
CBCL	Child Behavior Checklist
CDMI	Composite Disability Malingering Index
CE	consultative examination
CELF-4	Clinical Evaluation of Language Fundamentals-4

CFS	chronic fatigue syndrome
CIDI	Composite International Diagnostic Interview
CPP	Canada Pension Plan
CRPS	complex regional pain syndrome
CVLT-II	California Verbal Learning Test—second edition
DDS	Disability Determination Services
DIF	differential item functioning
DOM	Depression Outcomes Module
DSM	*Diagnostic and Statistical Manual of the American Psychiatric Association*
GAF	Global Assessment of Functioning Scale
GAO	Government Accountability Office
HVLT-R	Hopkins Verbal Learning Test-Revised
ICF	*International Classification of Functioning, Disability and Health*
ID	intellectual disability
IDES	Integrated Disability Examination System
IME	independent medical examination
IOM	Institute of Medicine
IQ	intelligence quotient
IRT	item response theory
MDI	medically determinable impairment
MEDCOM	U.S. Army Medical Command
M-FAST	Miller Forensic Assessment of Symptom Test
MINI	Mini International Neuropsychiatric Interview
MMPI	Minnesota Multiphasic Personality Inventory
MMY	Mental Measurements Yearbook
MRFC	Mental Residual Functional Capacity
MSVT	Medical Symptom Validity Test
NAN	National Academy of Neuropsychology
NIH	National Institutes of Health
NIM	Negative Impressionism
NPP	negative predictive power
NPRM	Notice of Proposed Rulemaking
NRC	National Research Council

OIDAP	Occupational Information Development Advisory Panel
OIS	Occupational Information System
OTSG	Office of the Surgeon General
P-3	Pain Patient Profile
PAI	Personality Assessment Inventory
PCE	psychological consultative examination
PDRT	Portland Digit Recognition Test
PHQ	Patient Health Questionnaire
POMS	Program Operations Manual System
PPP	positive predictive power
PTSD	posttraumatic stress disorder
PVT	performance validity test
RAVL	Rey Auditory Verbal Learning
RDS	Reliable Digit Span
RMFIT	Rey Memory for Fifteen Items Test
RMT	Recognition Memory Test
RMTF	Warrington Recognition Memory Test for Faces
SCAN	Schedule for Clinical Assessment in Neuropsychiatry
SCL-90-R	Symptom Checklist 90-Revised
SDM	single-decision-maker
SGA	substantial gainful activity
SIMS	Structured Inventory of Malingered Symptomology
SIRS	Structured Interview of Reported Symptoms
SSA	U.S. Social Security Administration
SSDI	Social Security Disability Insurance
SSI	Supplemental Security Income
SVT	symptom validity test
TBI	traumatic brain injury
TMJ	temporomandibular joint disorder
TOMM	Test of Memory Malingering
VA	U.S. Department of Veterans Affairs
VBA	Veterans Benefits Administration
VHA	Veterans Health Administration
WAIS	Wechsler Adult Intelligence Scale
WHO	World Health Organization
WISC	Wechsler Intelligence Scale for Children
WMS	Wechsler Memory Scale
WMT	Word Memory Test

Summary[1]

BACKGROUND

In 2012, the U.S. Social Security Administration (SSA) provided benefits to nearly 15 million disabled adults and children through two disability programs. The majority of beneficiaries, 8.8 million, received benefits through the Social Security Disability Insurance (SSDI) program for disabled individuals, and their dependent family members, who have worked and contributed to the Social Security trust funds. The remaining beneficiaries (4.9 million adults and 1.3 million children) received benefits through the Supplemental Security Income (SSI) program, which is a means-tested program based on income and financial assets for adults aged 65 years or older and disabled adults and children.

SSA disability determinations are based on the medical evidence and all evidence considered relevant by the examiners in an applicant's case record. Physical or mental impairments must be established by objective medical evidence consisting of medical signs and laboratory findings, which may include psychological tests and other standardized test results. SSA establishes the presence of a medically determinable impairment in individuals with mental disorders other than intellectual disability through the use of standard diagnostic criteria, which include symptoms and signs. Evidence for these mental impairment claims, as well as for many other categories of claims, such as those for certain musculoskeletal and connective tissue

[1] This summary does not include references. Citations to support text, conclusions, and recommendations made herein are provided in the body of the report.

conditions, relies less on standard laboratory tests than for some other categories of impairment.

SSA maintains a list of criteria for specific conditions that an applicant with one or more of those conditions must meet in order to receive disability benefits based solely on medical criteria. SSA currently requires psychological test results, specifically intelligence test results, in the listing criteria for intellectual disability in children and adults and in the criteria for cerebral palsy, convulsive epilepsy, and meningomyelocele and related disorders. SSA questions the value of purchasing psychological testing in cases involving mental disorders, other than for intellectual disability, and it does not require testing either to establish or to assess the severity of other mental disorders.

As noted, SSA indicates that *objective medical evidence* may include the results of standardized psychological tests. Given the great variety of psychological tests, some are more objective than others. Whether a psychological test is appropriately considered objective has much to do with the process of scoring. For example, unstructured measures that call for open-ended responding rely on professional judgment and interpretation in scoring; thus, such measures are considered less than objective. In contrast, standardized psychological tests and measures, such as those discussed in the report, are structured and objectively scored. In the case of non-cognitive self-report measures, the respondent generally answers questions regarding typical behavior by choosing from a set of predetermined answers. With cognitive tests, the respondent answers questions or solves problems, which usually have correct answers, as well as he or she possibly can. Such measures generally provide a set of normative data (i.e., norms), or scores derived from groups of people for whom the measure is designed (i.e., the designated population), to which an individual's responses or performance can be compared. Therefore, standardized psychological tests and measures rely less on clinical judgment and are considered to be more objective than those that depend on subjective scoring. Unlike measurements such as weight or blood pressure, standardized psychological tests require the individual's cooperation with respect to self-report or performance on a task. The inclusion of validity testing in the test or test battery allows for greater confidence in the test results. Standardized psychological tests that are appropriately administered and interpreted can be considered objective evidence.

As illustrated in Figure S-1, standardized psychological testing is one component of a full psychological assessment. Standardized psychological tests can be divided into *measures of typical behavior* and *tests of maximal performance*. Measures of typical behavior, such as personality, interests, values, and attitudes, may be referred to as *non-cognitive measures*. Tests of maximal performance ask people to answer questions and solve problems

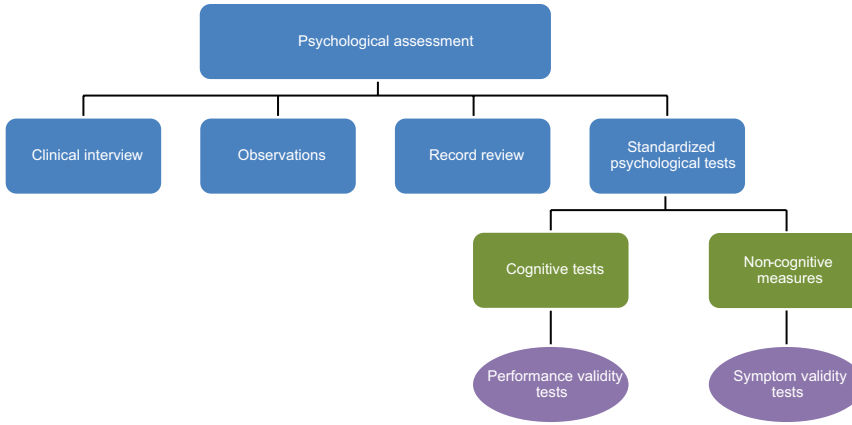


FIGURE S-1 Components of psychological assessment.
NOTE: Performance validity tests do not measure cognition but are used in conjunction with performance-based cognitive tests to examine whether the examinee is exerting sufficient effort to perform well and responding to the best of his or her capability. Similarly, symptom validity tests do not measure non-cognitive status but are used to examine whether a person is providing an accurate report of his or her actual symptom experience. Because cognitive tests frequently are performance based and non-cognitive measures generally involve self-report, performance validity tests and symptom validity tests are shown as being associated with these types of tests.

as well as they possibly can. Because tests of maximal performance typically involve cognitive performance, they are often referred to as *cognitive tests*. It is through these two lenses—non-cognitive measures and cognitive tests—that the committee examined psychological testing for the purpose of disability evaluation in this report. Intelligence tests and neuropsychological tests are examples of cognitive tests, while depression, anxiety, or personality inventories are examples of non-cognitive measures. Cognitive tests tend to be performance based, and non-cognitive measures tend to be based on self-report. Validity testing is an area of psychological testing. Performance validity tests (PVTs) provide information about an individual's effort on tests of maximal performance, such as cognitive tests. Symptom validity tests (SVTs) provide information about the consistency and accuracy of an individual's self-report of symptoms he or she is experiencing.

There are differences of opinion on the use of validity tests and their value for work disability evaluations. Current SSA policy precludes the purchase of validity tests as part of a consultative examination to supplement an applicant's medical evidence record, although applicants and their representatives sometimes submit validity test results in support of their claims.

BOX S-1
Statement of Task

An ad hoc committee will:

1. Perform a critical review of selected psychological tests, including symptom validity tests (SVTs), that could contribute to Social Security Administration (SSA) disability determinations;
2. Provide guidance on the general relevance and applicability of psychological tests, including SVTs, in the context of other relevant evidence to SSA disability determinations in claims involving physical and mental disorders; and
3. Provide guidance on how to use the results of psychological tests, including SVTs, in the context of disability determinations.

To accomplish these objectives, the committee shall consider the following topics: (1) use of psychological testing, (2) testing norms, (3) qualifications for administration of tests, (4) administration of tests, (5) reporting results, and (6) use of tests for the disability evaluation process.

Professional organizations of neuropsychologists and psychologists have issued position statements and guidance advocating for the use of validity tests in clinical and medicolegal contexts, and several have challenged SSA's institutional prohibition on ordering such tests. A September 2013 report from SSA's Office of the Inspector General concluded that although SSA does not allow the purchase of validity tests, "medical literature, national neuropsychological organizations, other federal agencies, and private disability insurance providers support the use of [validity tests] in determining disability claims."

It is within this context that SSA asked the Institute of Medicine (IOM) to convene a committee of relevant experts (e.g., adult and pediatric neuropsychology, psychology, psychiatry, disability medicine, behavioral economics, and economics) to review selected psychological tests, including validity tests, and to provide guidance on the use of such testing in the adjudication of claims submitted to the SSA Disability Programs (see Box S-1 for the statement of task). In carrying out this task, the Committee on Psychological Testing, Including Validity Testing, for Social Security Administration Disability Determinations was asked to address several specific topics, including testing norms, the administration of relevant tests and the qualifications for administering them, the interpretation and reporting of test results, and economic considerations.

COMMITTEE'S APPROACH TO ITS CHARGE

As part of its information-gathering process, the committee conducted an extensive review of the literature pertaining to the use of psychological tests, including PVTs and SVTs, in disability determinations. The committee supplemented its review of the literature with two public workshops to hear from neuropsychologists with expertise in performance validity and symptom validity testing in adults and children, the use of psychological and validity tests in culturally diverse populations, and the use of such tests in non-SSA disability determination contexts (e.g., private disability insurance programs, Canadian auto insurance, U.S. military disability or return-to-duty decisions, veterans' disability compensation). The committee also heard from SSA and Disability Determination Services (DDS) representatives about the SSA disability determination process and its current policies surrounding the use of psychological and validity testing. The committee commissioned two papers to provide additional critical analysis in areas relevant to the committee's work. The committee's work was further informed by previous IOM and National Research Council reports focused on different aspects of the SSA disability determination process.

COMMITTEE'S RECOMMENDATIONS

The committee identified three elements of SSA's disability determination process in which psychological testing could play a role: (1) identification of a "medically determinable impairment," (2) evaluation of functional capacity for work, and (3) assessment of the validity of applicants' psychological test results or the consistency of applicants' statements about self-reported symptoms. Although this report addresses all three elements, the committee focuses on the second and the third, for which questions about the use of psychological tests are more complex. As indicated in the following section, the committee found that the results of standardized psychological testing do provide information of value to each of the three elements.

Role of Psychological Testing in Social Security Administration Disability Programs

There currently is great variability in allowance rates for both SSI and SSDI among states that are not fully accounted for by differences in the populations of applicants. In addition, there is great variability in the disability determination appeal rulings among administrative law judges within and across states. Each state DDS agency, within the confines of SSA policy, issues its own rules regarding the tests that may be purchased as part of a consultative examination. Aside from the use of intelligence tests as

described in the listings for intellectual disability and certain neurological impairments, SSA does not require or specify the purchase of any type of (or individual) psychological test. SSA provides general guidance that good psychological tests are valid and reliable and have appropriate normative data. For this reason, there is variation among states about when and which standardized psychological tests can be purchased, with the exception of SVTs and PVTs, which are precluded from purchase by SSA except in rare cases such as a court order.

Although there currently are no data on the rates of false positives and false negatives in SSA disability determinations, systematic use of standardized psychological testing for a broader set of physical and mental impairments than is current practice is expected to improve the accuracy and consistency of disability determinations for applicants who allege cognitive impairment or whose allegation of functional impairment is based solely on self-report. The results of standardized cognitive and non-cognitive psychological tests that are appropriately administered, interpreted, and validated can provide objective evidence to help identify and document the presence and severity of medically determinable mental impairments at Step 2 of SSA's disability determination process. In addition, standardized cognitive test results can provide objective evidence to help identify and assess the severity of work-related cognitive functional impairment relevant to disability evaluations at the listing level (Step 3) and to mental residual functional capacity (Steps 4 and 5).

Current data on the prevalence of inconsistent reporting of symptoms or performing below one's capability on cognitive tests are very imprecise. In the context of SSA disability applicants, neither scenario rules out disability, but both suggest the need for additional assessment of the alleged impairment with the goal of making an accurate determination of disability. When a disability claim is based primarily on an applicant's self-report of symptoms and self-reported statements about their intensity, persistence, and limiting effects, SSA relies on an assessment of the consistency of the self-report with all of the evidence in the applicant's medical evidence record.

Although SSA's current policy precludes the purchase of SVTs and PVTs, these tests provide information about the validity of standardized non-cognitive and cognitive test results when administered as part of the test or test battery and therefore are an important addition to the medical evidence record in such cases. It is important that SVTs and PVTs only be administered in the context of a larger test battery and only be used to interpret information from that battery. Validity tests do not provide information about whether or not the individual is, in fact, disabled.

Standardized Non-Cognitive Psychological Measures and Symptom Validity Tests

The use of standardized non-cognitive psychological measures is essential to the determination of all cases in which an applicant's allegation of non-cognitive functional impairment meets each of three requirements:

- The applicant alleges a mental disorder (i.e., schizophrenic, paranoid, and other psychotic disorders; affective disorders; anxiety-related disorders; and personality disorders) unaccompanied by cognitive complaints or a disorder with somatic symptoms that are disproportionate to demonstrable medical morbidity (i.e., somatoform disorders, multisystem illnesses, and chronic idiopathic pain conditions).
- The presence and severity of impairment and associated functional limitations are based largely on applicant self-report.
- Objective medical evidence or longitudinal medical records sufficient to make a disability determination do not accompany the claim.

In certain instances, cognitive concerns may accompany the applicant's allegations, in which case cognitive testing, as discussed below, may be more appropriate. The committee recognizes there are a few chronic conditions (e.g., schizophrenia, chronic idiopathic pain, multisystem illnesses) that may generate potentially disabling, non-cognitive functional impairments but may not be accompanied by objective medical evidence. In such cases, the evidence provided by longitudinal medical records may be sufficient to substantiate the allegation.

Assessment of symptom validity, including the use of SVTs, analysis of internal data consistency, and other corroborative evidence, helps the evaluator to interpret the accuracy of an individual's self-report of behavior, experiences, or symptoms and responses on standardized non-cognitive psychological measures. For this reason, it is important to include an assessment of symptom validity when non-cognitive psychological measures are administered. Evidence of inconsistent self-report based on symptom validity measures is cause for concern with regard to self-reported symptoms but does not provide information about whether or not the individual is, in fact, disabled. A lack of validity on symptom validity testing alone is insufficient grounds for denying a disability claim, although additional information would be required to assess the applicant's allegation of disability.

Recommendation 1: The Social Security Administration should require the results of standardized non-cognitive psychological testing in the case record for all applicants whose claim of functional impairment relates either (1) to a mental disorder unaccompanied by cognitive complaints or (2) to a disorder in which the somatic symptoms are disproportionate to the medical findings. Testing should be required when the allegation is based primarily on applicant self-report and is not accompanied by objective medical evidence or longitudinal medical records sufficient to make a disability determination.

- All non-cognitive psychological evaluations should include a statement of evidence of the validity of the results, which could include symptom validity test results, analysis of internal data consistency (e.g., item response theory), and other corroborative evidence as well as discussion of the test norms relative to the individual being assessed.
- For cases in which validation is not achieved, SSA should pursue additional evidence of the applicant's allegation.

The committee intends *standardized non-cognitive psychological tests* to include measures of behavior, affect, personality, and psychopathology. By *objective medical evidence* in this and the following recommendation, the committee means medical signs and/or laboratory or test results that constitute clear objective medical evidence of a significant mental disorder and related functional impairment of sufficient severity to make a disability determination. An example would be a severe brain injury associated with significant functional deficits (e.g., minimally conscious state). By *longitudinal medical records* the committee means a documented history of a significant mental disorder or a chronic condition such as chronic idiopathic pain or multisystem illness and related functional impairment of sufficient severity and duration to make a disability determination. An example would be a well-documented history of repeated hospitalizations and treatments for a diagnosed mental disorder, such as an affective or personality disorder.

The committee intends the "statement of evidence of the validity of the results" specified in this and the following recommendation to reflect objective evidence that goes beyond the clinical opinion of the examiner. In addition to analysis of the results of SVTs or PVTs administered at the time of the testing and analysis of internal data consistency, evidence could include a pattern of test results that is inconsistent with the alleged condition, observed behavior, documented history, and the like. It is important to note that a finding of inconsistency between the test results and the areas specified is more informative than a finding of consistency would be.

The committee's recommendation here and in the following recommendation that SSA "pursue additional evidence of the applicant's allegation" for cases in which validation is not achieved means that the test results in those cases are an insufficient basis to make a determination regarding disability status.

Standardized Cognitive Tests and Performance Validity Tests

Standardized cognitive test results are essential to the determination of all cases in which an applicant's allegation of cognitive impairment is not accompanied by objective medical evidence. The results of cognitive tests are affected by the effort put forth by the test-taker. If an individual has not given his or her best effort in taking the test, the results will not provide an accurate picture of the person's neuropsychological or cognitive functioning. Performance validity indicators, which include PVTs, analysis of internal data consistency, and other corroborative evidence, help the evaluator to interpret the validity of an individual's neuropsychological or cognitive test results. For this reason, it is important to include an assessment of performance validity when cognitive testing is administered. It also is important that validity be assessed throughout the cognitive evaluation.

A PVT only provides information about the validity of an individual's cognitive test results that are obtained during the same evaluation. Evidence of invalid performance based on PVT results pertains only to the cognitive test results obtained and does not provide information about whether or not the individual is, in fact, disabled. A lack of validity on performance validity testing alone is insufficient grounds for denying a disability claim. In such cases, additional information is required to assess the applicant's allegation of disability.

Recommendation 2: The Social Security Administration should require the results of standardized cognitive testing be included in the case record for all applicants whose allegation of cognitive impairment is not accompanied by objective medical evidence.

- All cognitive evaluations should include a statement of evidence of the validity of the results, which could include performance validity test results, analysis of internal data consistency (e.g., item response theory), and other corroborative evidence as well as discussion of the test norms relative to the individual being assessed.
- For cases in which validation is not achieved, SSA should pursue additional evidence of the applicant's allegation.

Qualifications for Test Administration and Interpretation

Use of standardized procedures for the administration of standardized non-cognitive and cognitive psychological tests enables application of normative data to the individual being evaluated. Without standardized administration, the test-taker's performance may not accurately reflect his or her ability. It is important that any person administering cognitive or neuropsychological tests be well trained in the administration protocols for those particular tests, possess the interpersonal skills necessary to build rapport with the test-taker, and understand important psychometric properties, including validity and reliability, as well as factors that could emerge during testing to place either at risk.

Interpretation of standardized psychological test results is more than a report of the standardized test scores; it requires assigning meaning to the scores within the individual context of the specific examinee. As such, interpretation of test results requires a higher level of clinical training than does the administration alone of some psychological tests. Licensed psychologists and neuropsychologists are the specialists qualified to interpret the results of most standardized psychological and neuropsychological tests. Under close supervision and direction of licensed psychologists and neuropsychologists, it is standard practice for psychometrists or technicians with specialized training to administer and score tests. Test manuals specify the qualifications necessary for administration, scoring, and interpretation of the test or measure. It is important as well that the individual responsible for making the disability determination (disability examiner or administrative law judge) have the training and experience to understand and evaluate the report provided by the psychologist or neuropsychologist.

Recommendation 3: The Social Security Administration should ensure that psychological testing that is considered as part of a disability evaluation is performed by qualified specialists properly trained in the administration and interpretation of standardized psychological tests.

- "Qualified" means that the specialist must be currently licensed or certified to administer, score, and interpret psychological tests and have the training and experience to administer the test and interpret the results.
- This recommendation applies not only to standardized psychological testing that may be ordered in the course of a disability evaluation, but also to standardized psychological testing already in an applicant's medical evidence of record if the results are considered as part of the disability determination.

Economic Considerations

Systematic use of standardized psychological testing in SSA disability evaluations for a broader set of physical and mental impairments than is current practice will have financial implications. The average cost of testing services varies by the type of testing (e.g., psychological, neuropsychological), by the type of provider (e.g., psychologist or physician, technician), and by geographic area. The variation in pricing implies that the expected costs to SSA of requiring psychological testing will depend on exactly which tests are required, the qualifications mandated for testing providers, and the geographical location of the providers most in demand. Estimating the exact cost of broad use of psychological testing by SSA will require more detailed data on the exact implementation strategy.

At present, there do not appear to be any independently conducted studies regarding the accuracy of the disability determination process as implemented by DDS offices. Some published estimates of billions of dollars in potential cost savings to SSA associated with the use of symptom validity testing and performance validity testing are based on assumptions that if violated would substantially lower the estimated cost savings. Potential cost savings associated with testing vary considerably based on the assumptions about who it is applied to and how many individuals it detects and thus rejects for disability benefits. A full financial cost-benefit analysis of psychological testing will require SSA to collect additional data both before and after the implementation of the recommendations of this report.

Evaluation and Research

Based on its examination of the literature and dialogues with experts in a variety of areas, including psychological and neuropsychological testing, performance validity testing and symptom validity testing, and the disability evaluation process both within SSA and in other arenas, the committee recognizes many questions remain with regard to the use of standardized psychological testing in the disability determination process.

As part of its assessment of the use of standardized psychological tests for the disability evaluation process, the committee was asked to discuss the costs and cost-effectiveness of requiring a single test or a combination of tests. This report provides an initial framework for evaluating the economic costs and highlights the types of data that will be needed to accurately determine the financial impact of implementing the committee's first two recommendations. The following conclusions and recommendation relate to this enterprise.

Conclusions

- Accurate assessments of the net financial impact of psychological testing as recommended by the committee will require information on the current accuracy of DDS decisions and how the accuracy is affected by the increased use of standardized psychological testing.
- The absence of data on the rates of false positives and false negatives in current SSA disability determinations precludes any assessment of their accuracy and consistency.
- There currently is great variability in allowance rates for both SSI and SSDI among states that are not fully accounted for by differences in the populations of applicants. There also is great variability in the disability determination appeal rulings among administrative law judges within and across states. Although it is not possible to know definitively whether the large share of unexplained variation in state filing, award, and allowance rates is driven by variability in the federal disability determination process, there is some evidence that states differ in how they manage claims.
- In light of this unexplained variability, systematic use of standardized psychological testing as recommended by the committee is expected to improve the accuracy and consistency of disability determinations.

Recommendation 4: The Social Security Administration (SSA), in collaboration with other federal agencies, should establish a demonstration project(s) to investigate the accuracy and consistency of SSA's disability determinations with and without the use of recommended psychological testing.

- *Accuracy* refers to the rates of false negatives and false positives in SSA's disability determinations.
- *Consistency* means that adjudicators presented with the same evidence for comparable cases come to the same conclusion.

Recognizing that the costs and benefits of implementing the committee's recommendations go beyond the financial, the committee recommends that SSA evaluate the effect of implementing the committee's recommendations on its disability determination process using a number of different measures.

Recommendation 5: Following implementation of the committee's recommendations, the Social Security Administration should evaluate their impact on its disability determination process and end results. Measures of impact may include

- Number of backlogged cases;
- Efficiency of throughput or time to determination;
- Number of requests for appeals;
- Adherence to recommended evaluations;
- Effect on accuracy and consistency of disability determinations; and
- Effect on state-to-state variation in disability allowance rates and on appeal rulings among administrative law judges.

Over the course of the project, the committee identified two areas in particular in which it expects that the results of further research would help to inform disability determination processes as indicated in the following conclusions and recommendation.

Conclusions

- Additional research is needed on the use of SVTs and PVTs in populations representative of the pool of disability applicants, including in terms of gender, ethnicity, race, primary language, educational level, medical condition, and the like. In particular, additional research on the development of appropriate criterion or cutoff scores for PVTs and SVTs in these populations for the purposes of disability evaluation would be beneficial.
- The committee's task was to evaluate the usefulness of psychological testing in the disability determination process, as reflected in the foregoing recommendations. However, the committee recognizes that just as systematic use of standardized psychological testing is expected to improve the accuracy and consistency of disability determinations for applicants who allege cognitive impairment or whose allegation of functional impairment is based solely on self-report, the use of other standardized assessment tools also may be expected to improve the accuracy of disability determinations. The value of standardized assessment tools, including psychological tests, to assessments of individuals' work-related functional capacity is an area that would benefit from further research.

Recommendation 6: The Social Security Administration and other federal agencies should support a program of research to investigate the value of standardized assessment, including psychological testing, in disability determinations. Such a program should support original research on a variety of topics, including

- The effects of standardized psychological testing on the accuracy and consistency of disability determinations;
- The use of PVTs and SVTs with disability applicants; and
- The use of psychological tests, including PVTs and SVTs, in different populations with regard to fairness for members of all gender, ethnic, racial, language, educational levels, and other protected groups.

1

Introduction

The U.S. Social Security Administration (SSA) administers two disability programs: Social Security Disability Insurance (SSDI), for disabled individuals and their dependent family members, who have worked and contributed to the Social Security trust funds, and Supplemental Security Income (SSI), which is a means-tested program based on income and financial assets for adults aged 65 years or older and disabled adults and children (SSA, 2012a). Both programs require that applicants have a disability and meet specific medical criteria in order to qualify for benefits.

In 2012, SSA provided benefits to nearly 15 million disabled adults and children (see Table 1-1). The majority of beneficiaries, 8.8 million, received benefits through the SSDI program (SSA, 2013a, Table 20). The remaining beneficiaries received benefits through the SSI program; SSI paid benefits to 4.9 million adults and 1.3 million children (SSA, 2013b, Table 19).

Disability determinations are based on the medical evidence and all other evidence considered relevant by the examiners in a claimant's case record. Physical or mental impairments must be established by objective medical evidence consisting of medical signs and laboratory findings, which according to SSA may include psychological and other standardized test results (20 CFR § 404.1528). The presence of an impairment requires objective findings and cannot be based solely on an applicant's statement of symptoms and functional limitations, although such statements are treated as part of the overall evidence. SSA also considers the extent to which such self-reported claims of impairment and functional limitation are consistent with the observations by medical treating sources and collateral observers,

TABLE 1-1 Characteristics of SSDI and SSI Beneficiaries, 2012

Characteristic	SSDI Workers	SSI Adults—Disability	SSI Children
All	8,826,591	4,869,484	1,311,861
Age			
Under 30	2.50%	—	—
30–34	3.40%	—	—
35–39	4.60%	—	—
40–44	7.10%	—	—
45–49	11.00%	—	—
50–54	17.20%	—	—
55–59	23.20%	—	—
60–FRA	31.00%	—	—
18–21	—	7.49%	—
22–25	—	7.24%	—
26–29	—	6.43%	—
30–39	—	14.54%	—
40–49	—	20.07%	—
50–59	—	31.40%	—
60–64	—	12.83%	—
Under 5	—	—	14.90%
5–12	—	—	51.30%
13–17	—	—	34.00%
Gender			
Male	52.18%	46.50%	66.50%
Female	47.82%	53.50%	33.50%

NOTE: FRA = full retirement age; SSDI = Social Security Disability Insurance; SSI = Supplemental Security Income.
SOURCES: SSA, 2013a, Tables 19 and 20, 2013b, Table 19.

such as former employers, teachers, family, or acquaintances. After reviewing all of the evidence relevant to the claim, including medical evidence, the examiner makes a determination about what the evidence shows. In some situations, the examiner is unable to make a determination because the evidence in the case record is insufficient or inconsistent. In such cases, the examiner may ask the applicant to attend a consultative examination, which SSA purchases.[1]

SSA establishes the presence of a medically determinable impairment in individuals with mental disorders other than intellectual disability through the use of standard diagnostic criteria, which include symptoms and signs. Evidence for claims based on mental impairment, as well as for many other categories of claims, such as those for certain musculoskeletal and connective tissue conditions, relies less on standard laboratory tests than for some other categories of impairment. These impairments are established largely on reports of signs and symptoms of impairment and functional limitation.

[1] SSA guidelines for consultative examination reports are available (SSA, 2015).

TABLE 1-2 SSDI and SSI Beneficiaries by Diagnostic Category, 2012

Diagnostic Category	SSDI Workers (%)	SSI Adults—Disability (%)	SSI Children (%)
Congenital anomalies	0.20	0.81	5.40
Endocrine, nutritional, and metabolic diseases	3.40	2.68	0.70
Infectious and parasitic diseases	1.40	1.35	0.10
Injuries	4.10	2.62	0.50
Intellectual disability	4.20	19.15	9.60
Other mental disorder	27.60	38.41	57.90
Neoplasms	3.10	1.33	1.20
Disease—Blood and blood forming organs	0.30	0.40	1.10
Disease—Circulatory system	8.40	4.26	0.50
Disease—Digestive system	1.70	1.04	1.20
Disease—Genitourinary system	1.70	1.02	0.30
Disease—Musculoskeletal system and connective tissue	29.80	12.78	0.80
Disease—Nervous system and sense organs	9.30	7.68	7.80
Disease—Respiratory system	2.90	2.04	2.80
Disease—Skin and subcutaneous tissue	0.20	0.17	0.20
Other	0.20	0.27	7.80
Unknown	1.40	3.99	2.10

NOTE: SSDI = Social Security Disability Insurance; SSI = Supplemental Security Income.
SOURCES: SSA, 2013a, Table 21, 2013b, Tables 20, 35, 36.

SSA establishes the severity of functional limitations through a combination of self-reports on what an applicant can and cannot do in work and work-like settings and related reports from others. The consistency of such evidence with the evidence of signs, symptoms, and laboratory findings from other sources is what SSA uses to determine disability. Mental disorders other than intellectual disabilities and certain musculoskeletal system and connective tissue disorders together account for about 57 percent of SSDI claims, 41 percent of SSI adult claims, and 59 percent of SSI child claims (see Table 1-2) (SSA, 2013a, Table 21, 2013b, Tables 20, 35, 36).

SSA maintains a list of criteria[2] for specific conditions that an applicant with one or more of those conditions must meet in order to receive disability benefits based solely on medical criteria. SSA currently requires psychological test results, specifically intelligence test results, in the listing criteria for intellectual disability in children and adults and in the criteria for cerebral palsy, convulsive epilepsy, and meningomyelocele and related

[2] *Disability evaluation under Social Security—Part III Listing of Impairments*. http://www.ssa.gov/disability/professionals/bluebook/listing-impairments.htm (accessed October 3, 2014).

disorders. SSA questions the value of purchasing psychological testing in cases involving mental disorders, other than for intellectual disability, and it does not require testing either to establish or to assess the severity of other mental disorders.

Nevertheless, disability examiners and consultative examiners may request psychological testing, within the confines of the rules of each state's Disability Determination Services (DDS), if they think the test results would inform the adjudication of an individual's disability claim. Aside from the use of intelligence tests as described in the listings for intellectual disability and certain neurological impairments, SSA does not require or specify the purchase of any type of (or individual) psychological test. SSA provides general guidance that good psychological tests are valid and reliable and have appropriate normative data. Because each DDS issues its own rules regarding the tests that may be purchased, there is variation among states about when and which tests can be purchased.

When objective medical evidence cannot substantiate the credibility of an applicant's statements about his or her symptoms (and their effects on his or her functioning), SSA rules require disability examiners to consider all of the evidence in the case record. Examiners are directed to consider:

- The applicant's medical history, diagnosis, and prescribed treatment;
- The applicant's daily activities and efforts to work;
- Any other evidence showing how the applicant's impairment(s) and any related symptoms affect his or her ability to work (or, for a child, his or her ability to function compared to that of other children the same age who do not have impairments); and
- Any observations about the applicant recorded by SSA claims representatives during interview (in person or by telephone).[3]

Disability examiners are experts at assessing the consistency of all evidence and making a determination of its validity. As described more fully later in the chapter, there are two types of validity tests that might assist in this process. Performance validity tests (PVTs) provide information about an individual's effort on cognitive and other performance-based tests. Symptom validity tests (SVTs) provide information about the consistency and accuracy of an individual's self-report of symptoms he or she is experiencing. Both types of validity testing have generated controversy with respect to SSA policy.

There are differences of opinion on the use of validity tests and their value for work disability evaluations. SSA's current position is not to

[3] See Social Security Ruling (SSR) on the Evaluation of Symptoms in Disability Claims: Assessing the Credibility of an Individual's Statements (SSA, 1996).

purchase validity tests to address issues of credibility or malingering as part of a consultative examination. Although SSA does not purchase validity tests, claimants and their representatives sometimes submit them in support of their claims. Professional organizations of neuropsychologists and psychologists, such as the American Academy of Clinical Neuropsychology (AACN), the National Academy of Neuropsychology (NAN), the American Psychological Association (APA), the Association for Scientific Advancement in Psychological Injury and Law, and the British Psychological Society, have issued position statements and guidance advocating for the use of validity tests in clinical and medicolegal contexts (APA, 2013; British Psychological Society, 2009; Bush et al., 2005, 2014; Heilbronner et al., 2009). Two of these organizations, the AACN and the NAN, along with Division 40 (Neuropsychology) of the APA and the American Board of Professional Neuropsychology have challenged SSA's institutional prohibition on ordering validity tests (IOPC, 2013). In addition, a September 2013 report from SSA's Office of the Inspector General concluded that although SSA does not allow the purchase of validity tests, "medical literature, national neuropsychological organizations, other federal agencies, and private disability insurance providers support the use of [validity tests] in determining disability claims" (Office of the Inspector General, SSA, 2013, p. ii).

It is against this background that SSA asked the Institute of Medicine (IOM) to convene a committee of relevant experts to review selected psychological tests, including validity tests, and to provide guidance on the use of such testing in the adjudication of claims submitted to the SSA Disability Programs (see Box 1-1 for the statement of task). In carrying out this task, the Committee on Psychological Testing, Including Validity Testing, for Social Security Administration Disability Determinations was asked by the sponsor to address several specific topics, including testing norms, the administration of relevant tests and the qualifications for administering them, the interpretation and reporting of test results, and economic considerations relevant to the use of such tests for the disability evaluation process.[4] The 11-member committee included experts in the areas of adult and pediatric neuropsychology, psychology, psychiatry, disability medicine, behavioral economics, and economics (see Appendix B).

[4] In the project background material, the sponsor asked the committee to consider topics such as the cost of administering these tests, whether the cost varies by location, and the cost effectiveness (including cost per claim) of requiring a single test or a combination of tests in the disability evaluation process for physical and mental impairments (Revised project background, submitted by Joanna Firmin, Social Security Administration, May 23, 2014).

BOX 1-1
Statement of Task

An ad hoc committee will:

1. Perform a critical review of selected psychological tests, including symptom validity tests (SVTs), that could contribute to Social Security Administration (SSA) disability determinations;
2. Provide guidance on the general relevance and applicability of psychological tests, including SVTs, in the context of other relevant evidence to SSA disability determinations in claims involving physical and mental disorders; and
3. Provide guidance on how to use the results of psychological tests, including SVTs, in the context of disability determinations.

To accomplish these objectives, the committee shall consider the following topics: (1) use of psychological testing, (2) testing norms, (3) qualifications for administration of tests, (4) administration of tests, (5) reporting results, and (6) use of tests for the disability evaluation process.

COMMITTEE'S APPROACH TO ITS CHARGE

Terminology and Parameters of Study

Terminology that is fundamental to the committee's report, including the concept of disability, a variety of psychological terms, and the concept of credibility, is described in the following sections. Appendix C of the report contains a glossary of definitions for a number of terms that are particularly relevant to the committee's work.

Concept of Disability

SSA defines disability in adults as

> The inability to engage in any substantial gainful activity … by reason of any medically determinable physical or mental impairment(s) which can be expected to result in death or which has lasted or can be expected to last for a continuous period of not less than 12 months. (SSA, n.d., see also 2012b)

Substantial gainful activity is work that "involves doing significant and productive physical or mental duties" and "is done (or intended) for pay

or profit" (20 CFR § 416.910). A medically determinable physical or mental impairment is defined as "an impairment that results from anatomical, physiological, or psychological abnormalities which can be shown by medically acceptable clinical and laboratory diagnostic techniques" (SSA, n.d.).

Disability in children under 18 years of age is defined as

> a medically determinable physical or mental impairment or combination of impairments that causes marked and severe functional limitations, and that can be expected to cause death or that has lasted or can be expected to last for a continuous period of not less than 12 months. (SSA, n.d., see also 2012b)

The concept of disability is complex and reflects the interplay between an individual with a mental or physical health condition and all aspects of his or her biology, behavior, and environment. The World Health Organization (WHO) developed the *International Classification of Functioning, Disability and Health* (ICF) framework (WHO, 2001) "using a global consensus-building process that involved multiple stakeholders, including people with disabilities" (IOM, 2007b, p. 37). Endorsed by the World Health Assembly in May 2001, the ICF is a part of the WHO's family of International Classifications, which includes the *International Statistical Classification of Diseases and Related Health Problems, 10th Revision* (ICD-10) (IOM, 2007b, p. 37; WHO, 1992).

Consistent with previous disability frameworks, including those from prior IOM reports (IOM, 1991, 1997, 2007a) and Nagi (1965, 1976), "the ICF attempts to provide a comprehensive view of health-related states from a biological, personal, and social perspective" (IOM, 2007b, p. 37). Human functioning and disability are portrayed "as the product of a dynamic interaction between various health conditions and environmental and personal contextual factors" (IOM, 2007b, p. 37). The ICF framework differs from previous frameworks in that its components are described using both positive and negative terms (IOM, 2007b, p. 37) (see Box 1-2). Thus, it refers to health and functioning as well as disability.

As in the 1991 and 1997 IOM frameworks,

> the ICF identifies multiple levels of human functioning and disability: at the level of body or body parts, at the level of the whole person, and at the level of the whole person who is functioning in his or her environment. These levels, in turn, involve three aspects of human functioning that the ICF terms body functions and structures, activities, and participation. (IOM, 2007b, pp. 37–38)

Within the ICF, the term disability is used to denote decrements in all three aspects of human functioning, which are labeled impairments, activity limitations, and participation restrictions (IOM, 2007b, p. 38). For the purposes of SSA, disability in adults refers to the inability to work at any

BOX 1-2
Major Concepts in the *International Classification of Functioning, Disability and Health*

Health condition: Umbrella term for disease, disorder, injury, or trauma

Functioning: Umbrella term for body functions and structures, activities, and participation

Disability: Umbrella term for impairments, activity limitations, and participation restrictions

Body function: Physiological functions of body systems (including psychological functions)

Body structure: Anatomical parts of the body such as organs, limbs, and their components

Impairment: Problems in body function or structure such as a significant deviation or loss

Activity: Execution of a task or action by an individual

Activity limitations: Difficulties an individual may have in executing activities

Participation: Involvement in a life situation

Participation restriction: Problems an individual may experience in involvement in life situations

Environment: The physical, social, and attitudinal environment in which people live and conduct their lives

Personal factors: Contextual factors that relate to the individual such as age, gender, social status, and life experiences

SOURCE: WHO, 2001, pp. 10, 211–214. Reprinted from IOM, 2007b, p. 38.

job for a continuous period of 12 or more months. On this definition, disability refers to a participation restriction, namely, an inability to participate in work-related activity. Disability in children refers to "marked and severe functional limitations" relative to typically functioning peers of the same age.

Noteworthy is the dynamic interaction between the different components of the ICF model and various environmental (social and physical) and personal contextual (biological and behavioral) factors (see Figure 1-1) (IOM, 1991; WHO, 2001, p. 19). Movement among the components is mediated by these factors and may occur in either direction—disabling or enabling (IOM, 1991, 1997; WHO, 2001). Someone who lost a leg to disease or injury, for example, would then have a limitation with respect to walking, but that limitation might be reversed by the provision of a

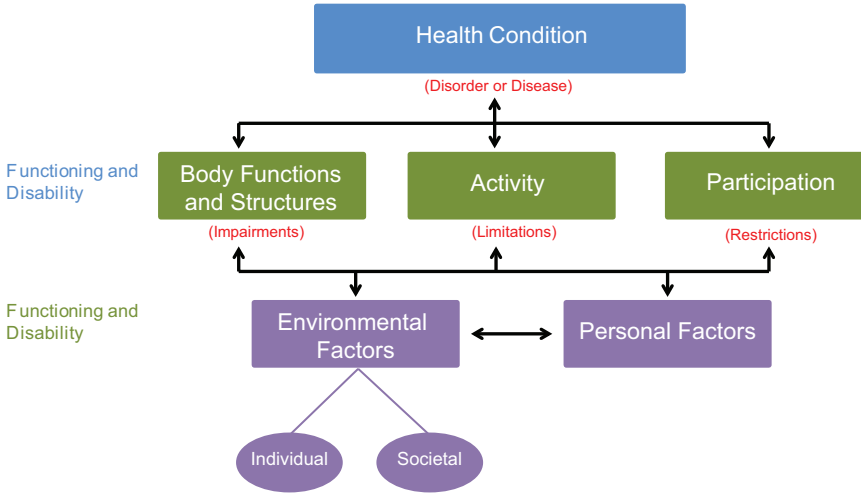


FIGURE 1-1 ICF model of disability and functioning.
SOURCE: Adapted from WHO, 2001, p. 18.

prosthetic leg. Similarly whether an individual is disabled as a result of his or her functional or activity limitations depends on the accommodations available to the individual that permit the person to engage in activities he or she otherwise would be unable to perform (IOM, 1997).

For this reason, disability is not tightly correlated with the presence of impairment. Both need to be evaluated, but the measures are fundamentally different, including objective measures (performance and anatomical) and self-report measures that help determine how usual roles are disrupted. The linkages among an individual's anatomy, diagnosis, and impairment are not sufficient to determine the presence of work disability. As the 2007 IOM report *Improving the Social Security Disability Decision Process* states with respect to work disability:

> Work disability … results from the interaction of individuals' impairments, functional limitations resulting from the impairments, assistive technologies to which they may have access, and attitudinal and other personal characteristics (such as age, education, skills, and work history) with the physical and mental requirements of potential jobs, accessibility of transportation, attitudes of family members and coworkers, and willingness of an employer to make accommodations. (IOM, 2007c, p. 26)

Given the complex interaction among the variety of factors that underlie a disability, it is clear that disability determinations are multidimensional and always involve some element of judgment (IOM, 1987). Although objective medical evidence can indicate the presence of physical or mental

impairments, the decision about whether those impairments result in a disability is an administrative or legal one (IOM, 1987; IOM and NRC, 2007).

Psychological Terms

Psychological assessment refers to

> the comprehensive integration of information from a variety of sources—including formal psychological tests, informal tests and surveys, structured clinical interviews, interviews with others, school and/or medical records, and observational data—to make inferences regarding the mental or behavioral characteristics of an individual or to predict behavior. (Furr and Bacharach, 2013; Hubley and Zumbo, 2013)

Psychological testing refers to "the use of formal, standardized procedures for sampling behavior that ensure objective evaluation of the test-taker regardless of who administers the test" (Furr and Bacharach, 2013; Hubley and Zumbo, 2013).

Major categories of psychological tests include (1) intelligence tests, (2) neuropsychological tests, (3) personality tests, (4) disorder-specific tests (e.g., depression, anxiety), (5) achievement tests, (6) aptitude tests, and (7) occupational or interests tests. The first four categories capture the tests that are most relevant to disability determinations. Standardized psychological tests can be divided into *measures of typical behavior* and *tests of maximal performance*. Measures of typical behavior, such as personality, interests, values, and attitudes, may be referred to as *non-cognitive measures*. Tests of maximal performance ask people to answer questions and solve problems as well as they possibly can. Because tests of maximal performance typically involve cognitive performance, they are often referred to as *cognitive tests*. It is through these two lenses—non-cognitive measures and cognitive tests—that the committee examined psychological testing for the purpose of disability evaluation in this report. Intelligence tests and neuropsychological tests are examples of cognitive-based measures, while depression, anxiety, or personality inventories are examples of non-cognitive measures. Psychological tests may also be categorized as performance based and self-report. Cognitive tests tend to be performance based, and non-cognitive measures tend to be based on self-report.

A variety of validity tests have been developed to assist examiners in interpreting the results of different psychological tests. The committee distinguishes in this report between *performance validity tests* (PVTs), which provide information about an individual's effort on tests of maximal performance, such as cognitive tests, and *symptom validity tests* (SVTs), which provide information about the consistency and accuracy of an individual's self-report of symptoms he or she is experiencing. PVTs are stand-alone or

embedded or derived measures that are used to assess whether an examinee is performing at a level consistent with his or her actual abilities (Larrabee, 2014). Measures of performance validity, often referred to as "effort" in the literature, generally are associated with neuropsychological or cognitive testing. As discussed in Chapter 5, PVTs help the examiner to interpret the validity of an individual's neuropsychological or cognitive test results. If an individual has not given his or her best effort in taking the test, the results may not provide an accurate picture of the person's neuropsychological or cognitive functioning. SVTs are measures embedded in non-cognitive psychological measures (e.g., personality, mood scales) that are used to assess whether an examinee is providing an accurate report of his or her actual symptom experience (Larrabee, 2014).

The distinction between performance validity and symptom validity was first introduced in the literature in 2012 (Larrabee, 2012). Prior to that time, the term *symptom validity* often encompassed the concept of performance validity as well as the consistency and accuracy of symptom self-report. The committee has made every effort to maintain the distinction between performance validity and symptom validity and to use the terms consistently throughout the report. In some cases, doing so required interpreting published literature, particularly older literature, in light of the revised terminology. For this reason, the report, when appropriate, may refer to performance validity when discussing a particular publication, despite the original source using the term *symptom validity*.

Table 1-3 provides a summary of the psychological terms discussed in this section, and Figure 1-2 shows the relationships among the different terms.

Credibility

In situations involving the potential for secondary gain—such as monetary gain from a SSA disability payment—there may be motivation for individuals intentionally to feign or exaggerate symptoms or to exert suboptimal effort on performance measures in order to present a stronger need for support or disability benefits. *Malingering* is the intentional presentation of false or exaggerated symptoms, intentionally poor performance, or a combination of the two, motivated by external incentives (American Psychiatric Association, 2013; Bush et al., 2005; Heilbronner et al., 2009). Two key elements of malingering are intention to deceive or mislead and motivation to do so for the purpose of achieving some type of secondary gain.

It is important to distinguish between malingering and the credibility or noncredibility of an individual's performance or symptom report, even in situations of potential secondary gain. Individuals might over- or underreport symptoms or not give their best effort on cognitive-based measures

TABLE 1-3 Definitions of Psychological Terms

Term	Definition	Description
Performance validity tests (PVTs)	Stand-alone or embedded/derived tests used to assess whether a test-taker is performing at a level consistent with his or her actual abilities	Assesses validity in tests of maximal performance, e.g., *cognitive tests*: • Intelligence tests[a] • Neuropsychological tests[a]
Psychological assessment	"The comprehensive integration of information from a variety of sources—including formal psychological tests, informal tests and surveys, structured clinical interviews, interviews with others, school and/or medical records, and observational data—to make inferences regarding the mental or behavioral characteristics of an individual or to predict behavior" (Furr and Bacharach, 2013; Hubley and Zumbo, 2013).	
Psychological tests	Formal, standardized procedures for sampling behavior that ensure objective evaluation of the test-taker regardless of who administers the test Can be divided into *cognitive tests* and *non-cognitive measures*	Major categories: • *Non-cognitive* o Personality tests[a] o Clinical/Diagnostic tests (e.g. depression, anxiety)[a] o Occupational or interest tests • *Cognitive* o Intelligence tests[a] o Neuropsychological tests[a] o Achievement tests o Aptitude tests
Symptom validity tests (SVTs)	Embedded in self-report psychological tests (e.g., personality, mood scales) and used to assess whether an examinee is providing an accurate report of actual symptom experience	Assesses validity in self-report measures, e.g., *non-cognitive measures*: • Personality tests[a] • Clinical/Diagnostic tests[a]

[a] Most relevant to disability determinations.

SOURCES: Bush et al., 2005; Furr and Bacharach, 2013; Hubley and Zumbo, 2013; Larrabee, 2014.

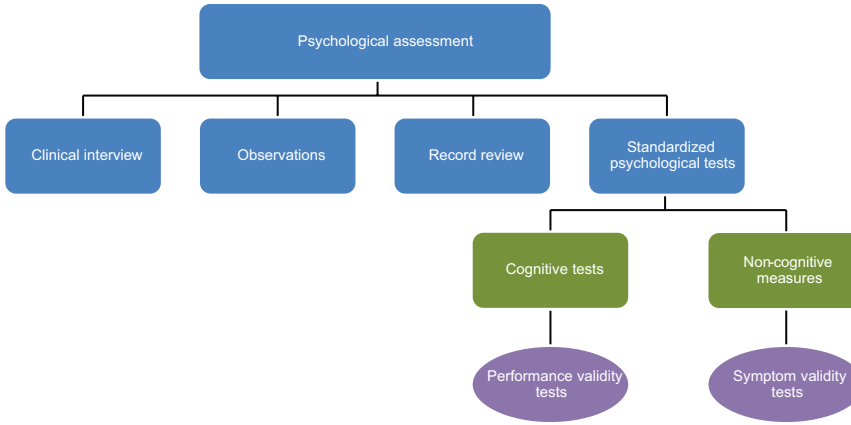


FIGURE 1-2 Components of psychological assessment.
NOTE: Performance validity tests do not measure cognition but are used in conjunction with performance-based cognitive tests to examine whether the examinee is exerting sufficient effort to perform well and responding to the best of his or her capability. Similarly, symptom validity tests do not measure non-cognitive status but are used to examine whether a person is providing an accurate report of his or her actual symptom experience. Because cognitive tests frequently are performance based and non-cognitive measures generally involve self-report, performance validity tests and symptom validity tests are shown as being associated with these types of tests.

for any number of reasons. SVTs and PVTs do not in themselves provide information about the motivations of an examinee[5] or the reasons why his or her performance or symptom report may appear to be noncredible. Throughout the report, the committee has avoided use of the term *malingering* when discussing the results of PVTs and SVTs, opting instead to refer to the credibility or accuracy of an individual's performance or symptom report. The committee intends such terms to be value-neutral with respect to the examinee, referring only to whether the examinee exerted sufficient effort for the test results to be considered valid and to the consistency and accuracy of the individual's statements about the experience of symptoms.

[5] Although below chance scores on a PVT can speak to an examinee's intention—the individual knew the answer and deliberately chose the wrong one—they cannot speak directly to the individual's motivation (reason) for intentionally choosing the wrong answer.

Study Focus

Although the report focuses primarily on the use of psychological tests in disability determinations in adults, the use of such tests in children is also addressed. There are three areas in SSA's disability determination process where psychological testing could be of value: (1) identification of a "medically determinable impairment," (2) evaluation of functional capacity for work, and (3) assessment of the validity of claimants' psychological test results or the accuracy of statements about self-reported symptoms. Although the report addresses all three areas, the committee focuses on the second and the third, where questions about the use of psychological tests are more complex.

In considering its task, the committee observed that the vast number (in the hundreds) of cognitive and non-cognitive psychological tests available for use precludes a detailed analysis of each specific test and recommendations about the use of specific tests. In addition, decisions about which specific tests are most appropriate for particular individuals in a particular set of circumstances properly fall in the realm of clinical decision making. Instead, the committee reviewed categories of psychological tests, including validity tests, and this report provides general guidance on the use of such tests in SSA disability determinations for claims involving physical and mental disorders.

It is important to note that SSA specifically requested that the committee not address the use of intelligence tests in making determinations about intellectual disability since that topic was previously examined in a 2002 National Research Council (NRC) report titled *Mental Retardation: Determining Eligibility for Social Security Benefits* (NRC, 2002). Consideration of intelligence tests with respect to embedded validity measures, however, was deemed to be within the committee's purview.

Information-Gathering Process

The committee conducted an extensive review of the literature pertaining to the use of psychological tests, including PVTs and SVTs, in disability determinations. The committee began with an English-language literature search of online databases, including PubMed, Embase, Medline, Web of Science, Scopus, PsychINFO, Government Accountability Office (GAO), Congressional Research Service, Google, Google Scholar, and Legistorm (GAO reports, congressional memorandums). Additional literature and other resources were identified by committee members and project staff using traditional academic research methods and online searches. Attention was given to consensus and position statements issued by relevant experts and professional organizations.

The committee used a variety of sources to supplement its review of the literature. It met in person five times and held two public workshops to hear

from invited experts in areas pertinent to the topic (see Appendix A for the open session agendas and speaker lists). Speakers included neuropsychologists with expertise in performance and symptom validity testing in adults and children, the use of psychological and validity tests in culturally diverse populations, and the use of such tests in non-SSA disability determination contexts (e.g., private disability insurance programs, Canadian auto insurance, U.S. military disability or return-to-duty decisions, veterans' disability compensation). The committee also heard from SSA and DDS representatives about the SSA disability determination process and its current policies surrounding the use of psychological and validity testing.

In addition, the committee commissioned two papers to provide additional critical analysis in areas relevant to the committee's work. One paper addresses issues of diversity (e.g., in terms of culture, language, gender and gender identity, educational or socioeconomic status) and multiculturalism in the use of psychological tests (self-report measures and performance-based cognitive tests as well as corresponding validity tests) in making disability determinations. The authors were asked to discuss the use of psychological tests in diverse populations in terms of their validity, fairness, and other characteristics. They also were asked to address whether, when, and/or how to use such measures, despite any limitations, in disability determinations for diverse populations in the United States.

Based on its review of the literature, the presentations from invited experts on PVT and SVT research at its open sessions, and the expertise of several of its members, the committee understood the arguments and evidence supporting the inclusion of validity tests in psychological and neuropsychological tests and test batteries. Because the committee found very little published literature critiquing the use of SVTs and PVTs, they felt it was important to seek more information about potential concerns or questions pertaining to their use. To this end, they commissioned a second paper and asked the author to address a number of questions designed to probe any challenges or cautions about the use of validity tests for disability determinations in different populations. The questions posed by the committee included the following:

- For whom are PVTs and SVTs useful for informing disability determinations? In what way?
- How or in what way do the results of PVTs or SVTs correlate with assessing functional limitations (such as limitations in a person's ability to do basic work activities, activities of daily living, social functioning, and concentration, persistence, or pace) due to an impairment?
- Given the historical context in which PVTs and SVTs were developed for forensic use in litigation settings, can they be adapted for

use in disability determinations? Discuss the transferability of PVTs and SVTs given the differences in evidence use and decision making among fields (legal versus mediated or negotiated).

- How should one interpret validity test scores or results in the "grey area" between clear failures (e.g., below chance scores) and clear passes on SVTs or PVTs? How many people fail completely versus at the margins?
- When interpreting PVT or SVT failures, particularly in the "grey zone," are there factors aside from malingering or intentionally poor performance that may explain the results (e.g., stems from symptoms, fatigue, apathy)?
- How does the current norming of SVTs and PVTs affect their usefulness in a variety of different populations (e.g., a diversity of race, ethnicity, culture, and educational or socioeconomic status)? Are there ways to resolve or mitigate the challenges posed by lack of norming for particular populations?

The committee's work was further informed by previous IOM and NRC reports, including *Pain and Disability: Clinical, Behavioral, and Public Policy Perspectives* (IOM, 1987); *Disability in America: Toward a National Agenda for Prevention* (IOM, 1991); *Enabling America: Assessing the Role of Rehabilitation Science and Engineering* (IOM, 1997); *PTSD Compensation and Military Service* (IOM and NRC, 2007); *The Future of Disability in America* (IOM, 2007b); *Improving the Social Security Disability Decision Process* (IOM, 2007c); *A 21st Century System for Evaluating Veterans for Disability Benefits* (IOM, 2007a); *Mental Retardation: Determining Eligibility for Social Security Benefits* (NRC, 2002); and *Survey Measurement of Work Disability: Summary of a Workshop* (NRC, 2000).

REPORT ORGANIZATION

Chapter 2 describes the current SSA disability determination process, focusing on areas relevant to the use of psychological tests. It also discusses the use of psychological tests in disability evaluations in non-SSA contexts. Chapter 3 provides an overview of psychological tests, including the different types of tests and their use, psychometrics and norms, and the administration of tests. Chapter 4 reviews the use of standardized psychological self-report measures and SVTs in the context of SSA disability determinations. Chapter 5 addresses standardized cognitive tests and the use of PVTs. Chapter 6 explores economic considerations related to the use of psychological testing in SSA disability determinations. Chapter 7 contains the committee's conclusions and recommendations.

REFERENCES

American Psychiatric Association. 2013. *American Psychiatric Association: Diagnostic and statistical manual of mental disorders, fifth edition (DSM-5)*. Arlington, VA: American Psychiatric Association.

APA (American Psychological Association). 2013. Specialty guidelines for forensic psychology. *American Psychologist* 68(1):7-19.

British Psychological Society. 2009. *Assessment of effort in clinical testing of cognitive functioning for adults*. Leicester, UK: British Psychological Society.

Bush, S. S., R. M. Ruff, A. I. Tröster, J. T. Barth, S. P. Koffler, N. H. Pliskin, C. R. Reynolds, and C. H. Silver. 2005. Symptom validity assessment: Practice issues and medical necessity. NAN Policy & Planning Committee. *Archives of Clinical Neuropsychology* 20(4):419-426.

Bush, S. S., R. L. Heilbronner, and R. M. Ruff. 2014. Psychological assessment of symptom and performance validity, response bias, and malingering: Official position of the Association for Scientific Advancement in Psychological Injury and Law. *Psychological Injury and Law* 7(3):197-205.

Furr, R. M., and V. R. Bacharach. 2013. *Psychometrics: An introduction*. Thousand Oaks, CA: Sage Publications, Inc.

Heilbronner, R. L., J. J. Sweet, J. E. Morgan, G. J. Larrabee, S. R. Millis, and 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.

Hubley, A. M., and B. D. Zumbo. 2013. Psychometric characteristics of assessment procedures: An overview. In *APA handbook of testing and assessment in psychology*, *Volume 1—Test theory and testing and assessment in industrial and organizational psychology*, edited by K. F. Geisinger, N. R. Kuncel, S. P. Reise, M. C. Rodriguez. Washington, DC: American Psychological Association.

IOM (Institute of Medicine). 1987. *Pain and disability: Clinical, behavioral, and public policy perspectives*. Washington, DC: National Academy Press.

IOM. 1991. *Disability in America: Toward a national agenda for prevention*. Washington, DC: National Academy Press.

IOM. 1997. *Enabling America: Assessing the role of rehabilitation science and engineering*. Washington, DC: National Academy Press.

IOM. 2007a. *A 21st century system for evaluating veterans for disability benefits*. Washington, DC: The National Academies Press.

IOM. 2007b. *The future of disability in America*. Washington, DC: The National Academies Press.

IOM. 2007c. *Improving the social security disability decision process*. Washington, DC: The National Academies Press.

IOM and NRC (National Research Council). 2007. *PTSD compensation and military service*. Washington, DC: The National Academies Press.

IOPC (Inter Organizational Practice Committee). 2013. Use of symptom validity indicators in SSA psychological and neuropsychological evaluations. Letter to Senator Tom Coburn. https://www.nanonline.org/docs/PAIC/PDFs/SSA%20and%20Symptom%20Validity%20Tests%20-%20IOPC%20letter%20to%20Sen%20Coburn%20-%202-11-13.pdf (accessed February 8, 2015).

Larrabee, G. J. 2012. Performance validity and symptom validity in neuropsychological assessment. *Journal of the International Neuropsychological Society* 18(4):625-630.

Larrabee, G. J. 2014. *Performance and symptom validity*. Presentation to the IOM Committee on Psychological Testing, Including Validity Testing, for Social Security Administration Disability Determinations, June 25, 2014, Washington, DC.

Nagi, S. Z. 1965. Some conceptual issues in disability and rehabilitation. In *Sociology and rehabilitation*, edited by M. B. Sussman. Washington, DC: American Sociological Association. Pp. 100-113.

Nagi, S. Z. 1976. An epidemiology of disability among adults in the United States. *Milbank Memorial Fund Quarterly Health and Society* 54(4):439-467.

NRC (National Research Council). 2000. *Survey measurement of work disability: Summary of a workshop*. Washington, DC: National Academy Press

NRC. 2002. *Mental retardation: Determining eligibility for social security benefits*. Washington, DC: The National Academies Press.

Office of the Inspector General, SSA (Social Security Administration). 2013. *The Social Security Administration's policy on symptom validity tests in determining disability claims*. Washington, DC: SSA. http://oig.ssa.gov/sites/default/files/audit/full/pdf/A-08-13-23094.pdf (accessed March 27, 2015).

SSA (Social Security Administration). 1996. *SSR 96-7p: Policy interpretation ruling Titles II and XVI: Evaluation of symptoms in disability claims: Assessing the credibility of an individual's statements*. http://www.socialsecurity.gov/OP_Home/rulings/di/01/SSR96-07-di-01.html (accessed October 3, 2014).

SSA. 2012a. DI 00115.001 Social Security Administration's (SSA) disability programs. *Program Operations Manual System (POMS)*. https://secure.ssa.gov/poms.nsf/lnx/0400115001 (accessed October 2, 2014).

SSA. 2012b. DI 00115.015 Definitions of disability. *Program Operations Manual System (POMS)*. https://secure.ssa.gov/poms.nsf/lnx/0400115015 (accessed October 3, 2014).

SSA. 2013a. *Annual statistical report on the Social Security Disability Insurance program, 2012*. http://www.socialsecurity.gov/policy/docs/statcomps/di_asr/2012/index.html (accessed September 26, 2014).

SSA. 2013b. *SSI annual statistical report, 2012*. http://www.socialsecurity.gov/policy/docs/statcomps/ssi_asr/2012/index.html (accessed September 26, 2014).

SSA. 2015. DI 22510.000 Development of consultative examinations (CE). *Program Operations Manual System (POMS)*. https://secure.ssa.gov/apps10/poms.nsf/lnx/0422510000 (accessed January 27, 2015).

SSA. n.d. *Disability evaluation under social security; Part I—General information*. http://www.ssa.gov/disability/professionals/bluebook/general-info.htm (accessed November 14, 2014).

WHO (World Health Organization). 1992. *International statistical classification of diseases and related health problems, 10th revision (ICD-10)*. Geneva: WHO.

WHO. 2001. *International classification of functioning, disability and health (ICF)*. Geneva: WHO.

2

Disability Evaluation and the Use of Psychological Tests

In 2013, the U.S. Social Security Administration (SSA) received approximately 2.6 million applications for Social Security Disability Insurance (SSDI) disabled worker benefits (SSA, n.d.-m), 1.6 million applications for the Supplemental Security Income (SSI) adult program (SSA, 2014a, p. 92, Table V.C.1), and 442,000 applications for the SSI child program (SSA, 2014a, p. 24, Table V.C.2). This chapter describes SSA's process for evaluating applications and determining the disability status of the applicants, including the use of psychological testing in SSA disability evaluations. It also provides an overview of base rates of "malingering" and a discussion of the benefits of formal, standardized data collection and actuarial data interpretation. The chapter concludes with an overview of the use of psychological tests in disability evaluations in non-SSA systems, including the U.S. military and the U.S. Department of Veterans Affairs (VA), private disability insurance, forensic assessments, and some international programs.

SOCIAL SECURITY ADMINISTRATION DISABILITY DETERMINATION PROCESS

The overall disability determination process (see Figure 2-1) is the same for both SSDI and SSI, although the specific steps of the process vary for adults (20 CFR § 416.920; see Figure 2-2) and children (20 CFR § 416.924; see Figure 2-3). For the average applicant, the initial determination process takes between 90 and 120 days from the date of filing. Decisions for applicants with certain medical conditions, incomplete medical records, or

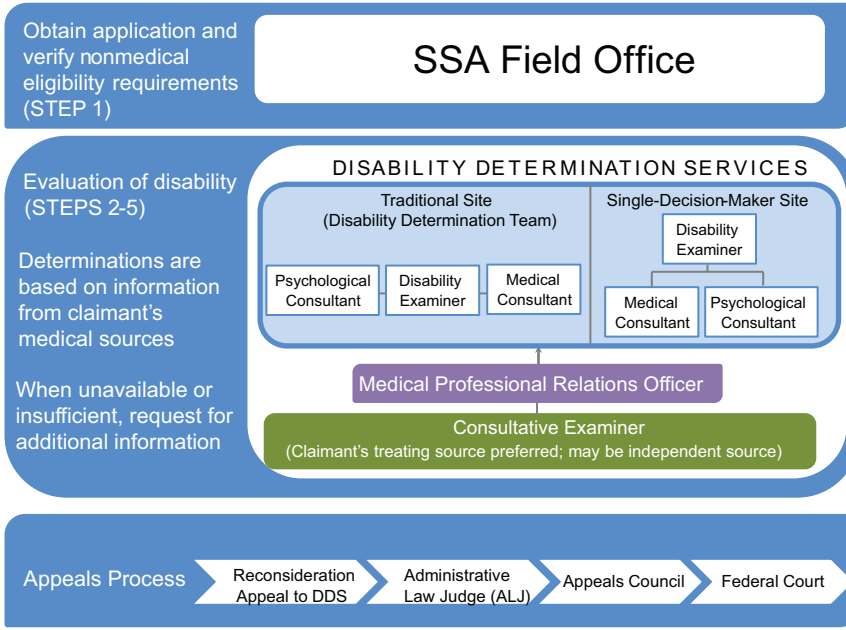
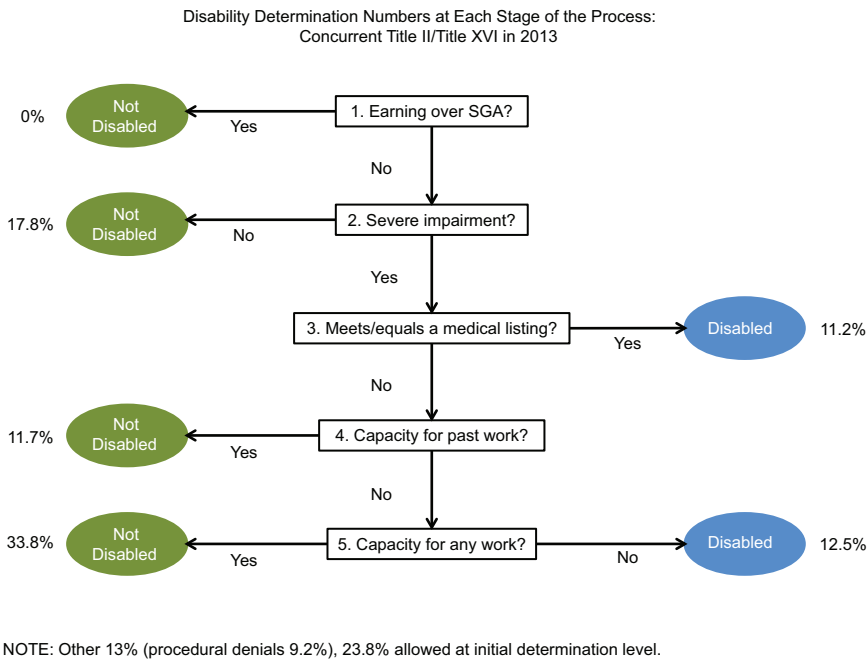
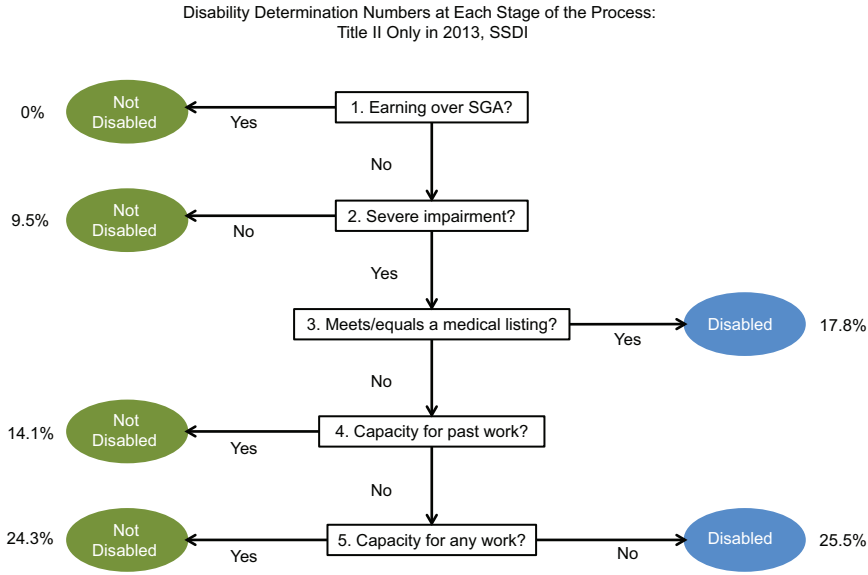


FIGURE 2-1 Overview of the SSA disability process.

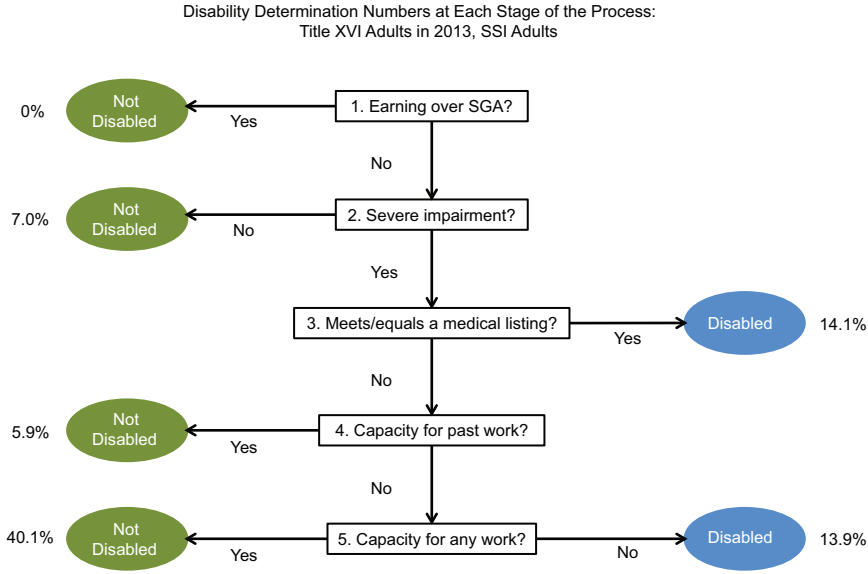


NOTE: Other 13% (procedural denials 9.2%), 23.8% allowed at initial determination level.

FIGURE 2-2 Disability determination process for adults by the numbers.
SOURCES: SSA, 2014d,h.

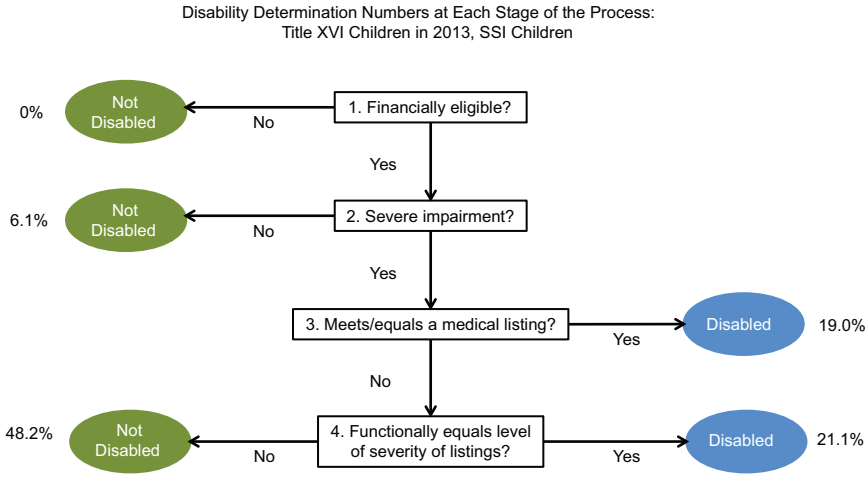


NOTE: Other 8.8%, 43.7% allowed at initial determination level.



NOTE: Other 19.0% (procedural denials 6.2%), 28.1% allowed at initial determination level.

FIGURE 2-2 Continued.
SOURCES: SSA, 2014d,h.



NOTE: Other 5.6%, 40.1% allowed at initial determination level.

FIGURE 2-3 Disability determination process for children by the numbers.
SOURCE: SSA, 2014h.

who appeal the initial decision can take far longer, in some cases stretching across several years (SSA, 2014i; SSDRC, n.d.).

Step 1: Nonmedical Eligibility?

Applications for disability benefits are made at a local SSA field office. During the first step of the disability determination process, officials in the SSA field offices verify applicants' financial and other nonmedical (e.g., age, work credits) eligibility requirements (SSA, 2012a). For SSDI and SSI applicants, the examiners first check to see if applicants are currently working and earning more than the substantial gainful activity (SGA) amount—$1,040 per month in 2013 for non-blind applicants (SSA, 2014m). For SSI applicants, examiners also verify that applicants meet the income and resource limits necessary to qualify for these means-tested benefits.[1] For concurrent SSDI/SSI adult applicants, financially eligibility is checked for both programs. If applicants fail on any of these financial criteria, the application is denied.

If an applicant meets the nonmedical eligibility requirements, the application is forwarded to the state Disability Determination Services (DDS)

[1] For SSI child applicants, the income test relates to the resources of the household.

agency, where a disability examiner develops and reviews the medical and other evidence[2] for the claim and makes an initial determination about disability. In 2013, state DDS offices evaluated approximately 2.8 million applications for disability benefits distributed as follows: 915,679 SSDI; 887,506 concurrent SSDI/SSI adult; 653,699 SSI adult; and 428,208 SSI child (SSA, 2014h). Before beginning the disability evaluation, DDS examiners recheck that applicants meet the financial and other nonmedical criteria for the disability programs. As shown in Figure 2-2, almost no cases that reach the DDSs are rejected at this step, because the SSA field offices have already screened the applicants on these criteria. If the financial criteria are met, the DDS agencies begin to develop the case.

DDS agencies follow either a traditional or a single-decision-maker (SDM) model (see Figure 2-1), depending on the state. In the traditional model, the disability examiner makes the determination in conjunction with a DDS psychological consultant or a medical consultant (20 CFR § 404.1615). In the SDM model (20 CFR § 404.906), disability examiners have the authority to make the initial disability determination. In most cases, the disability examiners prepare the assessments and have the authority to approve or deny claims without obtaining the signature of a medical or psychological consultant. The exception is denials for mental impairments, which must be reviewed by a psychological consultant. Medical and psychological consultants are always available to assist disability examiners in their review of claims.

Step 2: Severe Impairment?

The second step of the process is designed to screen out applicants whose medically determinable impairments are not considered to be "severe"—i.e., those who are clearly able to work at some sort of substantial gainful activity or whose impairment is expected to resolve within 12 months. A medically determinable physical or mental impairment or combination of impairments is considered severe "if it significantly limits an individual's physical or mental abilities to do basic work activities" (SSA, 1996a). The impairment also must either be expected to result in death or have lasted (or be expected to last) for 12 continuous months. An applicant is denied at this step if the medically determinable impairment or combination of impairments "has no more than a minimal effect on the ability to do basic work activities" (SSA, 1996a) or does not meet the duration criterion. In 2013, 9.5 percent of SSDI applicants, 17.8 percent of SSDI/SSI concurrent

[2] Types of evidence may include (1) objective medical evidence—i.e., medical signs and laboratory findings, (2) medical history and treatment records, (3) medical source opinions and statements, (4) statements from claimant or others, and (5) information from other sources—e.g., educational personnel, social welfare agency personnel (SSA, 2012b).

applicants, and 7.0 percent of SSI adult applicants were denied at this step (see Figure 2-2) (SSA, 2014h). If the applicant is found to have a severe impairment, the disability evaluation moves to the next step.

Step 3: Meets or Equals Medical Listings?

At Step 3, applicants' impairments are evaluated to determine whether they meet or equal the medical criteria codified in SSA's *Listing of Impairments* for adults (SSA, n.d.-c). The *Listing of Impairments* is organized by major body system and contains criteria to evaluate the severity of a listed impairment. These criteria may include assessments of work-related functioning[3] and are designed to identify individuals with impairments that are sufficiently severe to prohibit them from engaging in any kind of "gainful activity" (SSA, n.d.-b). In some cases, an individual has multiple impairments, none of which is, by itself, sufficiently severe to meet the listing criteria, or an impairment that is not included in the *Listing*. In such cases, the examiner considers whether the impairment or combination of impairments is medically equal to a listed impairment. If an applicant's impairment(s) meets or equals the listing criteria, the claim is allowed. In 2013, 17.8 percent of SSDI applicants, 11.2 percent of SSDI/SSI concurrent applications, and 14.1 percent of SSI adult applicants were allowed at this step of the disability screening process (see Figure 2-2) (SSA, 2014h). All remaining claims move to the fourth step in the evaluation process.

Step 4: Capacity for Past Work?

At this step, applicants are assessed with respect to their mental or physical "residual functional capacity" and the extent to which they can still perform activities related to jobs they have held in the past 15 years. Applicants who are found to meet the demands of "past relevant work" are denied. In 2013, 14.1 percent of SSDI applicants, 11.7 percent of SSDI/SSI concurrent applicants, and 5.9 percent of SSI adult applicants were denied at this step of the process (see Figure 2-2) (SSA, 2014h). Applicants who no longer are able to perform work they have done in the past are then assessed for their ability to perform any work in the national economy (Step 5).

[3] For mental disorders, functional limitations are used to assess the severity of the impairment. Paragraph B and C criteria in the *Listing of Impairments* for mental disorders describe the areas of function that are considered necessary for work (SSA, 2009).

Step 5: Capacity for Any Work?

At this step, applicants' residual functional capacity is evaluated along with the vocational factors of age, education, and previous work experience to determine whether they would be able to adjust to other work that exists in the national economy. Disability examiners consider increasing age, generally beginning at age 50; years of education or specialized job or vocational training; and transferability of skills from previous employment, along with an individual's residual physical and mental abilities, when determining whether the applicant could adjust to doing some sort of work (SSA, n.d.-j). For example, a 50-year-old applicant with less than a high school education, no skilled work experience, and a maximum sustained work capacity limited to sedentary work could be considered disabled, while the same 50-year-old applicant who has experience as a skilled worker could be denied. If an applicant is found unable to perform any work in the national economy, the claim is allowed; otherwise, the claim is denied. In 2013, 24.3 percent of SSDI applicants were denied benefits at this stage, and 25.5 percent were determined to be eligible for benefits (see Figure 2-2) (SSA, 2014h). Among SSDI/SSI concurrent applicants, 33.8 percent were denied at Step 5, and 12.5 percent were allowed (see Figure 2-2) (SSA, 2014h). Among SSI adult applicants, 40.1 percent were denied at Step 5, and 13.9 percent were allowed (see Figure 2-2) (SSA, 2014h). Notably, more than 50 percent of the initial determinations made at the DDS level in 2013 were made in this final step of the disability determination process, when medical-vocational factors are a primary component of the determination decision.[4]

SSA is in the process of updating its system for making medical-vocational decisions (SSA, n.d.-l). The medical-vocational decisions require up-to-date information about the occupations that exist in the national economy. Through an interagency agreement with the U.S. Bureau of Labor Statistics (BLS), SSA is working to develop an Occupational Information System (OIS). The OIS would include data elements of interest to SSA, including data elements that describe the mental and cognitive demands of work, on the full range of occupations available in the national economy.

At the end of the five-step determination process, 43.7 percent of SSDI applicants, 23.8 percent of SSDI/SSI adult concurrent applicants, and 28.1 percent of SSI adult applicants in 2013 were awarded benefits during the initial determination process (SSA, 2014h).[5] As described below, applicants

[4] The large number of cases determined on medical-vocational criteria is not unusual or unique to 2013.

[5] These figures are obtained by summing the percentages shown in Figure 2-2 for denied and allowed applicants across all stages. Applications for SSDI and SSI adult benefits may be initially denied at any point along the five-step determination process. Applications may be allowed only at Steps 3 and 5.

denied benefits during this initial evaluation process may be eligible for appeal. As such, the allowance rates from this initial evaluation stage are lower than the final allowance rates for all applicants.

Sequential Disability Determination Process for Children

The first two steps of the disability determination process are similar for children younger than 18 years of age and adults. As with SSDI and SSI adult applications, almost no applications are rejected at Step 1 due to prescreening of the nonmedical eligibility requirements by the SSA field offices. Step 2 for children involves a determination of whether the child has a medically determinable impairment or combination of impairments that causes more than "minimal functional limitations" rather than whether it precludes substantial gainful activity as in the adult cases (20 CFR § 416.924). In 2013, 6.1 percent of SSI child applications were denied at Step 2 (see Figure 2-3) (SSA, 2014h). As with adults, Step 3 involves a determination of whether a child's medically determinable physical or mental impairment(s) meets or medically equals the clinical criteria in SSA's *Listing of Impairments* for children (SSA, n.d.-d). If so, the claim is allowed. In 2013, 19 percent of SSI child applications were allowed at this stage (see Figure 2-3) (SSA, 2014h).

The primary difference between disability evaluations for children and adults is in an additional component of the evaluation at Step 3 for children whose impairments do not meet or medically equal the listings. In these cases, the examiner considers whether the impairment results in limitations that functionally equal the medical listings (20 CFR § 416.926a). To be functionally equal to the listings, the impairment must result in "marked" limitations in two of six domains of functioning or an "extreme" limitation in one of the domains.[6] The six domains considered are "(1) acquiring and using information, (2) attending and completing tasks, (3) interacting and relating with others, (4) moving about and manipulating objects, (5) caring for oneself, and (6) health and physical well-being" (20 CFR § 416.926a). In making the assessment, the examiner considers all of the information in the record about the interactive and cumulative effects of the impairments, including any that are not "severe," on the child's functioning during all activities at home, at school, and in the community. The assessment is based on how "appropriately, effectively, and independently" the child performs these activities compared to children of the same age who do not have

[6] A limitation is "marked" if it seriously interferes with the child's ability to independently initiate, sustain, or complete activities and is "extreme" if it very seriously interferes with the child's ability to independently initiate, sustain, or complete age-appropriate activities (20 CFR § 416.926a).

impairments (20 CFR § 416.926a). If the child's impairment functionally equals the severity of the medical listings, the application is approved. In 2013, 21.1 percent of applications were allowed and 48.6 percent were denied at this final step (see Figure 2-3) (SSA, 2014h).

The remaining steps of the disability determination process for adults, Steps 4 and 5, do not pertain to children. Summing the allowances in at Steps 2 and 3 (see Figure 2-3) brings total allowances in the initial determination stage to 40.1 percent (SSA, 2014h).The remaining cases were denied during the initial determination process. As with adults, denied applicants are allowed to appeal their decision, potentially increasing the final allowance rate for the program.

Medical and Other Evidence and Consultative Exams

The DDS uses the medical and other evidence in the applicants' files in making disability determinations. SSA recognizes different categories of evidence, including (1) objective medical evidence; (2) narrative medical records, opinions, and statements from treating and nontreating medical sources; (3) statements by the applicant for the file or made to medical sources or SSA field office or DDS representatives; and (4) information from other nonmedical sources (e.g., educational personnel, social welfare agency personnel). More generally the categories can be grouped as "objective medical evidence," applicant self-reports, and third-party reports (medical and nonmedical). According to SSA regulations, objective medical evidence refers to medical signs[7] and laboratory findings.[8] Laboratory findings must be demonstrated through "medically acceptable laboratory diagnostic techniques," among which SSA includes psychological tests (20 CFR § 404.1528).

SSA's use of the term *objective medical evidence* to refer to observable medical signs and laboratory or test results implies that the other types of evidence are "subjective" and therefore, perhaps, less reliable, which creates a tension among the different types of evidence that SSA considers.

[7] "Signs are anatomical, physiological, or psychological abnormalities which can be observed, apart from [self-reported symptoms]. Signs must be shown by medically acceptable clinical diagnostic techniques. Psychiatric signs are medically demonstrable phenomena that indicate specific psychological abnormalities, e.g., abnormalities of behavior, mood, thought, memory, orientation, development, or perception. They must also be shown by observable facts that can be medically described and evaluated" (20 CFR § 404.1528).

[8] "Laboratory findings are anatomical, physiological, or psychological phenomena which can be shown by the use of medically acceptable laboratory diagnostic techniques. Some of these diagnostic techniques include chemical tests, electrophysiological studies (electrocardiogram, electroencephalogram, etc.), roentgenological studies (X-rays), and psychological tests" (20 CFR § 404.1528).

This may arise particularly for categories of claims in which impairments are established and assessed primarily on reports of signs and symptoms of impairment and functional limitation (e.g., mental impairments other than intellectual disability, certain musculoskeletal conditions). It is important to note, as discussed in Chapter 4, that self-report measures can be valid assessment tools. In addition, SSA considers the consistency of all the evidence in a record to establish confidence in the validity of the claim of impairment and functional limitation.

If the information is insufficient to make a determination, the examiner generally tries to obtain additional information from the applicant's medical sources and, in some cases, other sources. Medical reports should include the applicant's medical history, clinical and laboratory findings, diagnosis, and prescribed treatment, including the applicant's response and prognosis. In addition, the report should include a statement about what the applicant can still do, including, for adults, the physical and/or cognitive ability to perform work-related activities. For children, the statement should discuss the child's functional limitations relative to other children of the same age (SSA, n.d.-a).

If the information requested from the applicant's treating and other sources is unavailable or remains insufficient (e.g., lacking in necessary detail or conflicting, inconsistent, or ambiguous) to make a determination, DDS may arrange for a consultative examination (CE) to obtain additional information needed to evaluate the claim (20 CFR § 404.1519a). In 2013, 45.1 percent of disability applicants received a CE as part of the initial disability determination process (SSA, 2014d). CEs were more commonly acquired for SSI and concurrent SSDI/SSI adult applicants than for SSDI applicants (SSA, 2014d). The minimum requirements for CE reports for mental disorders in adults and children can be found in the SSA's consultative examination guide for health professionals (SSA, n.d.-k). (See also for adults, SSA [2014e] and for children SSA [2012c].)

Appeals Process

If the DDS denies an application, the applicant can appeal the decision in turn to (1) the DDS (reconsideration), (2) an administrative law judge (ALJ), (3) the Appeals Council, and (4) a federal court.[9] Data on the number of applicants who appeal their decision at each stage are available from SSA. Because it takes time for denied applicants to move through the various stages of the appeal process, data are available through 2010. The data show that approximately 55 percent of those who applied for

[9] A 10-state pilot program begun in 1999 permits a claimant to bypass reconsideration by DDS and submit the appeal directly to an ALJ.

SSDI or concurrent worker benefits in 2010 and were denied during the initial evaluation, appealed the decision (calculation based on data from the 2013 Annual Statistical Report on the SSDI program, Tables 61 and 62 [SSA, 2014b]).[10] The rates of appeal were slightly lower for denied SSI applicants. Approximately 45 percent of 2010 SSI adult applicants and 30 percent of 2010 SSI child applicants who were rejected in the initial determination process appealed their decisions (calculations based on data from the 2013 Annual Statistical Report on the SSI program, Tables 70 and 71 [SSA, 2014k]).

The first level of appeal, which takes place within the DDS, is a reconsideration of the original claim or, for SSI, a review of an initial determination. Reconsideration involves a complete review of the initial claim by an examiner and, where applicable, a medical consultant who did not participate in the original evaluation. DDSs are reported to approve about 5 percent of reconsideration claims (Morton, 2014).

If the reconsideration is denied, the next level of appeal is a hearing before an ALJ. ALJs are employed by SSA and, on appeal, review the evidence in an applicant's file, including any new evidence submitted by the applicant. The ALJ also may interview the applicant and any witnesses brought by the applicant, as well as relevant medical or psychological consultants, other health care providers, or vocational experts. The applicant or a representative also may question any of the other witnesses. After considering all of the evidence and testimony, the ALJ issues a written decision (SSA, n.d.-i). If the ALJ finds that additional evidence is needed, he or she may order a CE or otherwise seek further development of the case file (SSA, 2012f). Reportedly about 67 percent of the claims reviewed by ALJs overall are approved, although the approval rate varies among ALJs and can be much higher (Morton, 2014; SSA, 2015).

Claims that are denied at the ALJ level may be brought to the Appeals Council, which serves as the final level of appeal within SSA. The Appeals Council considers each case brought to it and either denies the request for review, if it agrees with the ALJ's decision; sends it for review by another ALJ, if it finds a technical or procedural error with the ALJ's decision; or decides the case itself and grants benefits to the applicant (Laurence, 2015; SSA, n.d.-h). About 16 percent of requests for review are returned for rereview by an ALJ. In fiscal year 2014, the Appeals Council received more than 155,000 new requests for review. The council processed more than 162,280 requests that year. The processing time averaged 374 days.[11]

[10] This figure includes concurrent SSDI/SSI applicants.

[11] The numbers in this paragraph have been updated from those provided in the prepublication version of the report and were provided by SSA on May 13, 2015.

If the Appeals Council dismisses or does not reverse an unfavorable decision by the ALJ, the applicant may contest SSA's final decision by filing a civil suit in U.S. district court (SSA, n.d.-g). In fiscal year 2013, more than 18,700 new cases were filed (SSA, n.d.-g). The federal judge agrees with or overturns the decision of the ALJ and the Appeals Council, thereby denying or awarding benefits, or sends the case back for re-review by the ALJ.

Returning to data for 2010, by the end of all stages of the appeal process, 53 percent of SSDI or concurrent worker applicants who appealed their initial denial ultimately received an award (calculation based on data from the 2013 Annual Statistical Report on the SSDI program, Tables 62 and 63 [SSA, 2014b]). The rates are lower for SSI applicants: 40 percent of SSI adult applicants and 27 percent of child applicants in 2010 were ultimately awarded benefits after appeal (calculations based on data from the 2013 Annual Statistical Report on the SSI program, Tables 71 and 72 [SSA, 2014k]).

Final Outcomes of the Disability Determination Process

The final award rate, which includes initial and appealed decisions, varies across disability programs but is always higher than the initial award rates given in Figures 2-2 and 2-3. Based on data for applicants who filed for benefits in 2010, final award rates for disability benefit applicants are around 55 percent for SSDI workers, including concurrent applicants; 40 percent for SSI adult applicants; and 45 percent for SSI child applicants (SSA, 2014b, Tables 61, 62, 63, 2014k, Tables 70, 71, 72).[12]

Variability in Outcomes Across States

Although state DDS offices and SSA follow the same disability determination and appeals process, award rates vary significantly by state, reflecting variation in both filing rates (applications per eligible population) (see Figure 2-4) and allowance rates (allowances per DDS determinations) (see Figure 2-5). Variation in these rates stems, in part, from factors outside of the direct control of DDS offices or SSA. Such factors include state-level differences in population characteristics, such as age, education, and impairment type, as well as differences in local labor market

[12] In 2010, there were still applications pending final approval. Allowance rates for earlier years with smaller numbers of pending decisions were slightly higher than those referenced here for 2010.

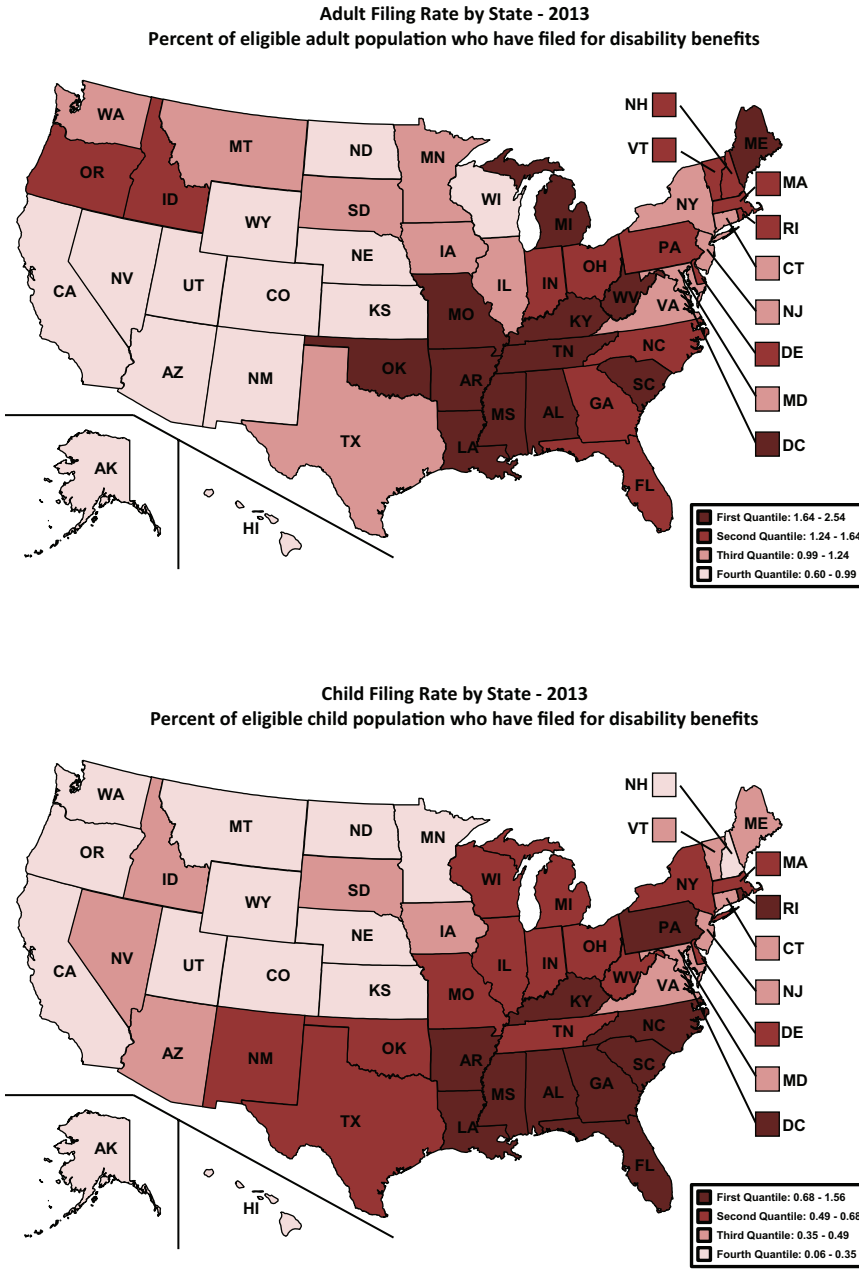


FIGURE 2-4 Filing rates by state, fiscal year 2013.
SOURCES: SSA, 2014b,k.

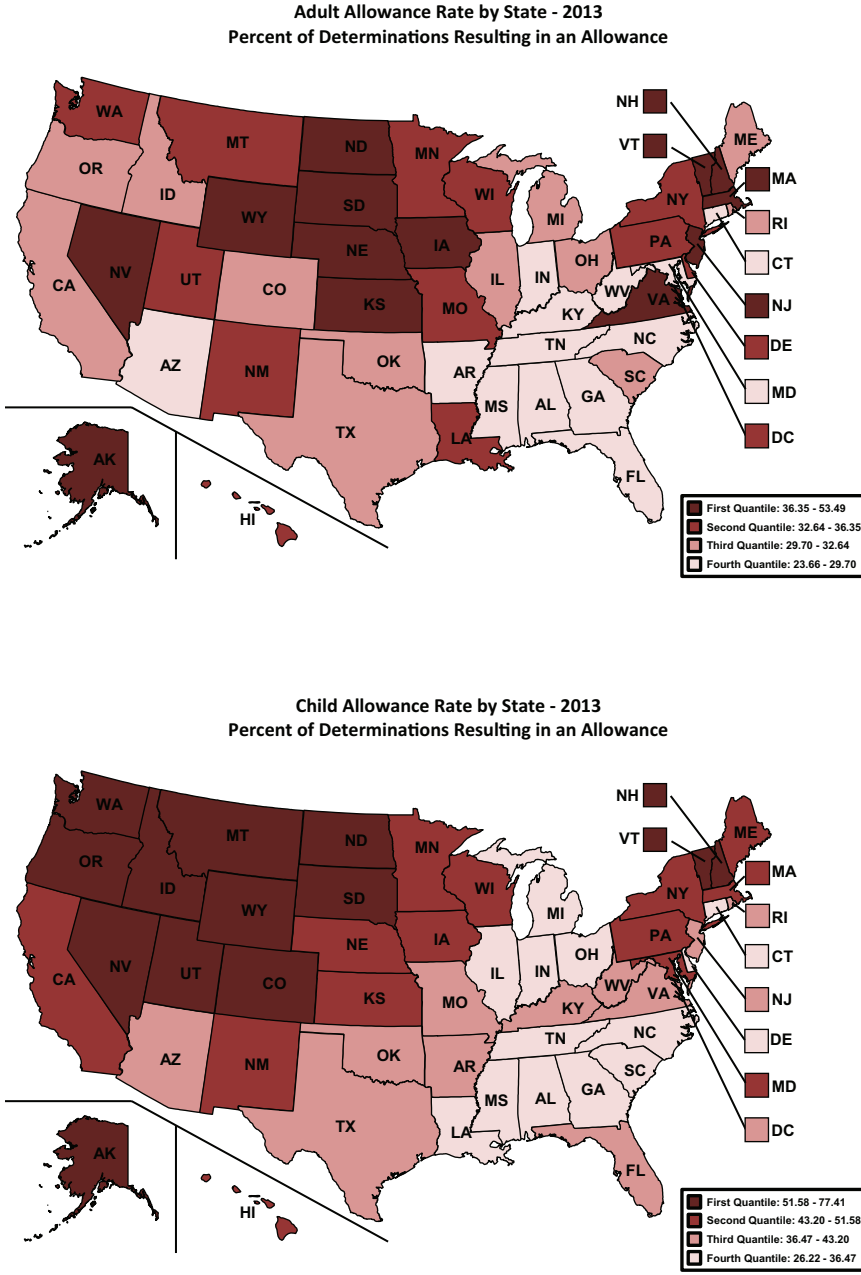


FIGURE 2-5 Allowance rates by state, fiscal year 2013.
SOURCES: SSA, 2014b,k.

conditions, such as the unemployment rate or mix of jobs available for workers with different skills.[13]

Several studies have attempted to quantify the degree to which state variation in application, allowance, and award rates is explained by these factors. In general the results suggest that observable state and individual characteristics account for half or more of the total variation. For example, Strand (2002) finds that controlling for state-level observables and year effects reduced variation in state-level allowance rates (1997–1999) by half. Soss and Keiser (2006) find similar reductions in variation for SSDI and SSI application rates.

Rupp (2012) decomposes overall cross-state variation in allowance rates for the 1993–2008 period and attributes it to one of four sources: (1) time-varying independent variables (unemployment rate and demographic and diagnostic criteria); (2) year fixed effects that capture national changes in economic conditions or policies affecting disability programs; (3) state fixed effects that capture unobservable, long-term differences across states that may or may not be related to DDS management; and (4) residual unexplained that captures the remaining variation not associated with any of the model variables (see Table 2-1).

The results show that time-varying independent variables explain a relatively small share of the state variation in allowance rates; about 10 percent for SSDI allowance rates and about 20 percent of variation in adult SSI and concurrent SSDI/SSI claims. Only 6 percent of the total variation in SSI child allowance rates is accounted for by the time-varying independent variables included in his model. Year fixed effects account for an additional small share of the variation in adult allowance rates (SSDI and SSI) but nearly 30 percent of the variation in SSI child allowances. Notably, between 40 and 50 percent of the overall variation in allowance rates across states is explained by long-term, unobservable state-specific differences. Combining these numbers with the amount unexplained by the model, the total variation in state allowance rates that cannot be traced back to observable variables outside of the DDS control is approximately 75 percent.

Although it is not possible to know definitively whether the large share of unexplained variation in state filing, award, and allowance rates is driven

[13] A long literature has documented the relationship between local labor market conditions, generally measured by the unemployment rate, and applications and awards for disability benefits. In general the results show that poor economic conditions/higher unemployment rates are associated with increased applications and awards for benefits (Autor and Duggan, 2003; Black et al., 2002; Burkhauser et al., 2002; Duggan and Imberman, 2008; Kreider, 1999; Rupp and Stapleton, 1995). Research on allowance rates and economic conditions (Rupp, 2012; Rupp and Stapleton, 1995; Strand, 2002) generally finds a negative relationship suggesting that SSA is able to screen out some marginally qualified candidates who might apply for the program in response to poor economic conditions.

TABLE 2-1 Components of Total Variation in Allowance Rates from Level Fixed-Effects OLS Regression Models, by SSA Program Group (in percent), 1993–2008

	Adult Program Group			
Component of Variation[a]	SSDI Only	SSI Only	Concurrent	SSI Child
State fixed effects	52	41	46	50
Year fixed effects	14	16	9	29
Time-varying independent variables (unemployment rate and demographic and diagnostic characteristics of applicants)	10	17	18	6
Unexplained[b]	24	25	27	16
Total	100	100	100	100

NOTES: A total of 12 regressions were estimated: 3 models for each of the 4 program groups. For each program group, independent variables were included in a sequential manner. The first model included only state fixed effects. The second model added year fixed effects. The third model added the time-varying variables. The results in this table reflect state-level OLS regression models. Totals may not sum to 100 because of rounding.

[a] The first row contains the R^2 from the first model for each program group. The subsequent two rows reflect the marginal increase in the R^2 arising from adding the given group of independent variables to the model. The total of the first three rows represents the R^2 for the third model that included all three groups of variables.

[b] The unexplained variation was calculated by subtracting the R^2 for the third model that included all of the predictors from 100 percent.

SOURCES: Data are based on 1,736,554 initial disability determinations in the 50 states and the District of Columbia for the 1993–2008 period, taken from SSA's National Disability Determination Services System File. State unemployment rate data are taken from the Current Population Survey. Reprinted with permission from Rupp, 2012, Table 9.

by variability in the federal disability determination process, there is some evidence that states differ in how they manage claims. For example, there are significant differences across states in the percentage of cases requiring a CE as part of the initial determination. Recall that nationally about 45 percent of initial determinations request a CE. By contrast, in low-CE states such as Hawaii, Missouri, and Virginia about one-quarter of cases receive a CE (SSA, 2014c). In high-CE states such as Indiana, Kentucky, and Tennessee about two-thirds of initial determinations request a CE (SSA, 2014c). That said, because the committee could locate no study of the variability of CE rates, this evidence is only suggestive of differences in case management across states.

COMPOSITION OF SSA BENEFICIARIES

Although there are no data on the composition of impairments affecting applicants, the data on allowed claims provide insight into the types of individuals seen at the state DDS offices. Figure 2-6 shows the composition of new beneficiaries in 2013 for SSDI and SSI adults and children. By far the largest two impairment categories for all three disability programs are mental disorders (excluding intellectual disabilities) and musculoskeletal and connective tissue disorders. In 2013, these two categories accounted for 52 percent of new SSDI awards, 53 percent of new SSI adult awards, and 58 percent of new SSI child awards. Within these two categories, a significant fraction of the applicants have conditions, including affective mood disorders and disorders of the back, for which the presence and severity of impairment and associated functional limitations are based largely on applicant self-report (SSA, 2014j,l).

The large share of these two categories in the flow of new beneficiaries indicates that DDS offices are evaluating a large number of cases that require more subjective judgment about the functional limitations the client faces. This is supported by the large number of adult cases that are determined on medical-vocational criteria at Steps 4 and 5 of the determination process: more than 50 percent of the initial DDS decisions and more than 80 percent of decisions at the hearing level (SSA, n.d.-l).

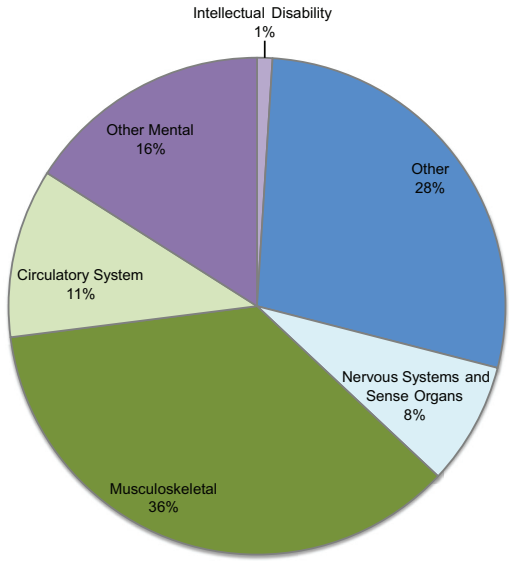
PSYCHOLOGICAL TESTING IN SSA DISABILITY EVALUATIONS

Policy Relevant to Evaluations of Disability for Mental Disorders

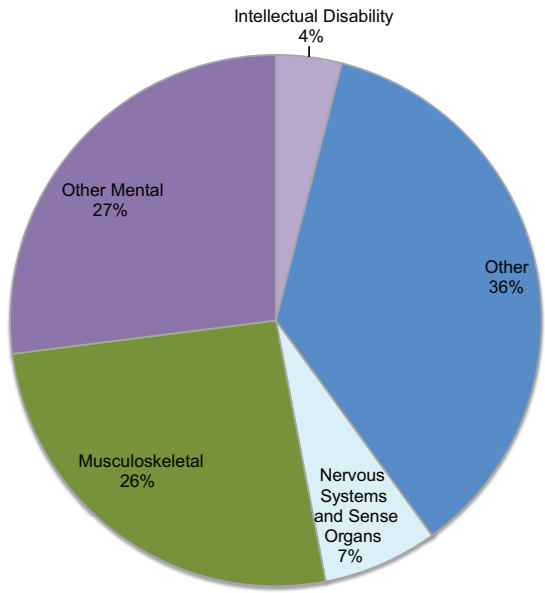
Adults who file for SSA disability on the basis of mental disorders and meet the nonmedical eligibility criteria are evaluated at Step 2 for the presence of a medically determinable mental impairment, the severity of the functional limitation it imposes on the individual's ability to work, and a determination that the impairment has lasted or will last for 12 or more continuous months (SSA, 2012d, n.d.-e). The DDS assesses the presence of a medically determinable mental impairment on the basis of the medical evidence, including relevant signs, symptoms, and laboratory or psychological test findings (SSA, 2012d).

The DDS assesses the severity of a medically determinable mental impairment on the basis of the functional limitations it imposes on the claimant's ability to engage in work-related activities. Functional limitations are assessed in four areas that are considered essential for work: (1) activities of daily living (ADLs); (2) social functioning; (3) concentration, persistence, or pace; and (4) episodes of decompensation in a work-like setting—or "the

New SSDI Worker Beneficiaries by Diagnostic Group, 2013



New SSI Adult Beneficiaries by Diagnostic Group, 2013



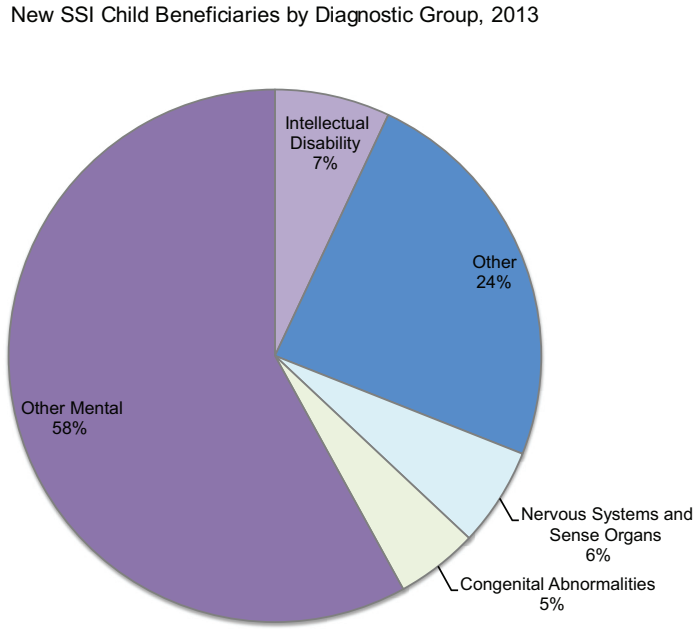


FIGURE 2-6 Composition of new beneficiaries in 2013 for SSDI and SSI adults and children.
SOURCES: SSA, 2014b,k.

ability to tolerate increased mental demands associated with competitive work" (SSA, 2009, section B). These areas correspond to the Paragraph B criteria,[14] which are part of the listings of impairments for mental disorders assessed at Step 3. A functional limitation is considered "marked" if it is "more than moderate but less than extreme"; in other words, the degree of limitation "interfere[s] seriously with [the claimant's] ability to function independently, appropriately, effectively, and on a sustained basis" (SSA, n.d.-e, section C).

ADLs and social functioning are evaluated within the contexts of (1) appropriateness, (2) independence, (3) sustainability, (4) quality, and (5) effectiveness (SSA, 2009). Information about the claimant's ADLs

[14] Under a notice of proposed rulemaking, SSA has proposed revised Paragraph B criteria to capture "the mental abilities an adult uses to function in a work setting" (SSA, 2010, p. 51340). The revised B criteria are the abilities to "understand, remember, and apply information"; "interact with others"; "concentrate, persist, and maintain pace"; and "manage oneself."

and social functioning is acquired through interview, self-report, observation, and other report. Concentration, persistence, or pace "refers to the ability to sustain focused attention sufficiently long to permit the timely completion of tasks commonly found in work settings" (SSA, 2009, section D). These functions may be assessed with a mental status exam or psychological tests, but such tests represent a point in time and do not necessarily reflect the ongoing stresses of a work environment. Clinical and test data should be supplemented by other evidence, such as observations of performance in a work or work-like setting.

The inability to tolerate the increased demands associated with work (deterioration or decompensation) is demonstrated by an increase in the signs or symptoms and the need for new or additional treatment or removal from the stressful environment. Generally to meet the criteria the claimant would have had at least three episodes, each lasting 2 weeks or longer, in the most recent year.

Step 2 is the first point at which the results of cognitive and noncognitive tests can help inform SSA's disability determination process. The results of such tests can help support the identification and documentation of the presence and severity of medically determinable mental impairments. It is important to note that an individual's level of functioning can fluctuate over time. To evaluate an individual's impairment accurately, it is important for DDS examiners to obtain evidence across a long enough timeframe (SSA, 2012d).

Applicants who meet the criteria at Step 2 are evaluated at Step 3 to determine whether they meet or equal the criteria in the *Listing of Impairments* for mental disorders (SSA, n.d.-e, n.d.-f). The listings for mental disorders include 9 diagnostic categories for adults[15] and 11 categories for children, of which the first 9 are similar to the adult listings:

1. Organic mental disorders
2. Schizophrenic, paranoid, and other psychotic disorders
3. Affective (mood) disorders
4. Intellectual disability disorders
5. Anxiety-related disorders
6. Somatoform disorders[16]
7. Personality disorders
8. Substance addiction disorders
9. Autistic disorder and other pervasive developmental disorders
10. Attention deficit hyperactivity disorder (children)

[15] Under the same notice of proposed rulemaking (SSA, 2010), SSA has proposed revised listing categories.

[16] Somatoform disorders are discussed separately in the following section.

11. Developmental and emotional disorders of newborn and younger infants (children)

For most of the diagnostic categories,[17] adult applicants will meet a listing if the impairment satisfies the following: (1) the diagnostic description of the mental disorder; (2) specified medical findings—e.g., symptoms (self-report), signs (medically demonstrable), laboratory findings (including psychological test findings)—(Paragraph A criteria); and (3) specified "impairment-related functional limitations that are incompatible with the ability to do any gainful activity" (Paragraph B or Paragraph C criteria) (SSA, n.d.-e). Paragraph A criteria, in conjunction with the diagnostic description, substantiate the presence of the specific mental disorder based on the medical evidence. Paragraph B and Paragraph C criteria list the functional limitations resulting from the mental impairment that preclude the ability to engage in gainful activity. Cognitive and non-cognitive test results can inform disability determinations at Step 3, particularly with respect to Paragraph A and B criteria.

If an applicant's impairment does not meet the diagnostic definition or the Paragraph A criteria of a listing but does result in the functional limitations specified in the Paragraph B or C criteria, the impairment is considered to equal the listing. Applicants whose impairments are severe but do not meet or equal any of the listings are not approved at Step 3. They move on to an evaluation of their residual function capacity at Steps 4 and 5 of the determination process. Residual functional capacity refers to the work-related capacities an applicant still possesses despite the impairment. Assessment of residual functional capacity is another area of the determination process that the results of psychological testing could inform.

The determination process differs somewhat for children at Step 3. In addition to asking whether the child's impairment(s) meets or medically equals one of the listings, a second question is posed if it does not: Does the impairment functionally equal the listings? By "functionally equal the listings," SSA means that "the impairment(s) must be of listing-level severity; i.e., it must result in 'marked' limitations in two domains of functioning or an 'extreme' limitation in one domain" (20 CFR § 416.926a). The functional limitations caused by the child's impairment(s) are assessed. In determining functional equivalence, SSA considers "the interactive and cumulative effects of all of the impairments for which [it has] evidence,

[17] The structure of the listing for intellectual disability and for substance addiction disorders differ from that of the other mental disorder listings. There are four sets of criteria (Paragraphs A through D) for the intellectual disability listing, and the listing for substance addiction disorders refers to which of the other listings should be used to evaluate the various physical or behavioral changes related to the disorder.

including any impairments [the child has] that are not 'severe' (see § 416.924(c))" (20 CFR § 416.926a). When assessing a child's functional limitations, it considers "how appropriately, effectively, and independently [the child] performs … activities compared to the performance of other children [the same] age who do not have impairments" (20 CFR § 416.926a).

Documentation

As previously described, the DDS uses all relevant evidence in an applicant's file in making a disability determination. The medical evidence in an applicant's file must be sufficiently complete and detailed to allow the DDS to make a determination. Medical evidence includes a history of the individual's mental impairment, the results of any mental status examinations and psychological tests, and the records of any treatments and hospitalizations provided by an "acceptable medical source" (SSA, 2014f, n.d.-e).

Although a full mental status exam, performed during a clinical interview, can be tailored to target the specific areas most relevant to the alleged impairment, a comprehensive exam generally would include "a narrative description of [the individual's] appearance, behavior, and speech; thought process (e.g., loosening of associations); thought content (e.g., delusions); perceptual abnormalities (e.g., hallucinations); mood and affect; sensorium and cognition (orientation, recall, concentration, intelligence); and judgment and insight" (SSA, n.d.-e, section D4).

Psychological Testing

SSA understands "standardized psychological tests" to be psychological test measures that have "appropriate validity, reliability, and norms" representative of relevant populations (SSA, n.d.-e, section D5). SSA characterizes a "good test" as one that is valid ("measures what it is supposed to measure") and reliable (use of the same test in the same individual yields consistent results over time) and has "appropriate normative data" and a "wide scope of measurement" (measures a broad range of elements of the domain being assessed) (SSA, n.d.-e, section D5).

SSA specifies the tests would be administered, scored, and interpreted by a "qualified" specialist—meaning someone "currently licensed or certified in the state to administer, score, and interpret psychological tests" with the "training and experience to perform the test" (SSA, n.d.-e, section D5). The types of specialists who are qualified to administer, score, and interpret standardized psychological tests are discussed in Chapters 3, 4, and 5. Observations of the test administrator—such as ability to concentrate, interact appropriately with test administrator, perform independently—would supplement the report of test results. The report would also address

the validity of the test results, including discussion of any discrepancies between the test results and "the individual's customary behavior and daily activities" (SSA, n.d.-e, section D5).

The results of standardized intelligence tests are built into the listings for intellectual disability and some neurological impairments. In addition, SSA notes that intelligence test results can help to confirm the presence of intellectual disability and organic mental disorders as well as the severity of cognitive impairment. SSA states that standardized personality measures (e.g., Minnesota Multiphasic Personality Inventory-2) or projective testing techniques (e.g., Rorschach) may provide useful data for the evaluation of disability "when corroborated by other evidence, including results from other psychological tests and information obtained in the course of the clinical evaluation" (SSA, n.d.-e, section D7). SSA also states that "comprehensive neuropsychological examinations may be used to establish the existence and extent of brain function, particularly in cases involving organic mental disorders" (SSA, n.d.-e, sections D6, D7, D8).

Psychological Consultative Examinations

SSA specifies the minimum content requirements for CE reports for adults with mental disorders (SSA, n.d.-k, Part IV, Mental Disorders). These requirements include the following: applicants' longitudinal, current, and past medical history; current medications; social and family history; physical examination; and mental status evaluation.[18] In addition, the report is to include interpretation of any psychological and/or clinical test results in relation to the history and examination findings as well as identification of the individual providing the interpretation if different from the provider signing the CE report (SSA, n.d.-k, Part IV, Mental Disorders, section H). The report also is to specify "a full multiaxial classification as set forth in the current *Diagnostic and Statistical Manual of Mental Disorders*" and prognosis and recommendations for treatment, if indicated (SSA, n.d.-k, Part IV, Mental Disorders, section I).

For applicants with intellectual impairments, current documentation of intelligence quotient (IQ) is required along with interpretation of the results, including an assessment of their validity, and consistency of the results "with the claimant's educational, vocational, and social background" (SSA, n.d.-k, Part IV, Mental Disorders, section I). Also required is "a

[18] Elements include "(1) manner and approach to evaluation; (2) dress, grooming, hygiene and presentation; (3) mood and affect; (4) eye contact; (5) expressive/receptive language; (6) recall/memory, including working, recent, and remote; (7) orientation in all four spheres; (8) concentration and attention; (9) thought processes and content; (10) perceptual abnormalities; (11) suicidal/homicidal ideation; (12) judgment/insight; and (13) estimated level of intelligence" (SSA, n.d.-k, Part IV, Mental Disorders, section G).

comprehensive and detailed description of adaptive behavior in the areas of personal, social, academic, and occupational functioning during the developmental period" (SSA, n.d.-k, Part IV, Mental Disorders, section I).

Additionally, SSA specifies that CE reports for mental disorders should include statements from the medical source regarding "the nature and extent of the mental disorder" and "an assessment of the claimant's abilities and limitations based on medical history, observations during examination, and results of relevant laboratory tests" as well as an opinion regarding the applicant's ability to carry out certain functions (SSA, n.d.-k, Part IV, Mental Disorders, section J). The report should discuss "any apparent discrepancies in medical history or in examination findings and how the discrepancies were resolved"; include "a statement regarding malingering, if applicable"; and "a statement regarding the [applicant's] capability to manage funds" (SSA, n.d.-k, Part IV, Mental Disorders, section J).

In practice, CEs for mental disorders generally consist of nonstandardized diagnostic interviews and mental status exams, with little to no standardized psychological testing other than intelligence testing (Chafetz, 2008; Chafetz et al., 2007; Griffin et al., 1996; Heiser, 2014; McLaren, 2014; Price, 2014; Ward, 2014).

Aside from the use of intelligence tests as described in the listings for intellectual disability and certain neurological impairments, SSA does not require or specify the purchase of any type of or individual psychological test. The primary guidance provided by SSA is that good psychological tests are valid, reliable, and appropriately normed, and have a wide scope of measurement, as previously described. In addition, as discussed later under *Use of Validity Tests*, current SSA policy precludes the purchase of validity tests except in rare cases, such as a court order.

Policy Relevant to Evaluations of Disability for Somatic Symptoms Disproportionate to Demonstrable Medical Morbidity

There are three distinct groups of applicants seeking disability compensation for somatic symptoms unaccompanied by demonstrable anatomical, biochemical, or physiological abnormalities: somatoform disorders (recently termed *somatic symptom disorders* in the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* [*DSM-5*]); multisystem illnesses; and chronic idiopathic pain conditions.

In all three of these types of conditions—somatoform disorder, multisystem illness, and chronic pain—the credibility, reliability, validity, or accuracy of the reported symptoms and/or impairment may be called into question. This is due to the absence of objective evidence or biomarkers that could explain or substantiate the applicant's report of subjective distress and disability. When relying on self-report of symptoms and impairment,

SSA policy states that applicants may not be found disabled solely on the basis of self-reported statements about pain or other symptoms (Social Security Act § 223(d)(5)(A), § 1614(a)(3)(D); 20 CFR 404.1508, 404.1529, 416.908, 416.929; SSA, 1996b, 2014g).

In cases where an individual's self-reported symptoms, including pain, suggest a greater degree of impairment than expected based on the objective medical evidence alone, other corroborative information from treating and nontreating medical sources and other sources is considered. Such information may include information about the individual's

> daily activities; the location, duration, frequency, and intensity of [the] pain or other symptoms; precipitating and aggravating factors; the type, dosage, effectiveness, and side effects of any medication … taken to alleviate [the] pain or other symptoms; treatment, other than medication …; any measures … used to relieve [the] pain or other symptoms …; and other factors concerning [the individual's] functional limitations and restrictions due to pain or other symptoms. (20 CFR 404, Subpart P, § 404.1529; 20 CFR 416, Subpart I, § 416.929)

SSA has issued guidance on its policy for evaluating claims involving chronic fatigue syndrome (CFS) (SSA, 2014g). This guidance explains how SSA determines the presence of a medically determinable impairment in an individual with CFS, including some of the possible medical signs and laboratory findings that would help to support such a finding. SSA then assesses whether the medically determinable impairment could reasonably be expected to produce the reported symptoms. In cases where objective medical evidence does not substantiate the person's statements, SSA considers the same types of evidence described for pain and other symptoms. SSA will also make a finding about the credibility of the person's statements as described in the following section.

Policy on the Evaluation of Credibility

Assessing Credibility of Statements About Pain and Other Symptoms

Given that symptoms—"individual's own description[s] of his or her physical or mental impairment(s)"—are insufficient under SSA regulations "to establish the existence of a physical or mental impairment or that the individual is disabled," the regulations provide a two-step process for evaluating statements about pain, fatigue, weakness, and other symptoms (SSA, 1996c). The first step is to determine whether the individual has a medically determinable impairment that could reasonably be expected to produce the symptoms. If so, the second step is to evaluate the intensity and persistence of the symptoms and their effect on the applicant's ability to function and perform work-related activities.

Given the subjective nature of symptoms such as pain, fatigue, nervousness, and the like, "objective medical evidence"—such as medical signs and laboratory findings—does not always substantiate the severity of an impairment as experienced by individuals and expressed in their self-reported symptoms. If the objective medical evidence does not support an individual's statements about the intensity, persistence, and limiting effects of the symptoms, the examiner must determine the credibility of the statements based on all of the information in the case record (SSA, 1996c).

When determining the credibility of an applicant's statements about symptoms, SSA states the examiner must consider specific indicators of credibility such as:

- Consistency, both internally (i.e., with other statements by the applicant) and with other information in the record (e.g., objective medical evidence, third-party reports and observations);
- The extent to which objective medical evidence may inform conclusions about the intensity and persistence of reported symptoms, even if the latter are not objectively measurable; and
- The individual's longitudinal medical record (history) of persistence and severity of reported symptoms.

SSA requires the examiner to articulate specific reasons for the credibility finding based on the medical and other evidence in the case record. It is important to note both that a credibility finding need not reflect complete acceptance or rejection of the individual's statements (i.e., the statements may be found to be partially credible) and that credibility concerns alone do not rule out the presence of disability (SSA, 1996c).

Use of Validity Tests

With rare exceptions, such as a court order, current SSA policy precludes the purchase of (validity) tests[19] to help inform determinations about the credibility of an individual's statements or about possible malingering (SSA, 2012e, 2013). It is SSA's position that "tests cannot prove whether a claimant is credible or malingering because there is no test that, when passed or failed, conclusively determines the presence of inaccurate self-reporting" (SSA, 2013, section D), although SSA acknowledges that the

[19] Such tests include the following: Rey-15 Item Memory Test (Rey-II), Miller Forensic Assessment of Symptoms Test (M-FAST), Millon Clinical Multiaxial Inventory, Minnesota Multiphasic Personality Inventory (MMPI), Minnesota Multiphasic Personality Inventory-2 (MMPI-2), Malingering Probability Scale, Structured Interview of Reported Symptoms, Test of Memory Malingering, and Validity Indicator Profile (SSA, 2008, 2013).

results of such tests "can provide evidence suggestive of poor effort or of intentional symptom manipulation" (SSA, 2008). Nevertheless, SSA will consider, along with all other relevant evidence, the results of symptom validity tests (SVTs) that are already in the claimant's file (SSA, 2013). According to a 2013 report from the Office of the Inspector General, SSA:

> The Agency disallowed the purchase of SVTs because of weaknesses in their psychometric properties and limited value in determining, with certainty, a claimant's credibility. In addition, SSA stated that in cases where there was a high likelihood of malingering, the circumstances did not preclude the person from having a genuine medically determinable impairment. (Office of the Inspector General, SSA, 2013)

There appears to be some confusion or inconsistency among SSA's statements regarding validity testing. On the one hand, SSA clearly rejects the purchase of performance validity tests (PVTs) and SVTs by DDS and consultative examiners with statements such as the following:

- "Malingering cannot be proven with tests";
- "Malingering is one aspect of the larger sphere of inaccurate self-reporting";
- "No test … conclusively determines the presence of inaccurate patient self-report"; and
- "Even a high likelihood of malingering does not preclude severe limitations resulting from a genuine medically determinable impairment."[20]

On the other hand, SSA acknowledges that validity test results can "provide evidence suggestive of poor effort or intentional symptom manipulation" and states that it will consider validity test results that are already in an applicant's file, along with all other relevant evidence. In fact, the statement that no one test "conclusively determines the presence of inaccurate patient self-report" seems to run counter to SSA's dedication to obtaining as much evidence as possible and taking account of all the information when making a disability determination. It is important to divorce the concept of "malingering" from that of validity testing. As introduced in the following section, and made clear later in this chapter and elsewhere in the report and appendixes, validity test results can speak to performance (on performance-based tasks) and to the consistency and accuracy of responses on self-report measures. However, they provide limited information about intentionality and none about motive. It is important, therefore, not to discount the potential usefulness of validity test results on the grounds

[20] Quotations are taken from SSA (2008).

that malingering cannot be proven with tests or that a high likelihood of malingering and the presence of severe limitations resulting from a genuine medically determinable impairment cannot coexist.

MALINGERING AND CREDIBILITY

Malingering Base Rates

As defined in Chapter 1, *malingering* is the intentional presentation of false or exaggerated symptoms, intentionally poor performance, or a combination of the two, motivated by external incentives (APA, 2015; Bush et al., 2005; Heilbronner et al., 2009). Base rates of "probable malingering and symptom exaggeration,"[21] as reported in a 2002 survey of members of the American Board of Clinical Neuropsychology, vary depending on the alleged impairment (e.g., mild head injury, depressive or anxiety disorders, seizure disorders, vascular dementia), the context (e.g., personal injury or disability, criminal, medical or psychiatric), and the referral source (e.g., plaintiff, defense) (Mittenberg et al., 2002). All of these factors make direct comparisons of the reported rates difficult. For this reason, the discussion in this section focuses on studies of "malingering" in the disability context.

The studies described here suggest that anywhere from 19 to 68 percent of SSA disability applicants may be performing below their capability on cognitive tests or inaccurately reporting their symptoms. A number of factors may account for the vast range, including differences in what precisely is being reported, differences in the tests administered or the indicators (e.g., patterns of performance, inconsistencies among different sources of information) being used, and differences in the populations being examined. It is notable that a number of these articles refer to "malingering," "probable malingering," or "definite malingering" (see, e.g., Chafetz et al., 2007; Larrabee, 2007; Mittenberg et al., 2002; Samuel and Mittenberg, 2005). What is being reported, however, are either failure rates at different levels (e.g., below chance, at chance, below cut score, failure on two or more validity measures) on various PVTs or SVTs or other indicators, such as inconsistencies or discrepancies in the evidence.

[21] Respondents were asked the extent to which each of the following supported such an assessment in their cases: "below empirical cut-off on forced-choice tests"; "below chance on forced-choice tests"; "below empirical cut-off on other malingering tests"; "pattern of cognitive test performance does not make neuropsychological sense (inconsistent with condition)"; "severity of cognitive impairment inconsistent with condition"; "implausible changes in test scores across repeated examinations"; "above validity scale cut-offs on objective personality tests"; "discrepancies among records, self-report, and observed behavior"; and "implausible self-reported symptoms in interview" (Mittenberg et al., 2002, p. 1102).

The following discussion, summarized in Table 2-2, focuses on the reported base rates of validity test failure in the context of disability claims and specifies what is being measured in each case.

In 1996, Griffin and colleagues reported on 167 SSA disability applicants alleging psychological impairment in Los Angeles County between December 1993 and December 1994 (Griffin et al., 1996). As part of their psychological evaluation, these applicants were administered the Composite Disability Malingering Index (CDMI), a research tool created from portions of the Minnesota Multiphasic Personality Inventory (MMPI), the M Test, the Millon Clinical Multiaxial Inventory-II, and the Beck Depression Inventory. Nineteen percent (n = 32) of the 167 applicants assessed scored at a level identified as "malingering." The CDMI scores for this group more closely resembled those of a group of disability examiners who were instructed to malinger than those of the comparison group of psychologically disabled individuals with no incentive to malinger. The subgroup identified as "malingering" differed from the rest of the disability applicant group only in the presence of a self-reported history of substance abuse.

In their 2002 survey, Mittenberg and colleagues (2002) found a base rate of "probable malingering or symptom exaggeration," as described in note 17, of approximately 30 percent (reported) to 33 percent (adjusted)[22] for disability or worker's compensation cases. The rate varied relative to the referral source, with patients referred by defense attorneys or insurers having a higher rate of "probable malingering or symptom exaggeration." Their estimates were based on a total of 33,532 cases reported in surveys returned by 131 of 375 possible respondents among the 388 members of the American Board of Clinical Neuropsychology. Eleven percent of the cases involved disability or worker's compensation (n = 3,688), 19 percent (n = 6,371) involved personal injury litigation, 4 percent (n = 1,341) involved criminal litigation, and 66 percent (n = 22,131) were medical or psychiatric cases not involving litigation or compensation. The reported base rate of "probable malingering or symptom exaggeration" in the last group was only 8 percent (Mittenberg et al., 2002, pp. 1095–1096).

In a sample of adult SSA disability applicants, Chafetz and Abrahams found that 13.8 percent scored below chance performance and 58.6 percent failed two or more validity indicators (Chafetz and Abrahams, 2005, reported in Larrabee, 2007). Miller and colleagues (2006) reported that more than 50 percent of 105 disability applicants failed "conservative criteria" for the Computerized Assessment of Response Bias.[23]

[22] The adjusted value is corrected to remove significant variation due to referral source.

[23] The information and data in this sentence have been revised from that provided in the prepublication version of the report.

TABLE 2-2 Summary of Reported Base Rates of Malingering

Source	Percent and Population	Definition	Tool
Griffin et al., 1996	19 percent Disability claimants reporting psychological impairment (n = 167)	Scored at a level defined as "malingering"	Composite Disability Malingering Index (CDMI): created from portions of the Minnesota Multiphasic Personality Inventory, M Test, Millon Clinical Multiaxial Inventory-II, and Beck Depression Inventory
Mittenberg et al., 2002	30–33 percent Disability or worker's compensation cases (n = 3,688)	"Probable malingering or symptom exaggeration" (see note 19)	Survey of members of the American Board of Clinical Neuropsychology
Chafetz and Abrahams, 2005, reported in Larrabee, 2007	13.8 percent 58.6 percent Social Security Administration (SSA) adult disability applicants	Below chance Failed two or more validity indicators	
Miller et al., 2006[a]	> 50 percent Disability applicants (n = 105)	Failed "conservative criteria" (< 90 percent correct)	Computerized Assessment of Response Bias
Chafetz et al., 2007	55.8 percent (adults); 28.3 percent (children) 12.4 percent (adults); 8.7 percent (children) 61.4 percent (adults); 37 percent (children) 12.3 percent (adults); 7.4 percent (children) 51.6–58.9 percent (adults); 34.6–43.8 percent (children) 20.5–30.4 percent (adults); 15.4–32.5 percent (children) SSA adult and child disability applicants, most with low cognitive functioning TOMM (n = 136 adults, 96 children) MSVT (n = 58 adults, 27 children)	Failed Below chance Failed Below chance Failed Below chance	Test of Memory Malingering (TOMM) Medical Symptom Validity Test (MSVT) Disability Determination Services (DDS) Malingering Rating Scale

TABLE 2-2 Continued

Source	Percent and Population	Definition	Tool
Chafetz, 2008	67.8 percent (adults)	Failed at least one	TOMM and/or DDS
	45.8 percent (adults)	Failed both	Malingering Rating Scale
	36.5 percent (adults)	At or below chance	
	68.4 percent (adults)	Failed at least one	MSVT and/or DDS
	59.7 percent (adults)	Failed both	Malingering Rating Scale
	47.4 percent (adults)	At or below chance	
	60 percent (children)	Failed at least one	TOMM and/or DDS
	26.3 percent (children)	At or below chance	Malingering Rating Scale
	48 percent (children)	Failed at least one	MSVT and/or DDS
	20 percent (children)	At or below chance	Malingering Rating Scale
	SSA adult and child disability applicants, most with low cognitive functioning		
	TOMM (n = 136 adults, 96 children)		
	MSVT (n = 58 adults, 27 children)		

[a] The information in this entry has been revised from that provided in the prepublication version of the report.

Chafetz and colleagues administered the Test of Memory Malingering (TOMM) or the Medical Symptom Validity Test (MSVT) to adult and child disability applicants, most with low cognitive functioning, who were referred for a psychological CE by the DDS (Chafetz et al., 2007). Based on their performance on the test, subjects' performance was scored as "below chance," "chance or below," or "failing." In this study, 55.8 percent of adults (n = 136) and 28.3 percent of children (n = 96) failed the TOMM, and 12.4 percent of adults and 8.7 percent of children scored below chance on the test. On the MSVT, 61.4 percent of adults (n = 58) and 37.0 percent of children (n = 27) failed, and 12.3 percent of adults and 7.4 percent of children scored below chance.

The same study was designed to validate a tool, the "DDS Malingering Rating Scale," developed by the authors to help psychologists assess and inform DDSs about the validity of their findings (Chafetz et al., 2007).[24] The rating scale was validated against the TOMM and the MSVT and was found to correlate well with "formal tests and indicators of effort in adults and children" (Chafetz et al., 2007, p. 11). Fifty-one point six (51.6) to 58.9 percent of adults and 34.6 to 43.8 percent of children failed the DDS

[24] To the committee's knowledge, the "DDS Malingering Rating Scale" has never been used or endorsed by any DDS agencies.

Malingering Rating Scale, and 20.5 to 30.4 percent of adults and 15.4 to 32.5 percent of children scored below chance (Chafetz et al., 2007, p. 10).

In a subsequent paper that draws on the research reported in Chafetz and colleagues (2007), Chafetz reports 67.8 percent of adults who were administered both the TOMM and the DDS Malingering Rating Scale failed at least one, 45.8 percent failed both, and 36.5 percent scored at or below chance. For adults who were administered both the MSVT and the rating scale, 68.4 percent failed at least one, 59.7 percent failed both, and 47.4 percent scored at or below chance on at least one of the SVT subtests. Sixty percent of children who were administered the TOMM and the rating scale failed at least one and 26.3 percent scored at or below chance. Of children who were administered the MSVT and the rating scale, 48 percent failed at least one, and 20 scored at or below chance on at least one of the SVT subtests (Chafetz, 2008).

In the context of SSA disability evaluations, it is important to note that even if an applicant performs below his or her capability on cognitive tests or inconsistently reports symptoms, neither scenario means the individual is not disabled. However, both scenarios suggest the need for additional assessment of the alleged impairment with the goal of making an accurate determination of disability. Doing so first requires identification of the individuals for whom additional assessment may improve the accuracy of the disability determination. As described in the section on assessing credibility, when a disability claim is based primarily on an applicant's self-report of symptoms and statements about their intensity, persistence, and limiting effects, SSA relies on an assessment of the consistency of the self-report with all of the evidence in the claimant's medical evidence record. As discussed, SSA policy currently precludes the purchase of validity tests by SSA (e.g., as part of a psychological CE). One question is whether the results of this type of standardized test could contribute to the evidence available for assessment. The following section discusses the potential value of adding standardized data collection and interpretation to clinical data collection and evaluation.

The Benefits of Mechanical Data Collection and Actuarial Data Interpretation

A robust literature demonstrates that people, including experts, are systematically overconfident in their ability to perform a wide range of tasks (Moore and Healy, 2008), from investing in the stock market (Scheinkman and Xiong, 2003) to estimating their level of general knowledge (Juslin, 1994; Oskamp, 1965). This overconfidence exists in large part because human judgment is influenced by biases that operate outside of conscious awareness (Kahneman, 2011). People believe they come to judgments by

rationally weighing evidence, unaware that other psychological forces are also influencing them.

This overconfidence extends to the judgment of practicing psychologists with obvious consequences for the accuracy of psychological evaluation (Oskamp, 1965). Clinicians may rely on clinical judgment alone to determine the degree of effort put forth on performance-based cognitive and behavioral tests and the credibility of an examinee's self-report, even though research has shown that when people have been coached to exaggerate the symptoms of neurocognitive impairment, most clinicians failed to detect such malingering (Faust et al., 1988a,b; Heaton et al., 1978; Oldershaw and Bagby, 1997).

The literature comparing clinical versus actuarial (statistical) judgment suggests the best approach will (1) collect both clinical and structured data, and (2) combine these data using actuarial methods. Of course, considerable research is needed to establish the exact actuarial approach to be used.

Defining Terms

Data collection Medical professionals often evaluate patients using a combination of what Wedding and Faust call clinical and mechanical data (Wedding and Faust, 1989). *Clinical data collection* includes all testing and examining that is variable depending on how the clinician performs the exam and/or on which aspects of the exam the clinician chooses to perform. For example, clinicians may interview patients to elicit their description of the symptoms of their illness; alternatively, clinicians may perform a physical exam. By contrast, *mechanical data collection* involves the use of standardized testing where the data collection is structured and the method typically does not vary from patient to patient. For example, if clinicians order a serum sodium level or MMPI tests on their patients, they are collecting mechanical data.

It should be noted that mechanical data collection is not completely divorced from clinical expertise. For example, clinicians may need to determine which mechanical data are relevant to collect in a given patient, making a judgment about whose diagnosis will be aided by a serum sodium level or an MMPI. In addition, the administration of mechanical tests can be affected by clinical skill. For example, a clinician who draws a patient's blood above an IV site will get a false sodium level. Similarly, a clinician who administers an MMPI test after the patient has been exhausted by previous examinations may also be collecting the data in a way that will reduce the value and accuracy of the test results.

Data interpretation Once data have been collected—whether clinical data, mechanical data, or some combination of both—they must be interpreted

to determine whether the patient has a specific health condition and to estimate how severe that condition is. Data interpretation generally takes one of two approaches: clinical or actuarial. In clinical data interpretation, a clinician looks at all the data and makes a judgment. ("Based on your age, family history, chest pain, and ECG [electrocardiogram], I think you are having a heart attack.") In actuarial data interpretation, data are entered into a diagnostic program and weighed according to a statistical procedure. ("The presence of chest pain, given your age, family history, and ECG changes, yields a risk score of x, which estimates the probability of a heart attack to be y.")

What Are the Evaluative Alternatives?

There is a range of possible approaches to the evaluation of people complaining of behavioral or cognitive impairments. At one extreme is a purely clinical evaluation, whereby expert clinicians collect clinical data from patients and then interpret what these data mean. In this example, no mechanical data are collected, and the judgment is not made actuarially. A more common approach is a clinical interpretation of mixed data, whereby a clinician examines clinical data on the patient (some combination of exam and interview) and also performs some standardized "mechanical" tests, perhaps administering an MMPI. Then the clinician interprets this combination of data to make a judgment about the person's condition. Studies suggest that both of these approaches—the purely clinical one and the clinical interpretation of mixed data—are typically less reliable and valid than approaches using actuarial methods to interpret that data (Ægisdóttir et al., 2006). If several pieces of clinical and mechanical data are available, for example, actuarial combination of this data performs better than clinical interpretation (Dawes et al., 1989). In fact, actuarial combination of just clinical data typically performs better than clinical interpretation of all the data. In short, actuarial combination of clinical data, mechanical data, and especially of both clinical and mechanical data performs better than clinical interpretation of clinical data, mechanical data, or even both kinds of data.

Why Are Actuarial Methods Controversial?

It is difficult for many clinicians to believe that an inflexible rule ("3 points for chest pain, 2 points for family history and heart disease, 2 points for change in the ST segment of the ECG leads to…") would perform better than an experienced clinician who could take advantage of information not included in the actuarial formula. Indeed, some clinicians recoil at actuarial methods for being too impersonal; for treating patients like numbers and not like unique individuals. Others criticize actuarial methods for ignoring

useful information available to clinicians. A famous criticism of actuarial methods is known as the "broken leg problem." In one version of this, Professor A goes to the movies almost every Tuesday night. Knowing that today is Tuesday, an actuarial table might predict that the probability of Professor A going to the movie tonight is 0.9. However, you might know that Professor A just broke his leg and cannot get out of the house. You will have a much more accurate estimate of tonight's chance of him going to the movie than the actuarial approach (Salzinger, 2005).

The psychological power of this counterexample is that it makes it seem obvious that a clinician, given actuarial information, can always improve on actuarial judgment by using additional information not available to the actuarial formula. In practice, however, few cases are as clear-cut as the broken leg example. Most additional information will not dramatically change likelihood estimates derived from validated actuarial methods. In addition, even when additional relevant data are available, clinicians may not make proper use of the data. They may give the data too much or too little weight (Dawes, 1979).

In summary, clinicians are trained to collect clinical data from patients and to make decisions about which mechanical data will aid in diagnoses as well as to interpret these clinical and mechanical data. However, clinicians are generally not as good at interpreting those data as are established actuarial methods (Grove and Meehl, 1996; Grove et al., 2000; Meehl, 1954). There is evidence that the use of clinical judgment alone to assess whether an individual is exerting sufficient effort on performance-based tests or is providing an accurate self-report of symptoms is unreliable (Faust et al., 1988a,b; Heaton et al., 1978; Oldershaw and Bagby, 1997), making it important for the evaluator to collect and consider relevant mechanical data along with other objective data in making such assessments.

USE OF PSYCHOLOGICAL TESTS IN NON-SSA DISABILITY EVALUATIONS

To better understand the potential role of standardized psychological testing, including validity testing, for SSA disability determinations, the committee looked at current practices surrounding the use of psychological testing in several other settings that involve, or might involve, an element of secondary gain. The VA provides disability benefits to veterans who qualify based on injuries or disease incurred or aggravated during active military service or postservice disabilities that are related or secondary to disabilities occurring during service or are presumed to be related to circumstances of military service. The U.S. military assesses active duty military personnel for fitness for return to duty following injury. Private disability insurance programs determine whether claimants under their plans meet the criteria

to receive benefits. The automobile insurance industry determines claims of injury following auto accidents. Finally, the forensic setting (i.e., criminal and civil judicial contexts) includes litigation for personal injury and determinations of competency to stand trial. Common to all of these settings is assessment of an individual's alleged impairments to determine whether the individual qualifies for an outcome that may benefit him or her (e.g., disability benefit, restriction of military duty, compensation for injury, incompetence to stand trial). Despite this common element, the context of the settings—the purposes for which the assessments are being conducted—differ in important ways as discussed in the following sections.

Military and Veterans Affairs

Mental and behavioral health conditions have become more prevalent and consume a larger portion of the military and VA budget than they did 5 years ago. Within the past 10 years, the VA has reached consensus about the compensability of behavioral health conditions (e.g., posttraumatic stress disorder [PTSD]).

Significant progress has been made in defining mental and behavioral diagnoses. Both the military and the VA have measures of mental and behavioral health, and both evaluation systems address function as a key determinant for disability, although for somewhat different purposes, as described in the following sections.

Military[25]

There are significant differences between policies and procedures followed by SSA and the military. In contrast to disability evaluations for SSA and the Veterans Benefits Administration (VBA), discussed in the following section, military assessments for mental and behavioral health are performed to assess combat or duty readiness. Assessing whether an individual is capable of performing his or her duty may be an issue of safety not only for the individual but also for others.

Fitness for duty and return-to-duty determinations are made by medical evaluation boards and physical evaluation boards. Mental health providers serve as consultants to the boards, providing them with reports of diagnostic impressions, assessment of degree of impairment and impact on military duty performance, prognosis, and recommendations. In contrast to SSA and the VBA, evaluations in the military are often performed by therapists and care professionals who are not "interrogators" but are considered

[25] Much of the information in this section is drawn from the presentation to the committee by Robert Seegmiller (2014).

advocates and treating professionals, which may present a conflict with respect to treatment goals versus determinations of fitness for duty. It also should be noted that Army behavioral health professionals "diagnose and treat and should not be in an adversarial role with patients in terms of disability processes" and "must approach with a soldier-centered focus that provides soldiers the benefit of the doubt." Providers "on the whole do support the patient/soldier on face value and advocate in every way for them; however, [they] lose credibility with both medical personnel and line units when [they] fail to properly investigate and obtain collateral information" (U.S. Army Medical Command [MEDCOM] Behavioral Health Training Day, June 12, 2012, reported in Seegmiller, 2014).

Evaluations typically include review of medical records, consideration of premorbid functioning (the Armed Services Vocational Aptitude Battery), clinical interview and behavioral observations, and information from collateral sources. Psychological or neuropsychological testing is required in cases involving reported traumatic brain injury (TBI), but not always in cases involving PTSD. The selection of specific tests is left to the discretion of the clinician performing the evaluation, as is the use of PVTs and SVTs, although most providers, particularly psychologists and neuropsychologists, recognize the importance of their use.

A previous Office of the Surgeon General (OTSG)/MEDCOM policy memo on the optimal use of psychological and neuropsychological assessment, notes (1) "psychological and neuropsychological assessments are valuable tools in quantifying patient deficits, clarifying diagnoses, informing treatment, and in making decisions regarding a soldier's continued fitness for military service" and (2) "certain clinical tests in use by neuropsychology are designed to evaluate level of effort on the part of the test-taker. Poor effort on cognitive symptom validity measures means only that the data is not valid to be fully interpreted, and invalid data can be due to a range of causes other than malingering" (Policy Memo 11-076: Optimal Use of Psychological/Neuropsychological Assessment [21 Sept 2011–2013], reported in Seegmiller, 2014). "Poor effort on psychological/neuropsychological tests does not equate malingering, which requires proof of intent, per OTSG/MEDCOM Policy 11-076. In addition, this diagnosis requires the signatures of two credentialed care providers, including a supervisor, Department Chief, or Deputy Commander for Clinical Services" (OTSG/MEDCOM Policy Memo 12-035: Policy Guidance on the Assessment and Treatment of Post-Traumatic Stress Disorder [10 Apr 12 thru 10 Apr 14], reported in Seegmiller, 2014).

In his discussion with the committee, Dr. Robert Seegmiller (2014) asserted that SVTs and PVTs are critical tools that provide valuable information about the validity of an individual's test results. When making decisions and recommendations about whether soldiers are fit for duty or

whether they need disability, Seegmiller noted the importance of ensuring that one has good information in order to make the decision and recommendation that is the fairest for them and best for the system in terms of returning to work or not. However, such tests are only one type of tool: clinician's performing the evaluation also review the individual's medical records, conduct a clinical interview, make behavioral observations, gather collateral information, and the like, and consider the consistency of all of the information with what the patient is reporting.

Veterans Health Administration[26]

The VBA is responsible for administering and delivering an array of federally authorized benefits and services to eligible veterans and their dependents and survivors. In fiscal year 2012, 3,536,802 veterans received compensation benefits. PTSD was the third most prevalent service-connected disability among veterans receiving compensation at the end of fiscal year 2012, and TBI has been widely reported as the hallmark injury of the wars in Afghanistan and Iraq. To be eligible for disability compensation, a veteran must have served under conditions other than dishonorable, and the disability must not be the result of misconduct by the veteran. In contrast to the military setting, in which service members are assessed in terms of fitness for duty, veterans' assessments are performed with the recognition that there is responsibility to care for individuals who served in the military.

Disability compensation is paid monthly and varies according to the degree of disability and the number of dependents. The rate of compensation is graduated from 10 percent to 100 percent disabling, in increments of 10 percent, according to the combined degree of the veteran's disabilities. This differs from SSA, which determines an individual to be either disabled (100 percent) or not. Also unlike SSA, recipients of veterans' disability benefits may work with no limit on their earnings.

Disability examinations are conducted by full-time employees of the Veterans Health Administration (VHA), fee-basis staff, and contracted staff. Initial evaluations can be conducted by

> (1) board-certified psychiatrists; (2) psychiatrists who have successfully completed an accredited psychiatry residency and who are appropriately credentialed and privileged; (3) licensed doctoral-level psychologist[s]; (4) nonlicensed doctoral-level psychologists working toward licensure under close supervision by a board-certified, or board-eligible, psychiatrist or a licensed doctoral-level psychologist; (5) psychiatry residents under close

[26] Much of the information in this section is drawn from the presentation to the committee by Stacey Pollack (2014).

> supervision by a board-certified, or board-eligible, psychiatrist or a licensed doctoral-level psychologist; and (6) psychology interns or residents under close supervision by a board-certified, or board-eligible, psychiatrist or a licensed doctoral-level psychologist. (VHA Directive 2012-021, August 27, 2012)

Under the close supervision of a board-certified or board-eligible psychiatrist or licensed doctoral-level psychologist, reviews and increase evaluations can be conducted by licensed clinical social workers, nurse practitioners or clinical nurse specialists, and physician assistants (VHA Directive 2012-021, August 27, 2012).

The VHA requires all examiners to complete general online training regarding compensation and pension (C&P). Some specialty examiners are required to take additional training related to specific disabilities (e.g., PTSD).

> The objective of a C&P mental disorder examination is to obtain competent, critical, objective, and unbiased evaluations. To ensure that examination providers are competent to provide findings and opinions that are valid and sufficient for rating purposes, individuals who conduct C&P mental disorder examinations have specific qualifications and must have completed the required training. (VHA Directive 2012-021, August 27, 2012)

Examiners conducting C&P examinations for mental disorders are instructed to:

- Diagnose mental disorders, including personality disorders, using the nomenclature in the most current edition of the *Diagnostic and Statistical Manual of Mental Disorders*; …
- Determine when clinician-administered psychometric testing is necessary and integrate the results of such testing into the examination reports; …
- When necessary, comment on the significance of the veteran's prior mental health assessments (as reported) with respect to symptoms, occupational history, social history, and global assessment of functioning. (VHA Directive 2012-021, August 27, 2012)

For all initial PTSD disability evaluations, the examiner is instructed to review the veteran's claims file (C-file) or any other available medical records prior to conducting the examination. For an Integrated Disability Examination System (IDES) examination, the examiner is required to review the service member's medical records. Examiners are instructed to obtain results from all pertinent studies, evaluations, and tests, and order or perform any further studies, evaluations, or tests needed to diagnose a

mental disorder before completing their report. In addition, examiners must assess the individual for functional impairment. The examination report is used along with all other evidence to determine what level of compensation may be awarded to the veteran or service member.

VHA policy requires mental health examiners to review all records provided by VBA as part of a comprehensive evaluation. These records typically include the claimant's medical record. If there are psychological tests in the claimant's medical record, these should be reviewed as part of the evidence used in a comprehensive examination. The option to order additional psychological tests, including validity tests, is left to the discretion of the examiner. VA policy neither requires nor prohibits the ordering or use of any specific tests or categories of tests to evaluate any mental health condition.

Private Disability Insurance

Unum is the largest commercial disability insurer in the United States for both short-term and long-term disability. The committee looked to its processes to gain an understanding of how private disability insurers approach the use of psychological testing in adjudicating claims.[27] In evaluating a claim, examiners, who are clinicians, are required to consider all of the information in the claimant's file, including the results of previously administered psychological and neuropsychological tests. Examiners will attempt to acquire the raw test materials—the actual reports, the actual scores, the actual tests with the questions and answers—to analyze those data independently and determine whether they match the conclusions of the clinician who administered the tests. The examiners also are mandated to speak to the claimant's attending physicians.

If an independent medical examination (IME), an umbrella term that includes psychological, neuropsychological, or psychiatric examinations, is needed to provide additional information, the practitioner conducting the examination may administer any psychological tests or measures he or she feels are most valid based on current scientific literature and research. IME examiners are required to include peer-reviewed, scientifically validated measures of symptom and performance validity in their evaluations.[28] Validity of test results is addressed through a three-tiered system, formal

[27] The information in this section is drawn from the presentation to the committee by Thomas McLaren (2014).

[28] This is consistent with the findings of the SSA Office of the Inspector General, which reports on the practices of three private disability insurance providers, all of which allow the purchase and use the results of validity tests in their disability claims processes. All three companies also indicated that validity test results are just one piece of data they consider when evaluating claims (Office of the Inspector General, 2013). The names of the companies are not released in the report.

effort by stand-alone validity measures, consideration of imbedded validity measures, and an examination of the pattern of testing—meaning, whether it makes neurologic or medical sense for the condition being evaluated.

Although validity testing is required by Unum, the results of such testing are data points, which when taken in isolation can be misconstrued. For this reason, examiners are mandated to look at all of the information collectively. Invalid results on validity measures indicate that the remaining test results are not valid for clinical interpretation. In such cases, the IME or claims examiner would seek information from other sources.

After collecting and examining all the data relevant to the claim, the claims examiner balances the data to make a decision on the outcome and the claimant's restrictions and limitations—i.e., what the person is unable to do and what the person should not do.

Forensic Assessment: Criminal and Civil Judicial Contexts

At its most basic, the role of the legal system is to adjudicate disputes based on factual evidence. To achieve this goal, the courts rely on the collection of facts from a multitude of sources that are directly relevant to a specific legal question. One such source of information is the testimony of witnesses, who may provide the court with factual evidence based on personal knowledge of the matter but are prohibited from testifying based on their own opinions or analysis (*Federal Rules of Evidence*, Rule 602). However, under certain circumstances, the law does allow for the provision of opinions by an expert based on facts or data in the case (*Federal Rules of Evidence*, Rule 703). According to Rule 702 of the *Federal Rules of Evidence*:

> A witness who is qualified as an expert by knowledge, skill, experience, training, or education may testify in the form of an opinion or otherwise if:
>
> (1) the expert's scientific, technical, or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue;
>
> (2) the testimony is based on sufficient facts or data;
>
> (3) the testimony is the product of reliable principles and methods; and
>
> (4) the expert has reliably applied the principles and methods to the facts of the case.

With the requirement that the expert witness be able to provide information that is directly relevant to the question at hand, such witnesses can come from a variety of fields, including mental health. Once established as an expert witness, a mental health professional (i.e., psychologist, psychiatrist, or social worker) may provide expert opinion to assist in answering the legal question at hand.

Psychological assessments may be used in a variety of contexts and at all stages of the judicial process. For example, one of the primary uses for psychological assessments is to assess competency. During pretrial information gathering, this includes competencies such as whether a defendant was competent to consent to search and seizure or to confess, or to answer questions regarding mental state at the time of the offense. Similarly, psychological assessments may be used during the trial phase to answer questions related to competence to plead guilty, waive the right to counsel, testify, or refuse an insanity defense. Following a guilty verdict, psychological assessment may help answer questions related to competency to be sentenced or executed. In civil contexts, psychological assessments may be used to help answer questions related to civil commitment, compensation for mental injuries, or questions of competency, such as for guardianship, making treatment decisions, or consenting to research.

Psychological assessment for the courts is typically based on a variety of information sources and methods of data collection, including psychometric testing. Establishing symptom, performance, and response validity[29] is of particular importance in forensic contexts, as the potential for secondary gain may lead to examinee attempts to minimize, exaggerate, or feign problems (Bush et al., 2014). As noted in a statement from the Association for Scientific Advancement in Psychological Injury and Law, "Measures of performance and symptom validity are still in their relative infancy … [and] methodological difficulties exist in validity assessment research" (Bush et al., 2014; see also Chapters 4 and 5 of this report). For example, Bush and colleagues (2014) note there are few PVT manuals or articles that provide data on test-retest reliability on how reliably volunteers fake poor performance or simulate performance of actual examinees in simulation studies used to create cut-off scores. In addition, some comparison groups consist of mixed patient samples or populations that are dissimilar to an examinee and may not allow for appropriate comparisons. Finally, such tests do not necessarily speak to the intentionality behind invalid results, which may be generated consciously or unconsciously. Even in cases in which there is evidence of intentionally poor performance, the test results alone do not explain why the examinee did so (Bush et al., 2014).

Although the results of psychometric testing may play a crucial role in the formulation of a mental health professional's expert opinion for the courts, it is important to note that such tests are rarely used in isolation,

[29] The Association for Scientific Advancement in Psychological Injury and Law has identified a third type of validity important to forensic psychological assessment, termed response validity, as "the accuracy of the examinee's responses to autobiographical questions (e.g., educational history, vocational history, legal history) and questions pertaining to the legal matter in question (e.g., the nature of, and events surrounding, an injury, crime, or traumatic event)" (Bush et al., 2014, p. 199).

with most tests requiring some degree of subjective interpretation (Cohen and Malcolm, 2005). As with psychometric testing, evaluation of validity also should not rely on test scores alone, but rather, employ a multimethod approach (Bush et al., 2014). In addition to psychometric testing, forensic psychological assessment is typically based on a variety of other information sources, such as clinical interview, observational methods, and interviews with third parties.

International Community

Canada

The Canada Pension Plan (CPP) provides disability benefits to eligible individuals using much the same criteria in its disability determination process as SSA does (Government of Canada, 2014) As in the United States, there are a number of different settings in which disability determinations are made. Settings in addition to the CPP include the Worker Safety Insurance Board, Veterans Affairs Canada, and the auto insurance industry. Psychologists and neuropsychologists do not work under the Canadian national health care system. As a result, they work in a number of other settings, such as auto insurance.

Brian Levitt (2014) presented to the committee on the use of psychological testing under private auto insurance in the province of Ontario as well as tort law in Ontario. In this setting as well, the decision of whether to administer psychological tests and, if so, which particular test to use is determined by the individual psychologists according to the practice standards in that area of inquiry. The Canadian Academy of Psychologists and Disability Assessment standards related to psychological testing include the following:

- A psychologist shall employ standardized psychometric tests whenever possible;
- Psychologists whenever possible shall employ psychometric procedures that measure response bias and symptom validity; and
- Psychologists shall address any apparent discrepancies between the results of psychometric tests and other information.

These standards are consistent with the message that the use of validity tests is important, but they constitute only one piece of data, which must be interpreted in the context of all the other information.

Europe

Merten and colleagues (2013) have reported that large-scale research on and use of SVTs and PVTs in Europe followed that in the United States by about a decade, beginning in earnest in the early 2000s. As in the United States, the setting or context (forensic, clinical, etc.) seems to matter (Dandachi-FitzGerald et al., 2013; McCarter et al., 2009; Merten et al., 2013). It is important to note that in the study by Dandachi-FitzGerald and colleagues (2013) the definition of SVT was left to the respondent. Everything from discrepancies between records and observed behavior, to more "objective" scales on personality and effort tests was included, making it very difficult to interpret the findings regarding the percentage of medical professionals using SVTs when contracted to assess work capacity due to claims of psychological disability. There also appear to be differences in SVT and PVT use across European countries, with practitioners in the Netherlands and Norway reporting the greatest use of such tests (Merten et al., 2013).

Closing Comments

SSA, the U.S. military, the VBA, private disability insurance providers, and forensic assessment in civil and criminal judicial contexts have different goals, needs, and approaches to the evaluation and determination of disability (see Table 2-3). All share common elements, including identification of the presence of impairment and evaluation of its effect on the individual's ability to function.

Although the use of psychological testing must be understood in the context of each system's goals, each of the systems encourages a comprehensive evaluation, as determined by the evaluator, in an effort to answer these questions and each permits a broad range of evaluations. Whether to order psychological tests and the selection of which tests to administer are left to the discretion of the professional performing the evaluation or examination. With the exception of SSA, all of the systems permit, or in some cases require, the use of validity testing to provide information about the validity of the results of other psychological tests being administered. Nevertheless, all agree that although validity tests yield important information, the results of such tests are only one piece of data that needs to be assessed and interpreted in the context of all the other information available.

TABLE 2-3 Psychological Testing in Different Settings

Setting	Who Performs the Assessments	What Are the Assessments	Psychological Tests Employed	Policy on Psychological or Neuropsychological Tests	Concerns/Conflicts
SSA	DDS disability examiners Consultative examiner psychologists	Medical record review Clinical interview Behavioral observations	Primarily intelligence tests Other standardized tests as determined by consultative examiner and paid for by state DDS agencies	Intelligence tests for intellectual disability claims Other tests at discretion of DDS and consultative examiner Disallows purchase of SVTs/PVTs	
VA	Psychiatrist Psychologist Under supervision: Residents NPs PAs Social workers	Clinical files IDES Lab studies/tests Functional evaluations Quality of life assessment	Any relevant, scientifically valid tests (as determined by evaluator)	None specifically required or prohibited SVTs/PVTs are neither required nor prohibited	Diagnostic listings are limited Inconsistency in the use of tests; not all VA medical centers use the same measures
Military	Medical Evaluation Boards Physical Evaluation Boards Consultants (provide reports to above boards) Psychologists Neuropsychologists Psychiatrists	Determination of degree of impairment Assessment of impact on duty assignment Review of all medical records Clinical interview with observation	Neuropsychological testing SVTs/PVTs used to validate data	Required for TBI Not required for PTSD PVTs/SVTs recommended when possibility of secondary gain Testing at providers' discretion	Sometimes evaluators are the treating physicians Each provider can select No uniformity/consistency Culture supports view that do not wish to offend those who sacrificed; hence, may not test or validate Malingering charge may lead to lengthy legal battle

continued

TABLE 2-3 Continued

Setting	Who Performs the Assessments	What Are the Assessments	Psychological Tests Employed	Policy on Psychological or Neuropsychological Tests	Concerns/Conflicts
Private	Disability evaluators: Neuropsychologists Psychologists Psychiatrists Social workers	Clinical files or records[a]	Any relevant, scientifically valid tests	Evaluator determines necessary testing PVTs/SVTs required	Industry has additional resources Each company makes its own policy
Forensic: Civil and Criminal	Mental health professionals hired by defense or prosecution: Psychologists Psychiatrists Social workers				Hired by defense or prosecution to support position favorable to that side

NOTE: DDS = Disability Determination Services; IDES = Independent Disability Examination System; NP = nurse practitioner; PA = physician assistant; PTSD = posttraumatic stress disorder; PVT = performance validity test; SVT = symptom validity test; TBI = traumatic brain injury.

[a] Some require standard tests, such as the AMA Guide (see, for example, Rondinelli, 2008).

FINDINGS

- There currently is great variability in allowance rates for both SSI and SSDI among states that is not fully accounted for by differences in the populations of applicants. There also is great variability in the disability determination appeal rulings among ALJs within and across states.
- Each state DDS agency, within the confines of SSA policy, issues its own rules regarding the tests that may be purchased as part of a CE. For this reason, there is variation among states about when and which standardized psychological tests can be purchased, with the exception of PVTs and SVTs, which are precluded from purchase by SSA.
- There currently are no data on the rates of false positives and false negatives in SSA disability determinations.
- Identification and documentation of the presence and severity of medically determinable mental impairments at Step 2 of SSA's disability determination process could be informed by results of standardized psychological tests.
- Identification and assessment of the severity of work-related functional impairment relevant to disability evaluations at the listing level (Step 3) and to mental residual functional capacity (Steps 4 and 5) are other points in SSA's disability determination process that could be informed by results of standardized psychological tests.
- Consultative examinations may be ordered by DDS examiners or ALJs to supplement evidence in a claimant's case record. Psychological tests could be administered as part of a CE.
- In some cases, SSA disability examiners must evaluate the credibility of statements by individuals about the intensity and persistence of their symptoms and the effect on the individual's ability to function and perform work-related activities.
- Current data on the prevalence of inconsistent reporting of symptoms or performing below one's capability on cognitive tests among SSDI and SSI applicant populations are limited.
- Current SSA policy precludes the purchase of (validity) tests—e.g., MMPI-2 and TOMM—to help inform determinations about the credibility of an individual's statements or about possible malingering.
- There is inconsistency among SSA's statements on validity testing:
  - Results can "provide evidence suggestive of poor effort or intentional symptom manipulation."

  - "Malingering cannot be proven with tests"; "malingering is one aspect of the larger sphere of inaccurate self-reporting."
  - "No test … conclusively determines the presence of inaccurate patient self-report."
  - "Even a high likelihood of malingering does not preclude severe limitations resulting from a genuine medically determinable impairment."
- Clinicians generally are not as good at interpreting clinical and mechanical data as are established actuarial methods.
- Each of the systems reviewed leave the question of whether to order psychological tests and the selection of which tests to administer to the discretion of the professional performing the evaluation or examination. With the exception of SSA, all of the systems permit, or in some cases require, the use of validity testing to provide information about the validity of the results of other psychological tests being administered. Nevertheless, all agree that although validity tests yield important information, the results of such tests are only one piece of data that needs to be assessed and interpreted in the context of all the other information available.

REFERENCES

Ægisdóttir, S., M. J. White, P. M. Spengler, A. S. Maugherman, L. A. Anderson, R. S. Cook, C. N. Nichols, G. K. Lampropoulos, B. S. Walker, G. Cohen, and J. D. Rush. 2006. The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction. *The Counseling Psychologist* 4(3):341-382.

APA (American Psychological Association). 2015. *Guidelines and principles for accreditation of programs in professional psychology: Quick reference guide to doctoral programs*. http://www.apa.org/ed/accreditation/about/policies/doctoral.aspx (accessed January 20, 2015).

Autor, D. H., and M. G. Duggan. 2003. The rise in the disability rolls and the decline in unemployment. *Quarterly Journal of Economics* 118(1):157-205.

Black, D., K. Daniel, and S. Sanders. 2002. The impact of economic conditions on participation in disability programs: Evidence from the coal boom and bust. *American Economic Review* 92(1):27-50.

Burkhauser, R., J. S. Butler, and R. Weathers II. 2002. How policy variables influence the timing of applications for Social Security Disability Insurance. *Social Security Bulletin* 64(1):52-83.

Bush, S. S., R. M. Ruff, A. I. Tröster, J. T. Barth, S. P. Koffler, N. H. Pliskin, C. R. Reynolds, and C. H. Silver. 2005. Symptom validity assessment: Practice issues and medical necessity. NAN Policy & Planning Committee. *Archives of Clinical Neuropsychology* 20(4):419-426.

Bush, S. S., R. L. Heilbronner, and R. M. Ruff. 2014. Psychological assessment of symptom and performance validity, response bias, and malingering: Official position of the Association for Scientific Advancement in Psychological Injury and Law. *Psychological Injury and Law* 7(3):197-205.

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., and J. Abrahams. 2005. Green's MACT helps identify internal predictors of effort in the social security disability exam. [Abstract.] *Archives of Clinical Neuropsychology* 20(7):889-890.

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

Cohen, A., and C. Malcolm. 2005. Psychological assessment for the courts. In *Psychology and law*, edited by C. Tredoux, D. Foster, A. Allan, A. Cohen, and D. Wassenaar. Lansdowne, South Africa: Juta and Co. Ltd.

Dandachi-FitzGerald, B., R. W. Ponds, and T. Merten. 2013. Symptom validity and neuropsychological assessment: A survey of practices and beliefs of neuropsychologists in six European countries. *Archives of Clinicial Neuropsychology* 28(8):771-783.

Dawes, R. 1979. The robust beauty of improper linear models in decision making. *American Psychologist* 34(7):571-582.

Dawes, R., D. Faust, and P. Meehl. 1989. Clinical versus actuarial judgments. *Psychological Science* 243(4899):1688-1674.

Duggan, M., and S. Imberman. 2008. Why are disability rolls skyrocketing? The contribution of population characteristics, economic conditions, and program generosity. In *Health at older ages: The causes and consequeses of declining disability among the elderly*, edited by D. Cutler and D. Wise. Chicago, IL: University of Chicago Press.

Faust, D., K. Hart, and T. Guilmette. 1988a. Pediatric malingering: The capacity of children to fake believable deficits on neuropsychological testing. *Journal of Consulting and Clinical Psychology* 56(4):578-582.

Faust, D., K. Hart, T. Guilmette, and H. Arkes. 1988b. Neuropsychologists' capacity to detect adolescent malingerers. *Professional Psychology: Research and Practice* 19(5):508-515.

Government of Canada. 2014. *How applications for disability benefits are assessed*. http://www.servicecanada.gc.ca/eng/services/pensions/cpp/disability/benefit/assessment.shtml (accessed January 4, 2015).

Griffin, G. A., J. Normington, R. May, and D. Glassmire. 1996. Assessing dissimulation among Social Security disability income claimants. *Journal of Consulting and Clinical Psychology* 64(6):1425-1430.

Grove, W. M., and P. E. Meehl. 1996. Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychology, Public Policy, and Law* 2(2):293-323.

Grove, W. M., D. H. Zald, B. S. Lebow, B. E. Snitz, and C. Nelson. 2000. Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment* 12(1):19-30.

Heaton, R. K., H. H. Smith, R. A. Lehman, and A. T. Vogt. 1978. Prospects for faking believable deficits on neuropsychological testing. *Journal of Consulting and Clinical Psychology* 46(5):892-900.

Heilbronner, R. L., J. J. Sweet, J. E. Morgan, G. J. Larrabee, S. R. Millis, and 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.

Heiser, N. 2014. *Disability Determination Services panel discussion with the committee*. Presentation to the IOM Committee on Psychological Testing, Including Validity Testing, for Social Security Administration Disability Determinations, August 11, 2014, Washington, DC.

Juslin, P. 1994. The overconfidence phenomenon as a consequence of informal experimenter-guided selection of almanac items. *Organizational Behavior and Human Decision Processes* 57(2):226-246.

Kahneman, D. 2011. *Thinking fast and slow*. New York: Farrar, Straus, and Giroux.

Kreider, B. 1999. Latent work disability and reporting bias. *Journal of Human Resources* 34(4):734-769.

Larrabee, G. J. 2007. Introduction: Malingering, research designs, and base rates. In *Assessment of malingered neuropsychological deficits,* edited by G. J. Larrabee. New York: Oxford University Press.

Laurence, B. 2015. *Third level of appeal for disability: Appeals council & remands*. http://www.disabilitysecrets.com/appeals-council.html (accessed October 20, 2014).

Levitt, B. 2014. Psychological disability evaluations under the Ontario auto insurance system and Ontario tort law. Presentation to the IOM Committee on Psychological Testing, Including Validity Testing, for Social Security Administration Disability Determinations, June 25, 2014, Washington, DC.

McCarter, R. J., N. H. Walton, D. N. Brooks, and G. E. Powell. 2009. Effort testing in contemporary U.K. neuropsychological practice. *The Clinical Neuropsychologist* 23(6):1050-1066.

McLaren, T. 2014. *Use of performance and symptom validity assessment within the independent disability insurer context*. Presentation to the IOM Committee on Psychological Testing, Including Validity Testing, for Social Security Administration Disability Determinations, June 25, 2014, Washington, DC.

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

Merten, T., B. Dandachi-FitzGerald, V. Hall, B. A. Schmand, P. Santamaría, and H. González-Ordi. 2013. Symptom validity assessment in European countries: Development and state of the art. *Clínica y Salud* 24(3):129-138.

Miller, L. S., M. C. Boyd, and A. Cohn. 2006. Prevalence of sub-optimal effort in disability applicants. [Abstract.] *Journal of the International Neuropsychological Society* 12(S1):159.

Mittenberg, W., C. Patton, E. M. Canyock, and D. C. Condit. 2002. Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology* 24(8):1094-1102.

Moore, D., and P. Healy. 2008. The trouble with overconfidence. *American Psychological Association* 115(2):502-517.

Morton, D. 2014. *Social Security disability: Four levels of appeal*. http://www.nolo.com/legal-encyclopedia/social-security-disability-appeal-levels-32398.html (accessed March 27, 2015).

Office of the Inspector General, SSA (Social Security Administration). 2013. *The Social Security Administration's policy on symptom validity tests in determining disability claims*. Washington, DC: SSA. http://oig.ssa.gov/sites/default/files/audit/full/pdf/A-08-13-23094.pdf (accessed March 27, 2015).

Oldershaw, L., and M. Bagby. 1997. *Children and deception*. New York: Guildford.

Oskamp, S. 1965. Overconfidence in case-study judgments. *Journal of Consulting Psychology* 29(3):261-265.

Pollack, S. 2014. *VA policies and/or practices surrounding the use of psychological tests and symptom validity tests in the disability determination process*. Presentation to the IOM Committee on Psychological Testing, Including Validity Testing, for Social Security Administration Disability Determinations, June 25, 2014, Washington, DC.

Price, J. H. 2014. *Disability Determination Services panel discussion with the committee*. Presentation to the IOM Committee on Psychological Testing, Including Validity Testing, for Social Security Administration Disability Determinations, August 11, 2014, Washington, DC.

Rondinelli, R. D., ed. 2008. *AMA guides to the evaluation of permanent impairment*, sixth edition. Chicago, IL: American Medical Association.

Rupp. 2012. Factors affecting initial disability allowance rates for the Disability Insurance and Supplemental Security Income programs: The role of the demographic and diagnostic composition of applicants and local labor market conditions. *Social Security Bulletin* 72(4):11-35. http://ssrn.com/abstract=2172488 (accessed February 4, 2015).

Rupp, K., and D. Stapleton. 1995. Determinants of the growth in the Social Security Administration's disability programs—An overview. *Social Security Bulletin* 58(4):43-70.

Salzinger, K. 2005. Clinical, statistical, and broken-leg predictions. *Behavior and Philosophy* 33:91-99.

Samuel, R. Z., and W. Mittenberg. 2005. Determination of malingering in disability evaluations. *Primary Psychiatry* 12(12):60-68.

Scheinkman, J. A., and W. Xiong. 2003. Overconfidence and speculative bubbles. *Journal of Political Economy* 111(6):1183-1220.

Seegmiller, R. 2014. *Use of psychological tests, including PVTs and SVTs, in select populations: The U.S. military*. Presentation to the IOM Committee on Psychological Testing, Including Validity Testing, for Social Security Administration Disability Determinations, June 25 2014, Washington, DC.

Soss, J., and L. R. Keiser. 2006. The political roots of disability claims: How state environments and policies shape citizen demands. *Political Research Quarterly* 59(1):133-148.

SSA (Social Security Administration). 1996a. *SSR 96-3p: Policy interpretation ruling. Titles II and XVI: Considering allegations of pain and other symptoms in determining whether a medically determinable impairment is severe*. http://www.socialsecurity.gov/OP_Home/rulings/di/01/SSR96-03-di-01.html (accessed August 20, 2014).

SSA. 1996b. *SSR 96-4p: Policy interpretation ruling. Titles II and XVI: Symptoms, medically determinable physical and mental impairments, and exertional and nonexertional limitations*. http://www.socialsecurity.gov/OP_Home/rulings/di/01/SSR96-04-di-01.html (accessed October 3, 2014).

SSA. 1996c. *SSR 96-7p: Policy interpretation ruling Titles II and XVI: Evaluation of symptoms in disability claims: Assessing the credibility of an individual's statements*. http://www.socialsecurity.gov/OP_Home/rulings/di/01/SSR96-07-di-01.html (accessed October 3, 2014).

SSA. 2008. *National Q&A, 08-003 rev 2, do tests of malingering have any value for SSA evaluations?* Washington, DC: SSA.

SSA. 2009. DI 22511.005 Documenting the impact of a medically determinable mental impairment on an individual's ability to work. *Program Operations Manual System (POMS)*. https://secure.ssa.gov/apps10/poms.nsf/lnx/0422511005 (accessed January 30, 2015).

SSA. 2010. Revised medical criteria for evaluating mental disorders. *Federal Register* 75(160):51336-51368.

SSA. 2012a. DI 00115.001 Social Security Administration's (SSA's) disability programs. *Program Operations Manual System (POMS)*. https://secure.ssa.gov/poms.nsf/lnx/0400115001 (accessed August 20, 2014).

SSA. 2012b. DI 22501.001 Disability case development for medical and other evidence. *Program Operations Manual System (POMS)*. https://secure.ssa.gov/poms.nsf/lnx/0422501001 (accessed October 3, 2014).

SSA. 2012c. DI 22510.048 Pediatric consultative examination (CE) report content guidelines—Mental disorders. *Program Operations Manual System (POMS)*. https://secure.ssa.gov/poms.nsf/lnx/0422510048 (accessed October 3, 2014).
SSA. 2012d. DI 22511.007 Sources of evidence. *Program Operations Manual System (POMS)*. https://secure.ssa.gov/apps10/poms.nsf/lnx/0422511007 (accessed December 30, 2014).
SSA. 2012e. *Disability Determination Services adminsistrative letter no. 866: Consulative examinations malingering & credibility tests—Information*. Washington, DC: SSA.
SSA. 2012f. *Social Security testimony before Congress*. Statement of Michael I. Astrue, Commisioner, Social Security Administration before the Committee on Ways and Means Subcommittee on Social Security, June 27, 2012. http://www.ssa.gov/legislation/testimony_062712.html (accessed October 20, 2014).
SSA. 2013. DI 22510.006 When not to purchase a consultative examination (CE). *Program Operations Manual System (POMS)*. https://secure.ssa.gov/poms.nsf/lnx/0422510006 (accessed October 3, 2014).
SSA. 2014a. *Annual report of the Supplemental Security Income program*. Baltimore, MD: SSA.
SSA. 2014b. *Annual statistical report on the Social Security Disability Insurance program, 2013*. Washington, DC. SSA. http://www.ssa.gov/policy/docs/statcomps/di_asr (accessed February 24, 2015).
SSA. 2014c. *DDS performance management report. Disability claims data. Consultative examination rates, fiscal year 2013*. Data prepared by ORDP, ODP, and ODPMI. Data submitted to the IOM Committee on Psychological Testing, Including Validity Testing, for Social Security Administration Disability Determinations by Joanna Firmin, Social Security Administration, on October 8, 2014.
SSA. 2014d. *Disability claims data (initial, reconsideration, continuing disability review) by adjudictive level and body system. SSDI, SSI, Concurrent, and Total Claims*. Data prepared by ORDP, ODP, and ODPMI. Submitted to the IOM Committee on Psychological Testing, Including Validity Testing, for Social Security Administration Disability Determinations by Joanna Firmin, Social Security Administration, on October 8, 2014.
SSA. 2014e. DI 22510.021 Consultative examination (CE) report content guidelines: Mental disorders. *Program Operations Manual System (POMS)*. https://secure.ssa.gov/poms.nsf/lnx/0422510021 (accessed October 3, 2014).
SSA. 2014f. DI 24515.008 Titles II and XVI: Considering opinions and other evidence from sources who are not "acceptable medical sources" in disability claims; considering decisions on disability by other governmental and nongovernmental agencies (SSR 06-03p). *Program Operations Manual System (POMS)*. https://secure.ssa.gov/poms.nsf/lnx/0424515008 (accessed February 24, 2015).
SSA. 2014g. DI 24515.075 Evaluating claims involving Chronic Fatigue Syndrome (CFS). *Program Operations Manual System (POMS)*. https://secure.ssa.gov/poms.nsf/lnx/0424515075 (accessed December 16, 2014).
SSA. 2014h. *National data: Title II—SSDI, Title XVI—SSI, & concurrent Title II/XVI initial disability determinations. By regulation basis code for adults and children (reason for decision), fiscal year 2013*. Data submitted to the IOM Committee on Psychological Testing, Including Validity Testing, for Social Security Administration Disability Determinations by Joanna Firmin, Social Security Administration, on October 23, 2014.
SSA. 2014i. *Open government initiative. Data on combined Title II disability and Title XVI blind/disabled average processing time (in days) (excludes technical denials)*. http://www.ssa.gov/open/data/Combined-Disability-Processing-Time.html (accessed December 16, 2014).

SSA. 2014j. *SSDI awards by diagnostic group and age of awardee under the age of 65, 2013 (preliminary data)*. Data submitted to the IOM Committee on Psychological Testing, Including Validity Testing, for Social Security Administration Disability Determinations by Joanna Firmin, Social Security Administration, on October 21, 2014.

SSA. 2014k. *SSI annual statistical report, 2013*. Washington, DC: SSA. http://www.ssa.gov/policy/docs/statcomps/ssi_asr (accessed February 24, 2015).

SSA. 2014l. *SSI awards by diagnostic group and age of awardee under the age of 65, 2013*. Data prepared by ORDP, ODP, and ODPMI. Data submitted to the IOM Committee on Psychological Testing, Including Validity Testing, for Social Security Administration Disability Determinations by Joanna Firmin, Social Security Administration, on October 21, 2014.

SSA. 2014m. *Substantial gainful activity*. http://www.socialsecurity.gov/oact/cola/sga.html (accessed December 15, 2014).

SSA. 2015. *Hearings and appeals. ALJ disposition data, fiscal year 2015 (for reporting purposes: 09/2/2014 through 01/20/2015)*. http://www.ssa.gov/appeals/DataSets/03_ALJ_Disposition_Data.html (accessed February 27, 2015).

SSA. n.d.-a. *Disability evaluation under Social Security—Part II: Evidentiary requirements*. http://www.ssa.gov/disability/professionals/bluebook/evidentiary.htm (accessed September 4, 2014).

SSA. n.d.-b. *Disability evaluation under Social Security—Part III: Listing of impairments*. http://www.ssa.gov/disability/professionals/bluebook/listing-impairments.htm (accessed October 3, 2014).

SSA. n.d.-c. *Disability evaluation under Social Security—Part III: Listing of impairments—Adult listings (Part A)*. http://www.ssa.gov/disability/professionals/bluebook/12.00-MentalDisorders-Adult.htm (accesseed October 3, 2014).

SSA. n.d.-d. *Disability evaluation under Social Security—Part III Listing of impairments—Childhood listings (Part B)*. http://www.ssa.gov/disability/professionals/bluebook/ChildhoodListings.htm (accessed October 7, 2014).

SSA. n.d.-e. *Disability evaluation under Social Security—Part III: Listing of impairments—Adult listings (Part A)—section 12.00 mental disorders*. http://www.ssa.gov/disability/professionals/bluebook/12.00-MentalDisorders-Adult.htm (accessed November 14, 2014).

SSA. n.d.-f. *Disability evaluation under Social Security—Part III: Listing of impairments—Childhood listings (Part B)—section 112.00 mental disorders*. http://www.ssa.gov/disability/professionals/bluebook/112.00-MentalDisorders-Childhood.htm (accessed October 3, 2014).

SSA. n.d.-g. *Hearings and appeals. Federal court review process*. http://www.socialsecurity.gov/appeals/court_process.html#a0=1 (accessed October 7, 2014).

SSA. n.d.-h. *Hearings and appeals. Information about requesting review of an administrative law judge's hearing decision*. http://www.socialsecurity.gov/appeals/appeals_process.html#a0=2 (accessed October 7, 2014).

SSA. n.d.-i. *Hearings and appeals. What you need to know to request a hearing before an administrative law judge*. http://www.socialsecurity.gov/appeals/hearing_process.html#a0=4&sb=3 (accessed October 7, 2014).

SSA. n.d.-j. *How we decide if you are disabled. Information we need about your work and education*. http://www.ssa.gov/disability/step4and5.htm (accessed October 7, 2014).

SSA. n.d.-k. *Medical/professional relations. Consultative examinations: A guide for health professionals*. http://www.ssa.gov/disability/professionals/greenbook (accessed October 16, 2014).

SSA. n.d.-l. *Occupational Information System project*. http://www.ssa.gov/disabilityresearch/occupational_info_systems.html (accessed December 30, 2014).

SSA. n.d.-m. *Selected data from Social Security's disability program*. http://www.ssa.gov/oact/STATS/dibStat.html (accessed January 27, 2015).

SSDRC (Social Security Disability and SSI [Social Security Insurance] Resource Center). n.d. *Applying for disability: How long does it take to get Soical Security Disability or SSI benefits?* http://www.ssdrc.com/disabilityquestions1-46.html (accessed December 15, 2014).

Strand, A. 2002. *Social Security disability programs: Assessing the variation in allowance rates. ORES working paper series, no. 98*. Washington, DC: Social Security Administration, Division of Policy Evaluation, Office of Research, Evaluation, and Statistics. http://socialsecurity.gov/policy/docs/workingpapers/wp98.pdf (accessed February 4, 2015).

Ward, T. A. 2014. *Disability Determination Services panel discussion*. Presentation to the IOM Committee on Psychological Testing, Including Validity Testing, for Social Security Administration Disability Determinations, August 11, 2014, Washington, DC.

Wedding, D., and D. Faust. 1989. Clinical judgement and decision making in neuropsychology. *Archives of Clinical Neuropsychology* 4(3):233-256.

3

Overview of Psychological Testing

Psychological assessment contributes important information to the understanding of individual characteristics and capabilities, through the collection, integration, and interpretation of information about an individual (Groth-Marnat, 2009; Weiner, 2003). Such information is obtained through a variety of methods and measures, with relevant sources determined by the specific purposes of the evaluation. Sources of information may include

- Records (e.g., medical, educational, occupational, legal) obtained from the referral source;
- Records obtained from other organizations and agencies that have been identified as potentially relevant;
- Interviews conducted with the person being examined;
- Behavioral observations;
- Interviews with corroborative sources such as family members, friends, teachers, and others; and
- Formal psychological or neuropsychological testing.

Agreements across multiple measures and sources, as well as discrepant information, enable the creation of a more comprehensive understanding of the individual being assessed, ultimately leading to more accurate and appropriate clinical conclusions (e.g., diagnosis, recommendations for treatment planning).

The clinical interview remains the foundation of many psychological and neuropsychological assessments. Interviewing may be structured,

semistructured, or open in nature, but the goal of the interview remains consistent—to identify the nature of the client's presenting issues, to obtain direct historical information from the examinee regarding such concerns, and to explore historical variables that may be related to the complaints being presented. In addition, the interview element of the assessment process allows for behavioral observations that may be useful in describing the client, as well as discerning the convergence with known diagnoses. Based on the information and observations gained in the interview, assessment instruments may be selected, corroborative informants identified, and other historical records recognized that may aid the clinician in reaching a diagnosis. Conceptually, clinical interviewing explores the presenting complaint(s) (i.e., referral question), informs the understanding of the case history, aids in the development of hypotheses to be examined in the assessment process, and assists in determination of methods to address the hypotheses through formal testing.

An important piece of the assessment process and the focus of this report, psychological testing consists of the administration of one or more standardized procedures under particular environmental conditions (e.g., quiet, good lighting) in order to obtain a representative sample of behavior. Such formal psychological testing may involve the administration of standardized interviews, questionnaires, surveys, and/or tests, selected with regard to the specific examinee and his or her circumstances, that offer information to respond to an assessment question. Assessments, then, serve to respond to questions through the use of tests and other procedures. It is important to note that the selection of appropriate tests requires an understanding of the specific circumstances of the individual being assessed, falling under the purview of clinical judgment. For this reason, the committee refrains from recommending the use of any specific test in this report. Any reference to a specific test is to provide an illustrative example, and should not be interpreted as an endorsement by the committee for use in any specific situation; such a determination is best left to a qualified assessor familiar with the specific circumstances surrounding the assessment.

To respond to questions regarding the use of psychological tests for the assessment of the presence and severity of disability due to mental disorders, this chapter provides an introductory review of psychological testing. The chapter is divided into three sections: (1) types of psychological tests, (2) psychometric properties of tests, and (3) test user qualifications and administration of tests. Where possible an effort has been made to address the context of disability determination; however, the chapter is primarily an introduction to psychological testing.

TYPES OF PSYCHOLOGICAL TESTS

There are many facets to the categorization of psychological tests, and even more if one includes educationally oriented tests; indeed, it is often difficult to differentiate many kinds of tests as purely psychological tests as opposed to educational tests. The ensuing discussion lays out some of the distinctions among such tests; however, it is important to note that there is no one correct cataloging of the types of tests because the different categorizations often overlap. Psychological tests can be categorized by the very nature of the behavior they assess (what they measure), their administration, their scoring, and how they are used. Figure 3-1 illustrates the types of psychological measures as described in this report.

The Nature of Psychological Measures

One of the most common distinctions made among tests relates to whether they are measures of *typical behavior* (often non-cognitive measures) versus tests of *maximal performance* (often cognitive tests) (Cronbach, 1949, 1960). A measure of typical behavior asks those completing the instrument to describe what they would commonly do in a given situation. Measures of typical behavior, such as personality, interests, values, and attitudes, may be referred to as *non-cognitive measures*. A test of maximal performance, obviously enough, asks people to answer questions and solve problems as well as they possibly can. Because tests of maximal performance typically involve cognitive performance, they are often referred to as *cognitive tests*. Most intelligence and other ability tests would be considered cognitive tests; they can also be known as ability tests, but this would be a more limited category. Non-cognitive measures rarely have correct answers per se, although in some cases (e.g., employment tests) there may be preferred responses; cognitive tests almost always have items that have correct answers. It is through these two lenses—non-cognitive measures and cognitive tests—that the committee examines psychological testing for the purpose of disability evaluation in this report.

One distinction among non-cognitive measures is whether the stimuli composing the measure are *structured* or *unstructured*. A structured personality measure, for example, may ask people true-or-false questions about whether they engage in various activities or not. Those are highly structured questions. On the other hand, in administering some commonly used personality measures, the examiner provides an unstructured projective stimulus such as an inkblot or a picture. The test-taker is requested to describe what they see or imagine the inkblot or picture to be describing. The premise of these projective measures is that when presented with ambiguous

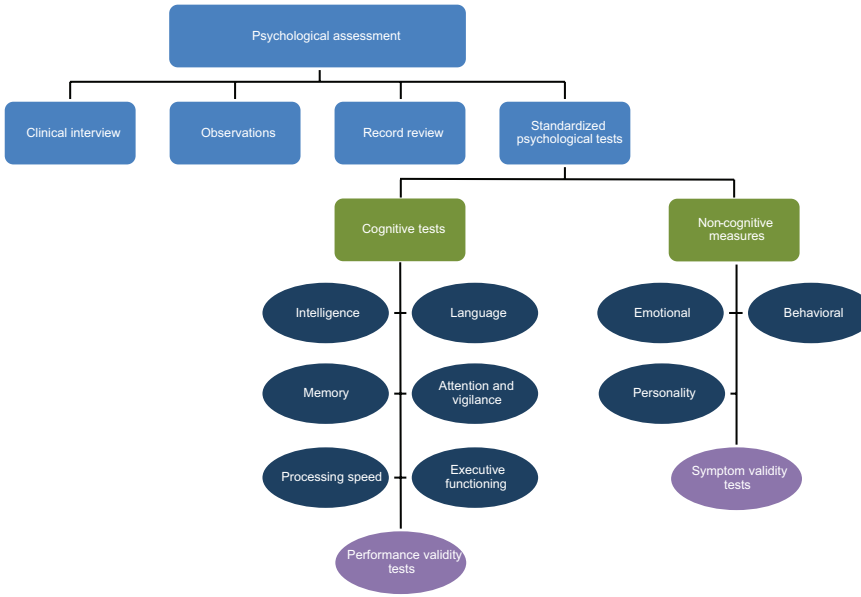


FIGURE 3-1 Components of psychological assessment.
NOTE: Performance validity tests do not measure cognition, but are used in conjunction with performance-based cognitive tests to examine whether the examinee is exerting sufficient effort to perform well and responding to the best of his or her capability. Similarly, symptom validity tests do not measure non-cognitive status, but are used to examine whether a person is providing an accurate report of his or her actual symptom experience. Because cognitive tests frequently are performance based and non-cognitive measures generally involve self-report, performance validity tests and symptom validity tests are shown as being associated with these types of tests.

stimuli an individual will project his or her underlying and unconscious motivations and attitudes. The scoring of these latter measures is often more complex than it is for structured measures.

There is great variety in cognitive tests and what they measure, thus requiring a lengthier explanation. Cognitive tests are often separated into tests of *ability* and tests of *achievement*; however, this distinction is not as clear-cut as some would portray it. Both types of tests involve learning. Both kinds of tests involve what the test-taker has learned and can do. However, achievement tests typically involve learning from very specialized education and training experiences; whereas, most ability tests assess learning that has occurred in one's environment. Some aspects of learning

are clearly both; for example, vocabulary is learned at home, in one's social environment, and in school. Notably, the best predictor of intelligence test performance is one's vocabulary, which is why it is often given as the first test during intelligence testing or in some cases represents the body of the intelligence test (e.g., the Peabody Picture Vocabulary Test). Conversely, one can also have a vocabulary test based on words one learns only in an academic setting. Intelligence tests are so prevalent in many clinical psychology and neuropsychology situations that we also consider them as neuropsychological measures. Some abilities are measured using subtests from intelligence tests; for example, certain working memory tests would be a common example of an intelligence subtest that is used singly as well. There are also standalone tests of many kinds of specialized abilities.

Some ability tests are broken into *verbal* and *performance* tests. Verbal tests, obviously enough, use language to ask questions and demonstrate answers. Performance tests on the other hand minimize the use of language; they can involve solving problems that do not involve language. They may involve manipulating objects, tracing mazes, placing pictures in the proper order, and finishing patterns, for example. This distinction is most commonly used in the case of intelligence tests, but can be used in other ability tests as well. Performance tests are also sometimes used when the test-taker lacks competence in the language of the testing. Many of these tests assess visual spatial tasks. Historically, nonverbal measures were given as intelligence tests for non-English speaking soldiers in the United States as early as World War I. These tests continue to be used in educational and clinical settings given their reduced language component.

Different cognitive tests are also considered to be *speeded tests* versus *power tests*. A truly speeded test is one that everyone could get every question correct if they had enough time. Some tests of clerical skills are exactly like this; they may have two lists of paired numbers, for example, where some pairings contain two identical numbers and other pairings are different. The test-taker simply circles the pairings that are identical. Pure power tests are measures in which the only factor influencing performance is how much the test-taker knows or can do. A true power test is one where all test-takers have enough time to do their best; the only question is what they can do. Obviously, few tests are either purely speeded or purely power tests. Most have some combination of both. For example, a testing company may use a rule of thumb that 90 percent of test-takers should complete 90 percent of the questions; however, it should also be clear that the purpose of the testing affects rules of thumb such as this. Few teachers would wish to have many students unable to complete the tests that they take in classes, for example. When test-takers have disabilities that affect their ability to respond to questions quickly, some measures provide extra time, depending upon their purpose and the nature of the characteristics being assessed.

Questions on both achievement and ability tests can involve either *recognition* or *free-response* in answering. In educational and intelligence tests, recognition tests typically include multiple-choice questions where one can look for the correct answer among the options, recognize it as correct, and select it as the correct answer. A free-response is analogous to a "fill-in-the-blanks" or an essay question. One must recall or solve the question without choosing from among alternative responses. This distinction also holds for some non-cognitive tests, but the latter distinction is discussed later in this section because it focuses not on recognition but selections. For example, a recognition question on a non-cognitive test might ask someone whether they would rather go ice skating or to a movie; a free recall question would ask the respondent what they like to do for enjoyment.

Cognitive tests of various types can be considered as *process* or *product* tests. Take, for example, mathematics tests in school. In some instances, only getting the correct answer leads to a correct response. In other cases, teachers may give partial credit when a student performs the proper operations but does not get the correct answer. Similarly, psychologists and clinical neuropsychologists often observe not only whether a person solves problems correctly (i.e., product), but how the client goes about attempting to solve the problem (i.e., process).

Test Administration

One of the most important distinctions relates to whether tests are *group administered* or are *individually administered* by a psychologist, physician, or technician. Tests that traditionally were group administered were *paper-and-pencil measures*. Often for these measures, the test-taker received both a test booklet and an answer sheet and was required, unless he or she had certain disabilities, to mark his or her responses on the answer sheet. In recent decades, some tests are administered using technology (i.e., computers and other electronic media). There may be some adaptive qualities to tests administered by computer, although not all computer-administered tests are adaptive (technology-administered tests are further discussed below). An *individually administered measure* is typically provided to the test-taker by a psychologist, physician, or technician. More faith is often provided to the individually administered measure, because the trained professional administering the test can make judgments during the testing that affect the administration, scoring, and other observations related to the test.

Tests can be administered in an *adaptive* or *linear* fashion, whether by computer or individual administrator. A linear test is one in which questions are administered one after another in a pre-arranged order. An adaptive test is one in which the test-taker's performance on earlier items affects

the questions he or she received subsequently. Typically, if the test-taker is answering the first questions correctly or in accordance with preset or expected response algorithms, for example, the next questions are still more difficult until the level appropriate for the examinee performance is best reached or the test is completed. If one does not answer the first questions correctly or as typically expected in the case of a non-cognitive measure, then easier questions would generally be presented to the test-taker.

Tests can be administered in *written* (keyboard or paper-and-pencil) fashion, *orally*, using an *assistive device* (most typically for individuals with motor disabilities), or in *performance format*, as previously noted. It is generally difficult to administer oral or performance tests in a group situation; however, some electronic media are making it possible to administer such tests without human examiners.

Another distinction among measures relates to who the respondent is. In most cases, the test-taker him- or herself is the respondent to any questions posed by the psychologist or physician. In the case of a young child, many individuals with autism, or an individual, for example, who has lost language ability, the examiner may need to ask others who know the individual (parents, teachers, spouses, family members) how they behave and to describe their personality, typical behaviors, and so on.

Scoring Differences

Tests are categorized as *objectively scored*, *subjectively scored*, or in some instances, both. An objectively scored instrument is one where the correct answers are counted and they either are, or they are converted to, the final scoring. Such tests may be scored manually or using optical scanning machines, computerized software, software used by other electronic media, or even templates (keys) that are placed over answer sheets where a person counts the number of correct answers. Examiner ratings and self-report interpretations are determined by the professional using a rubric or scoring system to convert the examinee's responses to a score, whether numerical or not. Sometimes subjective scores may include both quantitative and qualitative summaries or narrative descriptions of the performance of an individual.

Scores on tests are often considered to be *norm-referenced* (or *normative*) or *criterion-referenced*. Norm-referenced cognitive measures (such as college and graduate school admissions measures) inform the test-takers where they stand relative to others in the distribution. For example, an applicant to a college may learn that she is at the 60th percentile, meaning that she has scored better than 60 percent of those taking the test and less well than 40 percent of the same norm group. Likewise, most if not all intelligence tests are norm-referenced, and most other ability tests are as well.

In recent years there has been more of a call for criterion-referenced tests, especially in education (Hambleton and Pitoniak, 2006). For criterion-referenced tests, one's score is not compared to the other members of the test-taking population but rather to a fixed standard. High school graduation tests, licensure tests, and other tests that decide whether test-takers have met minimal competency requirements are examples of criterion-referenced measures. When one takes a driving test to earn one's driver's license, for example, one does not find out where one's driving falls in the distribution of national or statewide drivers, one only passes or fails.

Test Content

As noted previously, the most important distinction among most psychological tests is whether they are assessing cognitive versus non-cognitive qualities. In clinical psychological and neuropsychological settings such as are the concern of this volume, the most common cognitive tests are intelligence tests, other clinical neuropsychological measures, and performance validity measures. Many tests used by clinical neuropsychologists, psychiatrists, technicians, or others assess specific types of functioning, such as memory or problem solving. Performance validity measures are typically short assessments and are sometimes interspersed among components of other assessments that help the psychologist determine whether the examinee is exerting sufficient effort to perform well and responding to the best of his or her ability. Most common non-cognitive measures in clinical psychology and neuropsychology settings are personality measures and symptom validity measures. Some personality tests, such as the Minnesota Multiphasic Personality Inventory (MMPI), assess the degree to which someone expresses behaviors that are seen as atypical in relation to the norming sample.[1] Other personality tests are more normative and try to provide information about the client to the therapist. Symptom validity measures are scales, like performance validity measures, that may be interspersed throughout a longer assessment to examine whether a person is portraying him- or herself in an honest and truthful manner. Somewhere between these two types of tests—cognitive and non-cognitive—are various measures of adaptive functioning that often include both cognitive and non-cognitive components.

[1] This may be in comparison to a nationally representative norming sample, or with certain tests or measures, such as the MMPI, particular clinically diagnostic samples.

PSYCHOMETRICS: EXAMINING THE PROPERTIES OF TEST SCORES

Psychometrics is the scientific study—including the development, interpretation, and evaluation—of psychological tests and measures used to assess variability in behavior and link such variability to psychological phenomena. In evaluating the quality of psychological measures we are traditionally concerned primarily with test reliability (i.e., consistency), validity (i.e., accuracy of interpretations and use), and fairness (i.e., equivalence of usage across groups). This section provides a general overview of these concepts to help orient the reader for the ensuing discussions in Chapters 4 and 5. In addition, given the implications of applying psychological measures with subjects from diverse racial and ethnic backgrounds, issues of equivalence and fairness in psychological testing are also presented.

Reliability

Reliability refers to the degree to which scores from a test are stable and results are consistent. When constructs are not reliably measured the obtained scores will not approximate a true value in relation to the psychological variable being measured. It is important to understand that observed or obtained test scores are considered to be composed of true and error elements. A standard error of measurement is often presented to describe, within a level of confidence (e.g., 95 percent), that a given range of test scores contains a person's true score, which acknowledges the presence of some degree of error in test scores and that obtained test scores are only estimates of true scores (Geisinger, 2013).

Reliability is generally assessed in four ways:

1. *Test-retest*: Consistency of test scores over time (stability, temporal consistency);
2. *Inter-rater*: Consistency of test scores among independent judges;
3. *Parallel or alternate forms*: Consistency of scores across different forms of the test (stability and equivalence); and
4. *Internal consistency*: Consistency of different items intended to measure the same thing within the test (homogeneity). A special case of internal consistency reliability is split-half where scores on two halves of a single test are compared and this comparison may be converted into an index of reliability.

A number of factors can affect the reliability of a test's scores. These include time between two testing administrations that affect test-retest and alternate-forms reliability, and similarity of content and expectations of

subjects regarding different elements of the test in alternate forms, split-half, and internal consistency approaches. In addition, changes in subjects over time and introduced by physical ailments, emotional problems, or the subject's environment, or test-based factors such as poor test instructions, subjective scoring, and guessing will also affect test reliability. It is important to note that a test can generate reliable scores in one context and not in another, and that inferences that can be made from different estimates of reliability are not interchangeable (Geisinger, 2013).

Validity

While the scores resulting from a test may be deemed reliable, this finding does not necessarily mean that scores from the test have validity. *Validity* is defined as "the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests" (AERA et al., 2014, p. 11). In discussing validity, it is important to highlight that validity refers not to the measure itself (i.e., a psychological test is not valid or invalid) or the scores derived from the measure, but rather the interpretation and use of the measure's scores. To be considered valid, the interpretation of test scores must be grounded in psychological theory and empirical evidence that demonstrates a relationship between the test and what it purports to measure (Furr and Bacharach, 2013; Sireci and Sukin, 2013). Historically, the fields of psychology and education have described three primary types of evidence related to validity (Sattler, 2014; Sireci and Sukin, 2013):

1. *Construct evidence of validity*: The degree to which an individual's test scores correlate with the theoretical concept the test is designed to measure (i.e., evidence that scores on a test correlate relatively highly with scores on theoretically similar measures and relatively poorly with scores on theoretically dissimilar measures);
2. *Content evidence of validity*: The degree to which the test content represents the targeted subject matter and supports a test's use for its intended purposes; and
3. *Criterion-related evidence of validity*: The degree to which the test's score correlates with other measurable, reliable, and relevant variables (i.e., criterion) thought to measure the same construct.

Other kinds of validity with relevance to SSA have been advanced in the literature, but are not completely accepted in professional standards as types of validity per se. These include

1. *Diagnostic validity*: The degree to which psychological tests are truly aiding in the formulation of an appropriate diagnosis.

2. *Ecological validity*: The degree to which test scores represent everyday levels of functioning (e.g., impact of disability on an individual's ability to function independently).
3. *Cultural validity*: The degree to which test content and procedures accurately reflect the sociocultural context of the subjects being tested.

Each of these forms of validity poses complex questions regarding the use of particular psychological measures with the SSA population. For example, ecological validity is especially critical in the use of psychological tests with SSA given that the focus of the assessment is on examining everyday levels of functioning. Measures like intelligence tests have been sometimes criticized for lacking ecological validity (Groth-Marnat, 2009; Groth-Marnat and Teal, 2000). Alternatively, "research suggests that many neuropsychological tests have a moderate level of ecological validity when predicting everyday cognitive functioning" (Chaytor and Schmitter-Edgecombe, 2003, p. 181).

More recent discussions on validity have shifted toward an argument-based approach to validity, using a variety of evidence to build a case for validity of test score interpretation (Furr and Bacharach, 2013). In this approach, construct validity is viewed as an overarching paradigm under which evidence is gathered from multiple sources to build a case for validity of test score interpretation. Five key sources of validity evidence that affect the degree to which a test fulfills its purpose are generally considered (AERA et al., 2014; Furr and Bacharach, 2013; Sireci and Sukin, 2013):

1. *Test content*: Does the test content reflect the important facets of the construct being measured? Are the test items relevant and appropriate for measuring the construct and congruent with the purpose of testing?
2. *Relation to other variables*: Is there a relationship between test scores and other criterion or constructs that are expected to be related?
3. *Internal structure*: Does the actual structure of the test match the theoretically based structure of the construct?
4. *Response processes*: Are respondents applying the theoretical constructs or processes the test is designed to measure?
5. *Consequences of testing*: What are the intended and unintended consequences of testing?

Standardization and Testing Norms

As part of the development of any psychometrically sound measure, explicit methods and procedures by which tasks should be administered are determined and clearly spelled out. This is what is commonly known as *standardization*. Typical standardized administration procedures or expectations include (1) a quiet, relatively distraction-free environment, (2) precise reading of scripted instructions, and (3) provision of necessary tools or stimuli. All examiners use such methods and procedures during the process of collecting the normative data, and such procedures normally should be used in any other administration, which enables application of normative data to the individual being evaluated (Lezak et al., 2012).

Standardized tests provide a set of normative data (i.e., norms), or scores derived from groups of people for whom the measure is designed (i.e., the designated population) to which an individual's performance can be compared. Norms consist of transformed scores such as percentiles, cumulative percentiles, and standard scores (e.g., T-scores, Z-scores, stanines, IQs), allowing for comparison of an individual's test results with the designated population. Without standardized administration, the individual's performance may not accurately reflect his or her ability. For example, an individual's abilities may be overestimated if the examiner provides additional information or guidance than what is outlined in the test administration manual. Conversely, a claimant's abilities may be underestimated if appropriate instructions, examples, or prompts are not presented. When nonstandardized administration techniques must be used, norms should be used with caution due to the systematic error that may be introduced into the testing process; this topic is discussed in detail later in the chapter.

It is important to clearly understand the population for which a particular test is intended. The standardization sample is another name for the norm group. Norms enable one to make meaningful interpretations of obtained test scores, such as making predictions based on evidence. Developing appropriate norms depends on size and representativeness of the sample. In general, the more people in the norm group the closer the approximation to a population distribution so long as they represent the group who will be taking the test.

Norms should be based upon representative samples of individuals from the intended test population, as each person should have an equal chance of being in the standardization sample. Stratified samples enable the test developer to identify particular demographic characteristics represented in the population and more closely approximate these features in proportion to the population. For example, intelligence test scores are often established based upon census-based norming with proportional representation of

demographic features including race and ethnic group membership, parental education, socioeconomic status, and geographic region of the country.

When tests are applied to individuals for whom the test was not intended and, hence, were not included as part of the norm group, inaccurate scores and subsequent misinterpretations may result. Tests administered to persons with disabilities often raise complex issues. Test users sometimes use psychological tests that were not developed or normed for individuals with disabilities. It is critical that tests used with such persons (including SSA disability claimants) include attention to representative norming samples; when such norming samples are not available, it is important for the assessor to note that the test or tests used are not based on representative norming samples and the potential implications for interpretation (Turner et al., 2001).

Test Fairness in High-Stakes Testing Decisions

Performance on psychological tests often has significant implications (high stakes) in our society. Tests are in part the gatekeepers for educational and occupational opportunities and play a role in SSA determinations. As such, results of psychological testing may have positive or negative consequences for an individual. Often such consequences are intended; however, there is the possibility for unintended negative consequences. It is imperative that issues of test fairness be addressed so no individual or group is disadvantaged in the testing process based upon factors unrelated to the areas measured by the test. Biases simply cannot be present in these kinds of professional determinations. Moreover, it is imperative that research demonstrates that measures can be fairly and equivalently used with members of the various subgroups in our population. It is important to note that there are people from many language and cultural groups for whom there are no available tests with norms that are appropriately representative for them. As noted above, in such cases it is important for assessors to include a statement about this situation whenever it applies and potential implications on scores and resultant interpretation.

While all tests reflect what is valued within a particular cultural context (i.e., cultural loading), bias refers to the presence of systematic error in the measurement of a psychological construct. Bias leads to inaccurate test results given that scores reflect either overestimations or underestimations of what is being measured. When bias occurs based upon culturally related variables (e.g., race, ethnicity, social class, gender, educational level) then there is evidence of cultural test bias (Suzuki et al., 2014).

Relevant considerations pertain to issues of equivalence in psychological testing as characterized by the following (Suzuki et al., 2014, p. 260):

1. *Functional*: Whether the construct being measured occurs with equal frequency across groups;
2. *Conceptual*: Whether the item information is familiar across groups and means the same thing in various cultures;
3. *Scalar*: Whether average score differences reflect the same degree, intensity, or magnitude for different cultural groups;
4. *Linguistic*: Whether the language used has similar meaning across groups; and
5. *Metric*: Whether the scale measures the same behavioral qualities or characteristics and the measure has similar psychometric properties in different cultures.

It must be established that the measure is operating appropriately in various cultural contexts. Test developers address issues of equivalence through procedures including

- Expert panel reviews (i.e., professionals review item content and provide informed judgments regarding potential biases);
- Examination of differential item functioning (DIF) among groups;
- Statistical procedures allowing comparison of psychometric features of the test (e.g., reliability coefficients) based on different population samples;
- Exploratory and confirmatory factor analysis, structural equation modeling (i.e., examination of the similarities and differences of the constructs structure), and measurement invariance; and
- Mean score differences taking into consideration the spread of scores within particular racial and ethnic groups as well as among groups.

Cultural equivalence refers to whether "interpretations of psychological measurements, assessments, and observations are similar if not equal across different ethnocultural populations" (Trimble, 2010, p. 316). Cultural equivalence is a higher order form of equivalence that is dependent on measures meeting specific criteria indicating that a measure may be appropriately used with other cultural groups beyond the one for which it was originally developed. Trimble (2010) notes that there may be upward of 50 or more types of equivalence that affect interpretive and procedural practices in order to establish cultural equivalence.

Item Response Theory and Tests[2]

For most of the 20th century, the dominant measurement model was called classical test theory. This model was based on the notion that all scores were composed of two components: true score and error. One can imagine a "true score" as a hypothetical value that would represent a person's actual score were there no error present in the assessment (and unfortunately, there is always some error, both random and systematic). The model further assumes that all error is random and that any correlation between error and some other variable, such as true scores, is effectively zero (Geisinger, 2013). The approach leans heavily on reliability theory, which is largely derived from the premises mentioned above.

Since the 1950s and largely since the 1970s, a newer mathematically sophisticated model developed called item response theory (IRT). The premise of these IRT models is most easily understood in the context of cognitive tests, where there is a correct answer to questions. The simplest IRT model is based on the notion that the answering of a question is generally based on only two factors: the difficulty of the question and the ability level of the test-taker. Computer-adaptive testing estimates scores of the test-taker after each response to a question and adjusts the administration of the next question accordingly. For example, if a test-taker answers a question correctly, he or she is likely to receive a more difficult question next. If one, on the other hand, answers incorrectly, he or she is more likely to receive an easier question, with the "running score" held by the computer adjusted accordingly. It has been found that such computer-adaptive tests can be very efficient.

IRT models have made the equating of test forms far easier. Equating tests permits one to use different forms of the same examination with different test items to yield fully comparable scores due to slightly different item difficulties across forms. To convert the values of item difficulty to determine the test-taker's ability scores one needs to have some common items across various tests; these common items are known as anchor items. Using such items, one can essentially establish a fixed reference group and base judgments from other groups on these values.

As noted above, there are a number of common IRT models. Among the most common are the one-, two-, and three-parameter models. The one-parameter model is the one already described; the only item parameter is item difficulty. A two-parameter model adds a second parameter to the first, related to item discrimination. *Item discrimination* is the ability of the item to differentiate those lacking the ability in high degree from those holding it. Such two-parameter models are often used for tests like essay tests where

[2] The brief overview presented here draws on the works of De Ayala (2009) and DeMars (2010), to which the reader is directed for additional information.

one cannot achieve a high score by guessing or using other means to answer currently. The three-parameter IRT model contains a third parameter, that factor related to chance level correct scoring. This parameter is sometimes called the pseudo-guessing parameter, and this model is generally used for large-scale multiple-choice testing programs.

These models, because of their lessened reliance on the sampling of test-takers, are very useful in the equating of tests that is the setting of scores to be equivalent regardless of the form of the test one takes. In some high-stakes admissions tests such as the GRE, MCAT, and GMAT, for example, forms are scored and equated by virtue of IRT methods, which can perform such operations more efficiently and accurately than can be done with classical statistics.

TEST USER QUALIFICATIONS

The *test user* is generally considered the person responsible for appropriate use of psychological tests, including selection, administration, interpretation, and use of results (AERA et al., 2014). Test user qualifications include attention to the purchase of psychological measures that specify levels of training, educational degree, areas of knowledge within domain of assessment (e.g., ethical administration, scoring, and interpretation of clinical assessment), certifications, licensure, and membership in professional organizations. Test user qualifications require psychometric knowledge and skills as well as training regarding the responsible use of tests (e.g., ethics), in particular, psychometric and measurement knowledge (i.e., descriptive statistics, reliability and measurement error, validity and the meaning of test scores, normative interpretation of test scores, selection of appropriate tests, and test administration procedures). In addition, test user guidelines highlight the importance of understanding the impact of ethnic, racial, cultural, gender, age, educational, and linguistic characteristics in the selection and use of psychological tests (Turner et al., 2001).

Test publishers provide detailed manuals regarding the operational definition of the construct being assessed, norming sample, reading level of test items, completion time, administration, and scoring and interpretation of test scores. Directions presented to the examinee are provided verbatim and sample responses are often provided to assist the examiner in determining a right or wrong response or in awarding numbers of points to a particular answer. Ethical and legal knowledge regarding assessment competencies, confidentiality of test information, test security, and legal rights of test-takers are imperative. Resources like the *Mental Measurements Yearbook* (*MMY*) provide descriptive information and evaluative reviews of commercially available tests to promote and encourage informed test selection

(Buros, 2015). To be included, tests must contain sufficient documentation regarding their psychometric quality (e.g., validity, reliability, norming).

Test Administration and Interpretation

In accordance with the Standards for Educational and Psychological Testing (AERA et al., 2014) and the APA's Guidelines for Test User Qualifications (Turner et al., 2001), many publishers of psychological tests employ a tiered system of qualification levels (generally A, B, C) required for the purchase, administration, and interpretation of such tests (e.g., PAR, n.d.; Pearson Education, 2015). Many instruments, such as those discussed throughout this report, would be considered qualification level C assessment methods, generally requiring an advanced degree, specialized psychometric and measurement knowledge, and formal training in administration, scoring, and interpretation. However, some may have less stringent requirements, for example, a bachelor's or master's degree in a related field and specialized training in psychometric assessment (often classified level B), or no special requirements (often classified level A) for purchase and use. While such categories serve as a general guide for necessary qualifications, individual test manuals provide additional detail and specific qualifications necessary for administration, scoring, and interpretation of the test or measure.

Given the need for the use of standardized procedures, any person administering cognitive or neuropsychological measures must be well trained in standardized administration protocols. He or she should possess the interpersonal skills necessary to build rapport with the individual being tested in order to foster cooperation and maximal effort during testing. Additionally, individuals administering tests should understand important psychometric properties, including validity and reliability, as well as factors that could emerge during testing to place either at risk. Many doctoral-level psychologists are well trained in test administration; in general, psychologists from clinical, counseling, school, or educational graduate psychology programs receive training in psychological test administration. For cases in which cognitive deficits are being evaluated, a neuropsychologist may be needed to most accurately evaluate cognitive functioning (see Chapter 5 for a more detailed discussion on administration and interpretation of cognitive tests). The use of non-doctoral-level psychometrists or technicians in psychological and neuropsychological test administration and scoring is also a widely accepted standard of practice (APA, 2010; Brandt and van Gorp, 1999; Pearson Education, 2015). Psychometrists are often bachelor's- or master's-level individuals who have received additional specialized training in standardized test administration and scoring. They do not practice independently or interpret test scores, but rather work under

the close supervision and direction of doctoral-level clinical psychologists or neuropsychologists.

Interpretation of testing results requires a higher degree of clinical training than administration alone. Threats to the validity of any psychological measure of a self-report nature oblige the test interpreter to understand the test and principles of test construction. In fact, interpreting tests results without such knowledge would violate the ethics code established for the profession of psychology (APA, 2010). SSA requires psychological testing be "individually administered by a qualified specialist … currently licensed or certified in the state to administer, score, and interpret psychological tests and have the training and experience to perform the test" (SSA, n.d.). Most doctoral-level clinical psychologists who have been trained in psychometric test administration are also trained in test interpretation. SSA (n.d.) also requires individuals who administer more specific cognitive or neuropsychological evaluations "be properly trained in this area of neuroscience." As such, clinical neuropsychologists—individuals who have been specifically trained to interpret testing results within the framework of brain-behavior relationships and who have achieved certain educational and training benchmarks as delineated by national professional organizations—may be required to interpret tests of a cognitive nature (AACN, 2007; NAN, 2001).

Use of Interpreters and Other Nonstandardized Test Administration Techniques

Modification of procedures, including the use of interpreters and the administration of nonstandardized assessment procedures, may pose unique challenges to the psychologist by potentially introducing systematic error into the testing process. Such errors may be related to language, the use of translators, or examinee abilities (e.g., sensory, perceptual, and/or motor capacity). For example, if one uses a language interpreter, the potential for mistranslation may yield inaccurate scores. Use of translators is a nonpreferred option, and assessors need to be familiar with both the language and culture from which an individual comes to properly interpret test results, or even infer whether specific measures are appropriate. The adaptation of tests has become big business for testing companies, and many tests, most often measures developed in English for use in the United States, are being adapted for use in other countries. Such measures require changes in language, but translators must also be knowledgeable about culture and the environment of the region from which a person comes (ITC, 2005).

For sensory, perceptual, or motor abilities, one may be altering the construct that the test is designed to measure. In both of these examples, one could be obtaining scores for which there is no referenced normative

group to allow for accurate interpretation of results. While a thorough discussion of these concepts is beyond the scope of this report and is presented elsewhere, it may be stated that when a test is administered following a procedure that is outside of that which has been developed in the standardization process, conclusions drawn must recognize the potential for error in their creation.

PSYCHOLOGICAL TESTING IN THE CONTEXT OF DISABILITY DETERMINATIONS

As noted in Chapter 2, SSA indicates that *objective medical evidence* may include the results of standardized psychological tests. Given the great variety of psychological tests, some are more objective than others. Whether a psychological test is appropriately considered objective has much to do with the process of scoring. For example, unstructured measures that call for open-ended responding rely on professional judgment and interpretation in scoring; thus, such measures are considered less than objective. In contrast, standardized psychological tests and measures, such as those discussed in the ensuing chapters, are structured and objectively scored. In the case of non-cognitive self-report measures, the respondent generally answers questions regarding typical behavior by choosing from a set of predetermined answers. With cognitive tests, the respondent answers questions or solves problems, which usually have correct answers, as well as he or she possibly can. Such measures generally provide a set of normative data (i.e., norms), or scores derived from groups of people for whom the measure is designed (i.e., the designated population), to which an individual's responses or performance can be compared. Therefore, standardized psychological tests and measures rely less on clinical judgment and are considered to be more objective than those that depend on subjective scoring. Unlike measurements such as weight or blood pressure standardized psychological tests require the individual's cooperation with respect to self-report or performance on a task. The inclusion of validity testing, which will be discussed further in Chapters 4 and 5, in the test or test battery allows for greater confidence in the test results. Standardized psychological tests that are appropriately administered and interpreted can be considered objective evidence.

The use of psychological tests in disability determinations has critical implications for clients. As noted earlier, issues surrounding ecological validity (i.e., whether test performance accurately reflects real-world behavior) is of primary importance in SSA determination. Two approaches have been identified in relation to the ecological validity of neuropsychological assessment. The first focuses on "how well the test captures the essence of everyday cognitive skills" in order to "identify people who have difficulty

performing real-world tasks, regardless of the etiology of the problem" (i.e., verisimilitude), and the second "relates performance on traditional neuropsychological tests to measures of real-world functioning, such as employment status, questionnaires, or clinician ratings" (i.e., veridicality) (Chaytor and Schmitter-Edgecombe, 2003, pp. 182–183). Establishing ecological validity is a complicated endeavor given the potential effect of non-cognitive factors (e.g., emotional, physical, and environmental) on test and everyday performance. Specific concerns regarding test performance include (1) the test environment is often not representative (i.e., artificial), (2) testing yields only samples of behavior that may fluctuate depending on context, and (3) clients may possess compensatory strategies that are not employable during the testing situation; therefore, obtained scores underestimate the test-taker's abilities.

Activities of daily living (ADLs) and the client's likelihood of returning to work are important considerations in disability determinations. Occupational status, however, is complex and often multidetermined requiring that psychological test data be complemented with other sources of information in the evaluation process (e.g., observation, informant ratings, environmental assessments) (Chaytor and Schmitter-Edgecombe, 2003). Table 3-1 highlights major mental disorders, relevant types of psychological measures, and domains of functioning.

Determination of disability is dependent on two key factors: the existence of a medically determinable impairment and associated limitations on functioning. As discussed in detail in Chapter 2, applications for disability follow a five-step sequential disability determination process. At Step 3 in the process, the applicant's reported impairments are evaluated to determine whether they meet or equal the medical criteria codified in SSA's *Listing of Impairments*. This includes specific symptoms, signs, and laboratory findings that substantiate the existence of an impairment (i.e., Paragraph A criteria) and evidence of associated functional limitations (i.e., Paragraph B criteria). If an applicant's impairments meet or equal the listing criteria, the claim is allowed. If not, residual functional capacity, including mental residual functional capacity, is assessed. This includes whether the applicant has the capacity for past work (Step 4) or any work in the national economy (Step 5).

SSA uses a standard assessment that examines functioning in four domains: understanding and memory, sustained concentration and persistence, social interaction, and adaptation. Psychological testing may play a key role in understanding a client's functioning in each of these areas. Box 3-1 describes ways in which these four areas of core mental residual functional capacity are assessed ecologically. Psychological assessments often address these areas in a more structured manner through interviews, standardized measures, checklists, observations, and other assessment procedures.

TABLE 3-1 Listings for Mental Disorders and Types of Psychological Tests

Mental Disorder	Psychological Assessment Measures and Methods	Relevant Cognitive Domains of Functioning	Psychiatric Symptoms (per SSA [n.d.] Listings)
Organic mental disorders (e.g., delirium, dementia, amnestic)	Screening instruments (e.g., checklists, questionnaires) Memory and cognitive tests Interview Observations	Cognitive/intellectual ability Language and communication Memory acquisition Attention and distractibility Processing speed Executive functioning Adaptive functioning	Disorientation to time and place Memory impairment Perceptual or thinking disturbances Change in personality Disturbance in mood Emotional lability Loss of measured intellectual ability from premorbid levels or overall impairment
Schizophrenic, paranoid, and other psychotic disorders	Screening instruments Personality tests Interview Observations Cognitive tests	Cognitive/intellectual ability Language and communication Memory acquisition Attention and distractibility Processing speed Executive functioning	Delusions or hallucinations Catatonic or other grossly disorganized behavior Incoherence, loosening of associations, illogical thinking, or poverty of content of speech if associated with one of the following: • Blunt affect • Flat affect • Inappropriate affect • Emotional withdrawal and/or isolation

continued

TABLE 3-1 Continued

Mental Disorder	Psychological Assessment Measures and Methods	Relevant Cognitive Domains of Functioning	Psychiatric Symptoms (per SSA [n.d.] Listings)
Affective (mood) disorders	Personality tests Interview Observations Cognitive tests	Memory acquisition Attention and distractibility Processing speed Executive functioning	Depressive syndrome characterized by at least four of the following: • Anhedonia or pervasive loss of interest in almost all activities • Appetite disturbance with change in weight • Sleep disturbance • Psychomotor agitation or retardation • Decreased energy • Feelings of guilt or worthlessness • Difficulty concentrating or thinking • Thoughts of suicide • Hallucinations, delusions, or paranoid thinking Manic syndrome characterized by at least three of the following: • Hyperactivity • Pressure of speech • Flight of ideas • Inflated self-esteem • Decreased need for sleep • Easy distractibility • Involvement in activities that have a high probability of painful consequences that are not recognized • Hallucinations, delusions, or paranoid thinking Bipolar syndrome with a history of episodic periods manifested by the full symptomatic picture of both manic and depressive syndromes (and currently characterized by either or both syndromes)

TABLE 3-1 Continued

Mental Disorder	Psychological Assessment Measures and Methods	Relevant Cognitive Domains of Functioning	Psychiatric Symptoms (per SSA [n.d.] Listings)
Intellectual disability disorders	Cognitive tests	Cognitive/intellectual ability Language and communication Memory acquisition Attention and distractibility Processing speed Executive functioning Adaptive functioning	Mental incapacity evidenced by dependence on others for personal needs (e.g., toileting, eating, dressing, or bathing) and inability to follow directions, such that the use of standardized measures of intellectual functioning is precluded
Anxiety-related disorders	Personality tests Screening instruments Cognitive tests	Cognitive/intellectual ability Language and communication Memory acquisition Attention and distractibility Processing speed Executive functioning	Generalized persistent anxiety accompanied by three out of four of the following signs or symptoms: • Motor tension • Autonomic hyperactivity • Apprehensive expectation • Vigilance and scanning A persistent irrational fear of a specific object, activity, or situation that results in a compelling desire to avoid the dreaded object, activity, or situation Recurrent severe panic attacks manifested by a sudden unpredictable onset of intense apprehension, fear, terror, and sense of impending doom occurring on the average of at least once per week Recurrent obsessions or compulsions that are a source of marked distress Recurrent and intrusive recollections of a traumatic experience that are a source of marked distress

continued

TABLE 3-1 Continued

Mental Disorder	Psychological Assessment Measures and Methods	Relevant Cognitive Domains of Functioning	Psychiatric Symptoms (per SSA [n.d.] Listings)
Somatoform disorders	Personality tests Cognitive tests	Cognitive/intellectual ability Language and communication Memory acquisition Attention and distractibility Processing speed Executive functioning	A history of multiple physical symptoms of several years duration, beginning before age 30, that have caused the individual to take medicine frequently, see a physician often, and alter life patterns significantly Persistent nonorganic disturbance of one of the following: • Vision • Speech • Hearing • Use of a limb • Movement and its control (e.g., coordination disturbance, psychogenic seizures, akinesia, dyskinesia) • Sensation (e.g., diminished or heightened) Unrealistic interpretation of physical signs or sensations associated with the preoccupation or belief that one has a serious disease or injury

TABLE 3-1 Continued

Mental Disorder	Psychological Assessment Measures and Methods	Relevant Cognitive Domains of Functioning	Psychiatric Symptoms (per SSA [n.d.] Listings)
Personality disorders	Personality tests		Deeply ingrained, maladaptive patterns of behavior associated with one of the following: • Seclusiveness or autistic thinking • Pathologically inappropriate suspiciousness or hostility • Oddities of thought, perception, speech, and behavior • Persistent disturbances of mood or affect • Pathological dependence, passivity, or aggressivity • Intense and unstable interpersonal relationships and impulsive and damaging behavior
Substance addiction disorders	Interviews Screening instruments	Memory acquisition Attention and distractibility Processing speed Executive functioning	Behavioral changes or physical changes associated with the regular use of substances that affect the central nervous system
Autistic disorder and other pervasive develop-mental disorders	Observations Screening instruments Checklists Rating scales Cognitive tests	Cognitive/intellectual ability Language and communication Memory acquisition Attention and distractibility Processing speed Executive functioning	Qualitative deficits in reciprocal social interaction Qualitative deficits in verbal and nonverbal communication and in imaginative activity Markedly restricted repertoire of activities and interests
Attention deficit hyperactiv-ity disorder (children)	Observations Screening instruments Checklists Rating scales Cognitive tests	Cognitive/intellectual ability Memory acquisition Attention and distractibility Processing speed Executive functioning	Developmentally inappropriate degrees of inattention, impulsiveness, and hyperactivity

continued

TABLE 3-1 Continued

Mental Disorder	Psychological Assessment Measures and Methods	Relevant Cognitive Domains of Functioning	Psychiatric Symptoms (per SSA [n.d.] Listings)
Developmental and emotional disorders of newborns and infants	Interviews with parents/ caregivers Observations, scales of infant development	Cognitive/intellectual ability Language and communication	Deficit or lag in social functioning Apathy, overexcitability, or fearfulness, demonstrated by an absent or grossly excessive response to one of the following: • Visual stimulation • Auditory stimulation • Tactile stimulation
RELATED DIAGNOSTIC ENTITIES			
Traumatic brain injury	Cognitive tests	Cognitive/intellectual ability Language and communication Memory acquisition Attention and distractibility Processing speed Executive functioning	
Cognitive dysfunction	Cognitive tests	Cognitive/intellectual ability Language and communication Memory acquisition Attention and distractibility Processing speed Executive functioning	

BOX 3-1
Descriptions of Tests by Four Areas of Core Mental Residual Functional Capacity*

Understanding and Memory	• Remember location and work-like procedures • Understand and remember very short and simple instructions • Understand and remember detailed instructions
Sustained Concentration and Persistence	• Carry out very short and simple instructions • Carry out detailed instructions • Maintain attention and concentration for extended periods • Perform activities within a schedule, maintain regular attendance, and be punctual within a customary tolerance • Sustain an ordinary routine without special supervision • Work in coordination with and proximity to others without being distracted by them • Make simple work-related decisions • Complete a normal workday and workweek without interruptions from psychologically based symptoms, and perform at a consistent pace without an unreasonable number or length of rest periods
Social Interaction	• Interact appropriately with the general public • Ask simple questions or request assistance • Get along with co-workers or peers without distracting them or exhibiting behavioral extremes • Maintain socially appropriate behavior, and adhere to basic standards of neatness and cleanliness
Adaptation	• Respond appropriately to changes in the work setting • Be aware of normal hazards, and take appropriate precautions • Travel to unfamiliar places, or use public transportation • Set realistic goals, or make plans independently of others

* Adapted from Form SSA-4734-F4-SUP: Mental Residual Functional Capacity Assessment.

This chapter has identified some of the basic foundations underlying the use of psychological tests including basic psychometric principles and issues regarding test fairness. Applications of tests can inform disability determinations. The next two chapters build on this overview, examining the types of psychological tests that may be useful in this process, including a review of selected individual tests that have been developed for measuring validity of presentation. Chapter 4 focuses on non-cognitive, self-report measures and symptom validity tests. Chapter 5 then focuses on cognitive tests and associated performance validity tests. Strengths and limitations of various instruments are offered, in order to subsequently explore the relevance for different types of tests for different claims, per category of disorder, with a focus on establishing the validity of the client's claim.

REFERENCES

AACN (American Academy of Clinical Neuropsychology). 2007. AACN practice guidelines for neuropsychological assessment and consultation. *Clinical Neuropsychology* 21(2):209-231.

AERA (American Educational Research Association), APA (American Psychological Association), and NCME (National Council on Measurement in Education). 2014. *Standards for educational and psychological testing*. Washington, DC: AERA.

APA. 2010. *Ethical principles of psychologists and code of conduct*. http://www.apa.org/ethics/code (accessed March 9, 2015).

Brandt, J., and W. van Gorp. 1999. American Academy of Clinical Neuropsychology policy on the use of non-doctoral-level personnel in conducting clinical neuropsychological evaluations. *The Clinical Neuropsychologist* 13(4):385-385.

Buros Center for Testing. 2015. *Test reviews and information*. http://buros.org/test-reviews-information (accessed March 19, 2015).

Chaytor, N., and M. Schmitter-Edgecombe. 2003. The ecological validity of neuropsychological tests: A review of the literature on everyday cognitive skills. *Neuropsychology Review* 13(4):181-197.

Cronbach, L. J. 1949. *Essentials of psychological testing*. New York: Harper.

Cronbach, L. J. 1960. *Essentials of psychological testing*. 2nd ed. Oxford, England: Harper.

De Ayala, R. J. 2009. *Theory and practice of item response theory*. New York: Guilford Publications.

DeMars, C. 2010. *Item response theory*. New York: Oxford University Press.

Furr, R. M., and V. R. Bacharach. 2013. *Psychometrics: An introduction*. Thousand Oaks, CA: Sage Publications, Inc.

Geisinger, K. F. 2013. Reliability. In *APA handbook of testing and assessment in psychology*. Vol. 1, edited by K. F. Geisinger (editor) and B. A. Bracken, J. F. Carlson, J. C. Hansen, N. R. Kuncel, S. P. Reise, and M. C. Rodriguez (associate editors). Washington, DC: APA.

Groth-Marnat, G. 2009. *Handbook of psychological assessment*. Hoboken, NJ: John Wiley & Sons.

Groth-Marnat, G., and M. Teal. 2000. Block design as a measure of everyday spatial ability: A study of ecological validity. *Perceptual and Motor Skills* 90(2):522-526.

Hambleton, R. K., and M. J. Pitoniak. 2006. Setting performance standards. *Educational Measurement* 4:433-470.

ITC (International Test Commission). 2005. ITC guidelines for translating and adaptating tests. Geneva, Switzerland: ITC.

Lezak, M., D. Howieson, E. Bigler, and D. Tranel. 2012. *Neuropsychological assessment*. 5th ed. New York: Oxford University Press.

NAN (National Academy of Neuropsychology). 2001. *NAN definition of a clinical neuropsychologist: Official position of the National Academy of Neuropsychology.* https://www.nanonline.org/docs/PAIC/PDFs/NANPositionDefNeuro.pdf (accessed November 25, 2014).

PAR (Psychological Assessment Resources). 2015. *Qualifications levels*. http://www4.parinc.com/Supp/Qualifications.aspx (accessed January 5, 2015).

Pearson Education. 2015. *Qualifications policy.* http://www.pearsonclinical.com/psychology/qualifications.html (accessed January 5, 2015).

Sattler, J. M. 2014. *Foundations of behavioral, social, and clinical assessment of children*. 6th ed. La Mesa, CA: Jerome M. Sattler, Publisher, Inc.

Sireci, S. G., and T. Sukin. 2013. Test validity. In *APA handbook of testing and assessment in psychology*. Vol. 1, edited by K. F. Geisinger (editor) and B. A. Bracken, J. F. Carlson, J. C. Hansen, N. R. Kuncel, S. P. Reise, and M. C. Rodriguez (associate editors). Washington, DC: APA.

SSA (Social Security Administration). n.d. *Disability evaluation under social security—Part III: Listing of impairments—Adult listings (Part A)—section 12.00 mental disorders*. http://www.ssa.gov/disability/professionals/bluebook/12.00-MentalDisorders-Adult.htm (accessed November 14, 2014).

Suzuki, L. A., S. Naqvi, and J. S. Hill. 2014. Assessing intelligence in a cultural context. In *APA handbook of multicultural psychology.* Vol. 1, edited by F. T. L. Leong, L. Comas-Diaz, G. C. Nagayama Hall, V. C. McLoyd, and J. E. Trimble. Washington, DC: APA.

Trimble, J. E. 2010. Cultural measurement equivalence. In *Encyclopedia of cross-cultural school psychology*. New York: Springer. Pp. 316-318.

Turner, S. M., S. T. DeMers, H. R. Fox, and G. Reed. 2001. APA's guidelines for test user qualifications: An executive summary. *American Psychologist* 56(12):1099.

Weiner, I. B. 2003. The assessment process. In *Handbook of psychology*, edited by I. B. Weiner. Hoboken, NJ: John Wiley & Sons.

4

Self-Report Measures and Symptom Validity Tests

Allegations of disability are sometimes made on the basis of self-report, with few, if any, medical signs or laboratory findings to substantiate such claims. Often in these cases a medical source or consultative examiner may corroborate a claimant's history and allegations, finding them consistent with a medically determinable impairment that causes a particular level of functional limitation; however, the claim is still based primarily on self-report. Currently, such evidence may be deemed sufficient to grant disability benefits, albeit via a somewhat inconsistent process that varies from one state to another. A more systematic approach to assessing and verifying such claims would improve the consistency and reliability of the determination process in these cases.

To receive benefits, applicants must prove the existence of a medically determinable physical or mental impairment and associated functional limitations that result in an inability to engage in any substantial gainful activity. The U.S. Social Security Administration (SSA) (n.d.-b) defines a *medically determinable impairment* as

> an impairment that results from anatomical, physiological, or psychological abnormalities which can be shown by medically acceptable clinical and laboratory diagnostic techniques … [and] must be established by medical evidence consisting of signs, symptoms, and laboratory findings—not only by the individual's statement of symptoms.

Following establishment of a medically determinable impairment, the overall degree of functional limitation is evaluated based on the extent to which the applicant's impairment interferes with his or her "ability to

function independently, appropriately, effectively, and on a sustained basis" (20 CFR § 416.920a). SSA definitions of symptoms, signs, and laboratory findings are provided in Box 4-1.

The current chapter focuses on the potential role of non-cognitive psychological measures, often characterized as self-report measures, in SSA disability determinations. It begins with an examination of potential domains for which psychological self-report measures may provide information to assist in identifying a claimant's medically determinable impairment and determining the level of functional limitation. Following this, procedures and qualifications for administering tests and interpreting test results are presented. Finally, the chapter concludes with an examination of related symptom validity tests (SVTs).

ASSESSING SELF-REPORT OF SYMPTOMS

For claims based entirely on self-report, it is important to use a systematic method for identifying and documenting a medically determinable impairment and assessing the severity of associated functional limitations. A variety of standardized self-report measures exist that could further systematize SSA's disability determination process. Before delving into such measures, it is important to briefly address the distinction between self-report of

BOX 4-1
SSA Definitions of Symptoms, Signs, and Laboratory Findings

Symptoms: Your own description of your physical or mental impairment.

Signs: Anatomical, physiological, or psychological abnormalities that can be observed, apart from your statements (symptoms). Signs must be shown by medically acceptable clinical diagnostic techniques. Psychiatric signs are medically demonstrable phenomena that indicate specific psychological abnormalities, e.g., abnormalities of behavior, mood, thought, memory, orientation, development, or perception. They must also be shown by observable facts that can be medically described and evaluated.

Laboratory findings: Anatomical, physiological, or psychological phenomena that can be shown by the use of medically acceptable laboratory diagnostic techniques. Some of these diagnostic techniques include chemical tests, electrophysiological studies (electrocardiogram, electroencephalogram, etc.), roentgenological studies (X-rays), and psychological tests.

SOURCE: 20 CFR § 404.1528.

symptoms and self-report measures. As noted above, SSA defines symptoms as "the claimant's own description of [his or her] physical or mental impairment, [which] alone are not enough to establish that there is a physical or mental impairment" (20 CFR § 404.1528). In some cases, such as with children, symptoms may be reported by a third party, for example, a parent or a teacher. The committee refers to this as *self-report of symptoms*. Alternatively, there exist standardized instruments that rely on self-report (for example, of symptoms, behaviors, personality characteristics and/or traits, interests, values, and attitudes) with population-based normative data that allow the examiner to compare an individual's reported behaviors or symptoms with an appropriate comparison group (e.g., those of the same age group, sex, education level, and/or race/ethnicity). According to SSA regulations, such instruments may be considered medically acceptable laboratory diagnostic techniques, and thus provide signs and laboratory findings that corroborate the claimant's self-report of symptoms. The committee refers to these instruments as *self-report measures*.

Among these self-report measures are those that traditionally have been referred to as psychological tests, such as personality, multiscale, or single syndrome inventories and standardized psychiatric diagnostic interviews. These measures generally assess non-cognitive psychological complaints, and are therefore referred to as *non-cognitive measures*.[1] However, it is also important to note that some standardized self-report measures that might be useful to SSA in such cases are *not* considered psychological tests or measures. Examples may include standardized measures of pain, fatigue, sleep, or adaptive living. Some of these may contain internal validity measures, and indeed may be useful to SSA in the disability determination process; however, these measures are considered outside the scope of the committee and this report. Figure 4-1 delineates between psychological (or non-cognitive) self-report measures and nonpsychological self-report measures.

PSYCHOLOGICAL SELF-REPORT MEASURES AND DISABILITY EVALUATION

As discussed in Chapter 3, psychological assessment generally begins with a referral question followed by a clinical interview, the purpose of which is to explore presenting complaints (self-report of symptoms) and develop an understanding of the case, which may include a history of symptom development and an assessment of current status and impact on daily functioning. From this understanding, the next steps typically include the

[1] Note that when the committee refers to non-cognitive measures, it is referring to standardized psychological self-report measures.

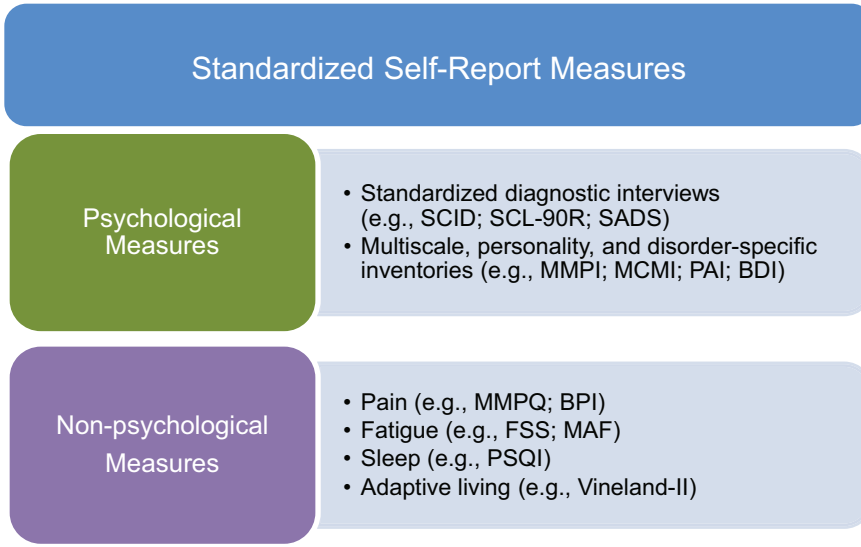


FIGURE 4-1 Psychological versus nonpsychological self-report measures.
NOTE: BDI = Beck Depression Inventory; BPI = Brief Pain Inventory; FSS = Fatigue Severity Scale; MAF = Multidimensional Assessment of Fatigue; MCMI = Millon Clinical Multiaxial Inventory; MMPI = Minnesota Multiphasic Personality Inventory; MMPQ = McGill-Melzack Pain Questionnaire; PAI = Personality Assessment Inventory; PSQI = Pittsburgh Sleep Quality Index; SADS = Schedule for Affective Disorders and Schizophrenia; SCID = Structured Clinical Interview for DSM Disorders; SCL-90R = Symptom Checklist 90 Revised; Vineland-II = Vineland Adaptive Behavior Scales.

identification of hypotheses to be examined and postulation of methods to assess these hypotheses. The primary goal of such methods is to provide corroborative evidence for the presenting complaints and their integration into case understanding. This may include the longitudinal history (which may provide evidence of internal consistency, such as refractoriness to treatment, chronicity, and severity); objective medical evaluation; direct observation of the claimant; and information from third parties such as family members, employers, and teachers. The use of non-cognitive measures may be another source of corroborative information, with the potential to inform the existence of a medically determinable impairment and/or functional limitations. Because of the potential for gain associated with disability determinations, a systematic method for assessing the validity of claims based primarily on self-report would prove valuable. In some cases, the use of non-cognitive psychological testing may contribute to achieving these goals.

Areas of Symptom Complaint

In the realm of disability evaluation, the committee identified two primary areas of impairment in which psychological self-report measures may prove beneficial to SSA disability determinations: mental disorders and somatic symptoms disproportionate to demonstrable medical morbidity. Each of these are discussed in turn, followed by a discussion on the ability of psychological self-report measures to provide useful information in confirming a medically determinable impairment and assessing functional capacity in these areas. A variety of non-cognitive measures, such as multiscale personality measures, disorder-specific inventories, and standardized diagnostic interviews, are provided as illustrative examples, and not an endorsement of any specific test.

Mental Disorders

Within its mental health listings, SSA (n.d.-a) identifies nine diagnostic categories (see Chapter 3, Table 1). Of these nine, the committee identified five categories for which non-cognitive measures may provide useful information: (1) schizophrenic, paranoid, and other psychotic disorders; (2) affective disorders; (3) anxiety-related disorders; (4) personality disorders; and (5) somatoform disorders.[2] Box 4-2 contains the SSA descriptions of each of the first four mental disorders categories.

These categories of mental disorders are well-established psychiatric diagnoses with distinct diagnostic criteria. In clinical settings, diagnosis in these categories often relies on self-report of symptoms, which are then weighed against criteria in the *Diagnostic and Statistical Manual of the American Psychiatric Association* (*DSM-5*). However, the method for assessing symptom report may vary, from a simple, unstructured clinical interview to more systematic approaches, such as the use of standardized psychiatric diagnostic schedules and interviews or formal psychological self-report measures. The use of such systematic approaches may help corroborate and validate a patient's symptom report.

There are also 11 mental disorder diagnostic categories listed by SSA specifically for children. The structure and organization of these categories is parallel to mental disorder listings shown for adults. The categories that contain conditions typically first diagnosed in childhood contain intellectual disability, autistic disorder and other pervasive developmental disorders, and attention deficit hyperactivity disorder. In addition, conduct disorder and oppositional defiant disorder are contained in the SSA listing for personality disorders.

[2] Although somatoform disorders are included in the SSA mental health listings, the committee focuses on these in the next section on disproportionate somatic symptoms, alongside multisystem illnesses and chronic idiopathic pain conditions.

BOX 4-2
SSA Definitions of Relevant Mental Disorders

Schizophrenic, paranoid, and other psychotic disorders	Characterized by the onset of psychotic features with deterioration from a previous level of functioning.
Affective disorders	Characterized by a disturbance of mood, accompanied by a full or a partial manic or depressive syndrome. Mood refers to a prolonged emotion that colors the whole psychic life; it generally involves either depression or elation.
Anxiety-related disorders	In these disorders anxiety is either the predominant disturbance or it is experienced if the individual attempts to master symptoms; for example, confronting the dreaded object or situation in a phobic disorder or resisting the obsessions or compulsions in obsessive compulsive disorders.
Personality disorders	A personality disorder exists when personality traits are inflexible and maladaptive and cause either significant impairment in social or occupational functioning or subjective distress. Characteristic features are typical of the individual's long-term functioning and are not limited to discrete episodes of illness.

SOURCE: SSA, n.d.-a.

Similar to those listed for adults, mental disorders present in childhood are well-established conditions listed in the *DSM-5* (American Psychiatric Association, 2013). These conditions are diagnosed in clinical settings based on report of symptoms, often by parents or others who interact with the child (e.g., teachers), as well as behavioral observations and the completion of standardized or systematic approaches, such as questionnaires, tests, and age-appropriate self-report instruments. Many conditions diagnosed in children are reevaluated when a child reaches majority age.

Disproportionate Somatic Symptoms

The committee identified three distinct groups of applicants seeking disability compensation for somatic symptoms unaccompanied by demonstrable

anatomical, biochemical, or physiological abnormalities: somatoform disorders (recently termed *somatic symptom disorders* in the *DSM-5*), multisystem illnesses, and chronic idiopathic pain conditions. Brief descriptions of these disorders are provided in Box 4-3.

Somatoform (or somatic symptom) disorders are diagnosable psychiatric disorders with distinct, well-elaborated diagnostic criteria (American Psychiatric Association, 2013); as such, they are among the listed mental disorders that are eligible for SSA disability compensation. These disorders appear to be medical disorders because their clinical presentation is characterized by somatic or physical symptoms, but on further examination they are best understood and treated as psychiatric conditions. They include somatic symptom disorder (formerly termed *somatization disorder*), hypochondriasis or illness anxiety disorder, and conversion disorder. These diagnoses require clinically significant and persistent bodily symptoms and a substantial degree of associated distress and functional impairment.

Multisystem illnesses (also termed *functional somatic syndromes*) share a common, nonspecific symptom pool, that includes fatigue, weakness, lightheadedness, dizziness, sleep difficulties, headache, problems of memory and attention, blurry vision, gastrointestinal complaints (e.g., heartburn,

BOX 4-3
Definitions of Relevant Disorders with Disproportionate Somatic Symptoms

Somatoform disorders[a]	Physical symptoms for which there are no demonstrable organic findings or known physiological mechanisms.
Multisystem illnesses[b]	Characterized by multiple, widespread, nonspecific, often diffuse symptoms that involve several different organ systems and anatomical locations, for which no consistent biochemical, anatomical, or physiological abnormality can be demonstrated. Hence the medical and psychiatric status of these conditions remains unclear.
Chronic idiopathic pain conditions[c]	The only or predominant symptom is bodily pain, most commonly musculoskeletal pain, that is disproportionate to (incompletely explained by) tissue injury or disease.

[a] American Psychiatric Association, 2013.
[b] Barsky and Borus, 1999; Henningsen et al., 2007.
[c] Vranceanu et al., 2009.

bloating), palpitations, shortness of breath, sore throats, and urinary frequency. Chronic fatigue syndrome, repetitive strain injury, toxic building syndrome, multiple chemical sensitivity, and chronic Lyme disease are among these conditions. Other apparently related illnesses include interstitial cystitis, chronic whiplash (cervical hyperextension), multiple food allergies, and hypoglycemia. These conditions are considered together as a group because they appear to share a number of characteristics: the same individual over time is frequently diagnosed with more than one of these conditions; they share extensive phenomenological overlap and common epidemiological characteristics; there is a higher than expected prevalence of psychiatric comorbidity; and they are marked by a refractoriness to the usual symptomatic medical treatments and standard palliative measures (Barsky and Borus, 1999; Henningsen et al., 2007).

The only or predominant symptom of *chronic idiopathic pain disorders* is bodily pain, most commonly musculoskeletal pain, that is disproportionate to (incompletely explained by) tissue injury or disease (Vranceanu et al., 2009). These conditions account for a large fraction of all disability payments; musculoskeletal pain accounts for 25 to 35 percent of adult disability claims. Low back pain is one of the most common single sources of disability compensation, but other pain conditions in which pain may be disproportionate to medical findings include fibromyalgia, complex regional pain syndrome, carpal tunnel syndrome, and temporomandibular joint disorder. There is often an acute precipitating injury or illness or procedure, after which the individual experiences chronic, intense, and severe pain that impairs their physical and role functioning.

Confirming the Existence of a Disability

As noted above, a disability determination requires a *medically determinable impairment* that affects an applicant's *ability to function* in a work setting. Such a determination must be confirmed with observable signs and laboratory findings. Included among acceptable laboratory findings are psychological tests (20 CFR § 404.1528).

Standardized non-cognitive measures are developed, interpreted, and evaluated in accordance with *psychometrics*, the scientific study of tests and measures used to assess variability in behavior and link such variability to psychological phenomena. Psychometrics also considers measurement theory (e.g., classical test theory and item response theory) and its applicability to measures. In evaluating the quality of psychological measures, psychometrics is primarily concerned with test reliability (i.e., consistency) and validity (i.e., accuracy).[3] Therefore, standardized psychological self-report

[3] See Chapter 3 for an in-depth discussion on psychometrics.

measures that demonstrate good psychometric properties can provide scientific laboratory findings that corroborate self-report of psychological symptoms.

The systematic use of standardized psychological self-report measures can help identify and document the presence and severity of a medically determinable impairment in each of the areas outlined above. Broad personality and multiscale inventories can provide medical evidence of a wide variety of mental disorders. The most prominent example of such measures is the Minnesota Multiphasic Personality Inventory (MMPI) (Hathaway and McKinley, 1940, 1943), along with more recent editions. The instrument was originally created more than 70 years ago and has been through two normative revisions. The MMPI, MMPI-2 (Butcher et al., 1989), and MMPI-2RF (Ben-Porath et al., 2008) all consist of a self-report inventory of symptoms and personal characteristics. Items are statements for which the test-taker responds in a dichotomous fashion (i.e., True/False) as the content applies to his or her own functioning. The current version of this assessment, the MMPI-2RF, comprises 338 items that are part of 51 different scales and was normed on a U.S. population (n = 2,227) of men and women ages 18–80. Other widely used multiscale inventories include the Millon Clinical Multiaxial Inventory (MCMI-III) (Millon et al., 2009) and the Personality Assessment Inventory (PAI) (Morey, 2007). The MCMI-III is a 175-item test normed largely on individuals seeking psychiatric services. The PAI contains 344 items and was developed on a U.S. normative sample of 1,000 adults matched to the census; additionally, 1,265 patients and 1,051 college students completed the test in the standardization process.

Standardized psychiatric diagnostic schedules, interviews, and inventories may also provide scientific medical findings across a broad range of psychiatric symptoms and diagnoses. The Symptom Check-List 90 Revised (SCL-90R) (Derogatis, 1994), a broad-based measure designed for individuals 13 years and older, contains a list of symptoms commonly associated with psychological difficulties and psychiatric disorders. Written at a sixth-grade level, the test measures nine primary symptom dimensions (i.e., somatization, obsessive-compulsive disorder, interpersonal sensitivity, depression, anxiety, hostility, phobic anxiety, paranoid ideation, and psychoticism), assessing symptom presence and frequency and severity across a 1-week period of time. There is also a 53-item version of the scale, the Brief Symptom Inventory (BSI) (Derogatis and Spencer, 1993). Designed specifically to measure subjective symptom report, the SCL-90R has separate norms for nonpatient adults, adult psychiatric outpatients, adult psychiatric inpatients, and nonpatient adolescents. Some reviewers suggest that this instrument is best used to screen for global psychological distress, as the individual symptom dimensions have not always been identified in studies examining the psychometric properties of the scale. Another broad

symptom inventory, the Patient Health Questionnaire (PHQ) (Spitzer et al., 1999), was developed for use in primary care settings and normed against this population. From the original test, scales to measure symptoms of depression (PHQ-9), anxiety (GAD-7), and somatic symptom severity (PHQ-15) have been constructed, along with a derivate scale, the PHQ-SADS that measures convergence of psychiatric symptoms often seen in primary care patients: depression, anxiety, and somatic complaints.

Many disorder-specific scales, such as the Beck Depression Inventory, second edition (Beck et al., 1996), Hamilton Depression Rating Scale (Hamilton, 1980), Beck Anxiety Inventory (Beck and Steer, 1993), and PTSD (posttraumatic stress disorder) Checklist (Weathers et al., 1994) may also provide medical evidence to corroborate patients' identification and report of symptoms.

Confirming the diagnosis of disproportionate somatic symptoms may be more difficult, as the first step involves ruling out the presence of demonstrable anatomical, biochemical, or physiological abnormalities as the sole cause for symptom presentation and severity. Note that this does not rule out the existence of such abnormalities, but that reported symptom severity is disproportionate to the diagnosis. Additionally, the lack of a medical explanation does not automatically equal a psychiatric diagnosis (American Psychiatric Association, 2013). There are a variety of self-report questionnaires to assess somatization and somatoform disorders, which examine the number, nature, intensity, persistence, and severity of physical symptoms. These instruments include the PHQ-15, the somatization subscale of the SCL-90R, the Somatic Symptom Inventory (SSI), and the MMPI-2-RF. There are also several structured diagnostic interviews containing modules for diagnosing somatoform disorders, including the Composite International Diagnostic Interview (CIDI) (WHO, 1993), the Structured Clinical Interview for DSM (SCID) (First et al., 2012; Gibbon et al., 1997), the Mini International Neuropsychiatric Interview (MINI) (Sheehan et al., 1998), and the Schedule for Clinical Assessment in Neuropsychiatry (SCAN) (Wing et al., 1990).

There are a great many self-report inventories for assessing the severity, character, location, and chronicity of pain; the nonpsychological nature of such measures place them outside of the committee's scope. However, there are non-cognitive measures that are used to identify and assess psychological factors related to pain, such as the Pain Patient Profile (P-3) (Tollison and Langley, 1995), which comprises three clinical scales measuring depression, anxiety, and somatization.

The second criterion in disability determinations is the impact of the medically determinable impairment on the applicant's ability to function in a work setting, what SSA refers to as the *Paragraph B criteria*. In the realm of mental disorders, SSA currently assesses functioning in four categories:

(1) activities of daily living (ADLs); (2) social functioning; (3) concentration, persistence, or pace; and (4) episodes of decompensation. However, SSA (2010) published a Notice of Proposed Rulemaking (NPRM)[4] for its mental disorders listings, which among other changes, would alter the functional categories on which disability determinations would be based, increasing focus on the relation of functioning to the work setting. Proposed functional domains in the NPRM are the abilities to (1) understand, remember, and apply information; (2) interact with others; (3) concentrate, persist, and maintain pace; and (4) manage oneself.[5] Definitions of each of these domains are presented in Box 4-4. With SSA's move in this direction and the greater focus on functional abilities as they relate to work, the committee will examine the relevance of psychological self-report measures to the proposed functional domains.

Although non-cognitive assessments do not provide direct evidence of functional capacity, information obtained from these measures allows for the corroboration of symptoms as presented, which can lead to greater diagnostic accuracy. For example, self-report instruments allow for a standardized method of obtaining information that is normed against other clinical and nonclinical groups, adding to the ability of a clinician to offer accurate diagnoses. In addition, some of these instruments have validity scales, which measure test-taking strategies, as discussed in detail below. Understanding these presentation approaches (i.e., over- or underreporting of symptoms) is helpful in identifying conditions accurately. From obtaining an accurate diagnosis, the ability to generate more accurate prognostic indicators increases and thereby provides greater ability to discern the chronicity of conditions presented.

ADMINISTRATION AND INTERPRETATION OF NON-COGNITIVE PSYCHOLOGICAL MEASURES

One of the most important aspects of administration of non-cognitive measures is selection of the appropriate measures to be administered. That

[4] Public comments are still under review and a final rule has yet to be published as of the publication of this report.

[5] These proposed domains align closely with the recommendations of the Mental Cognitive Subcommittee of the Occupational Information Development Advisory Panel (OIDAP), which conceptualized psychological abilities essential to work in four categories: (1) neurocognitive functioning, (2) initiative and persistence, (3) interpersonal functioning, and (4) self-management. Note that with this first category, neurocognitive functioning, the Mental Cognitive Subcommittee's recommendation goes into greater detail; this will be discussed further in the following chapter, which focuses on cognitive testing. The Mental Cognitive Subcommittee was assembled to advise OIDAP about what psychological abilities of disability applicants should be included in the Content Model and Classification Recommendations made to SSA.

BOX 4-4
SSA Proposed Functional Domains

Understand, remember, and apply information	The ability to acquire, retain, integrate, access, and use information to perform work activities. You use this mental ability when, for example, you follow instructions, provide explanations, and identify and solve problems.
Interact with others	The ability to relate to and work with supervisors, co-workers, and the public. You use this mental ability when, for example, you cooperate, handle conflicts, and respond to requests, suggestions, and criticism.
Concentrate, persist, and maintain pace	The ability to focus attention on work activities and to stay on task at a sustained rate. You use this mental ability when, for example, you concentrate, avoid distractions, initiate and complete activities, perform tasks at an appropriate and consistent speed, and sustain an ordinary routine.
Manage oneself	The ability to regulate your emotions, control your behavior, and maintain your well-being in a work setting. You use this mental ability when, for example, you cope with your frustration and stress, respond to demands and changes in your environment, protect yourself from harm and exploitation by others, inhibit inappropriate actions, take your medications, and maintain your physical health, hygiene, and grooming.

SOURCE: SSA, 2010.

is, selection of measures is dependent on examination of the normative data collected with each measure and consideration of the population on which the test was normed. Normative data are typically gathered on generally healthy individuals who are free from significant mental impairments. Data are generally gathered on samples that reflect the broad demographic characteristics of the United States including factors such as age, gender, and educational status. There are some measures that also provide specific comparison data on the basis of race and ethnicity.

As discussed in detail in Chapter 3, the use of psychological testing requires the examiner to follow standardized procedures for the

administration of the tests. Administration instructions for non-cognitive measures are contained in the respective test manuals. Although unique to each test, an overarching concern is the selection of a test for which there have been procedures developed for the characteristics of the person being examined. For example, the majority of non-cognitive measures require that the individual be able to complete a self-report inventory, a task that requires reading and responding to a list of dichotomous (e.g., True/False) or Likert scale items. To complete a task like this, one must have the ability to attend, read, comprehend, and respond to a series of items. For example, the MMPI-2-RF was developed with a fifth-grade reading level, while the MCMI-3 and the PAI both require an eighth-grade reading level. Although some tests have alternative methods of administration (e.g., standardized audio tape administration, computerized administration), ensuring that the examinee is able to understand information at a content level equivalent to the items on the test and has the capacity to attend to and respond to items is generally recommended. In addition, the capacity of the individual to work on an activity with similar characteristics for the development of normative data must be considered. Additionally, consideration of the examinee's language and administration of a test that has been translated and normed within the language is generally recommended.

SSA requires psychological testing be "individually administered by a qualified specialist," defining *qualified* as "currently licensed or certified in the state to administer, score, and interpret psychological tests and have the training and experience to perform the test" (SSA, n.d.-a). It is important to note here, as discussed in Chapter 3, the different qualification levels that may be necessary for administration and interpretation. It is common practice for psychometrists or technicians with specialized training to administer and score psychological tests, under the close supervision and direction of doctoral-level clinical psychologists. Interpretation of testing results requires a higher degree of clinical training than administration alone. Most psychological tests require interpretation by doctoral-level psychologists with a high level of expertise in psychometric test administration and interpretation.[6] Threats to the validity of any psychological measure of a self-report nature oblige the test interpreter to understand the test and principles of test construction. In fact, interpreting tests results without such knowledge would violate the ethics code established for the profession of psychology (APA, 2010). Finally, it is important for the person interpreting the test results to address in the assessment report the reliability and validity of test scores and test norms relative to the individual being assessed.

[6] These are commonly referred to as level C tests. Some tests have less stringent qualifications (level B) or no special qualifications (level A) necessary for purchase, administration, and interpretation. See Chapter 3 for additional information on different qualification levels.

ASSESSING THE VALIDITY OF NON-COGNITIVE SYMPTOM REPORT

Because much of psychological assessment relies heavily on self-report, assessing the accuracy of symptomatic complaint, or *symptom validity,* is critical. Symptom validity may be assessed in a number of ways. For example, an examinee's self-report may be evaluated alongside data from a number of outside sources, such as behavioral observations, interviews with corroborative sources (e.g., family members, friends, teachers), and review of historical records (e.g., medical, educational, occupational, legal), or a formal analysis of internal data consistency. Symptomatic complaint may also be considered against typical diagnostic considerations, such as onset, symptom presentation, course, and response to treatment (Heilbronner et al., 2009). And, as presented in this chapter, formal non-cognitive psychological testing can provide scientific evidence that may support a patient's self-report; however, as these measures also rely on self-report, assessing their validity is necessary. For this reason, formal SVTs exist to objectively assess the validity of data obtained during psychological assessment.

The initial step in interpreting results on self-report measures or questionnaires is to examine protocol validity. Multiple threats to validity are possible on most self-report measures. These threats include item responses that are not content based, such as omissions of items, provision of more than one response per item, or random responding. Such response styles may occur for a variety of reasons, for example, limited ability to read and process information, random human error (e.g., mismarking the answer sheet), or confusion or thought disorganization. Alternatively, invalid item responding may be content based, depending on the test-taker's motivations. While unintentional random response may be due to confusion and thought disorganization, content-based response patterns are thought to be due to defensiveness or other characteristics on the part of the test-taker. Content-based response threats occur when the test-taker intentionally skews his or her approach to responding to items and presents an impression that may or may not be convergent with his or her true characteristics. Such a response style may include exaggeration by intentionally over-reporting symptoms, which may occur in settings where there are benefits to being seen as impaired. For this reason, an examination of the measure's protocol validity scales is often undertaken.

Many of the self-report measures discussed in this chapter contain formal measures of the credibility and consistency of examinee response. These SVTs are measures used to assess whether an examinee is providing an accurate or consistent report of his or her actual symptom experience (Larrabee, 2014). Such tests have recently been distinguished from performance validity tests (PVTs) (Bigler, 2012; Larrabee, 2012; Van Dyke et al.,

2013), which assess whether a test-taker is attempting to perform at a level consistent with his or her actual abilities and generally focus on measures of cognition; such tests will be examined in Chapter 5. SVTs are constructed to assess the accuracy of the test-taker's responses on non-cognitive measures. Ultimately, such tests provide information on the interpretability and usefulness of results obtained from psychological tests and measures.

SVTs use a variety of approaches to examine response patterns that affect the accuracy of self-report on non-cognitive measures, which generally fall into three broad categories: consistency of response, negative self-presentation, and positive self-presentation. Consistency of response generally refers to whether a test-taker responds in a fixed or a random fashion or answers similar pairs of items in the same way. SVTs assess negative self-presentation in a variety of ways. Often, test-takers are presented with questions about infrequent or unlikely behaviors or symptoms; SVTs look for patterns of overreporting or amplification on these items, as compared to some population (e.g., general, psychiatric for mental complaints, medical patients for somatic complaints). For example, these measures generally contain items to which an individual is asked to respond with respect to concerns or symptoms, such as, "I have difficulty remembering what I had for breakfast" or "I see things around me that others do not see." There are diagnostic conditions for which an endorsement of either of these individual items would be appropriate. However, many scales use items that are conceptually divergent, minimizing the likelihood of multiple items being endorsed, even if a diagnosis is present. Positive self-presentation is assessed in a similar fashion, but generally examines underreporting or minimization of symptoms or difficulties in an attempt to assert better psychological adjustment. An example of an item in this category might be "I never missed a day of school due to being ill." While possible, the likelihood of positively endorsing multiple items when the scale consists of low base-rate behaviors is not high.

Scores on SVTs are typically generated by a summation of items and conversion to generate a standardized total score. Total scores are then compared to established cut-off scores, based on normative data on the scale. Norms may be based on nationally representative samples or subpopulations of relevance to the particular patient concern. For example, the MMPI-2-RF contains a validity scale that compares reports of emotional distress and psychiatric illness with psychiatric populations (i.e., Infrequent Psychopathology Responses [Fp-r]) and another that compares reporting of somatic complaints with medical patient populations (i.e., Infrequent Somatic Responses [Fs]). Norms may also include specific diagnostic groups that illuminate particular profiles on the test that may be indicative of a particular diagnosis. Cut-off scores are established to identify the presence of a response set that is either incongruent with known diagnoses or suggestive

of responding employing an alternative response set (e.g., overendorsement of symptoms). Such response sets are commonly seen as invalid and dependent on the test. The scale(s) are interpreted using clinical judgment by the examiner taking into consideration the referral questions, history of the examinee, and context of the evaluation.

Types of SVTs

Many SVTs are scales within larger personality or multiscale inventories assessing test-taker response styles used in completing the battery. These scales may be designed as such and embedded or later derived from existing items and scales based on typical response patterns, including those of specific populations. For example, each of the personality measures discussed earlier in this chapter (i.e., MMPI-2-RF, MCMI-III, and PAI) contains validity scales that examine consistency of response, negative self-presentation, and positive self-presentation to varying degrees. Box 4-5 lists the negative self-presentation SVTs included in each of these measures.

Though fewer in number, stand-alone SVTs also exist to assess potential exaggeration or feigning of psychological and neuropsychological symptoms. These include a number of structured interviews, such as the Structured Interview of Reported Symptoms (Rogers et al., 1992), the Structured Inventory of Malingered Symptomatology (Widows and Smith, 2005), and the Miller Forensic Assessment of Symptom Test (Miller, 2001). Like the embedded/derived measures, these SVTs examine accuracy of symptom report in a variety of ways. As this is their sole purpose, they are often used in conjunction with other measures that do not contain tests of validity. Box 4-6 lists the scales related to negative self-presentation in stand-alone SVTs.

Symptom Validity and the Disability Determination Process

When an applicant's medical record is based primarily on self-report, assessment of symptom validity helps the evaluator assess the accuracy of an individual's self-report of behavior, experiences, or symptoms. For this reason, it is important to include an assessment of symptom validity in the medical evidence of record. Such assessment may include the analysis of internal data consistency, examination of corroborative evidence, and formal SVTs.

There has been strong advocacy for the assessment of symptom validity—including the use of SVTs when administering non-cognitive measures—in forensic contexts in which examinees may be more likely to exaggerate symptoms. Organizations such as the Association for Scientific Advancement in Psychological Injury and Law (ASAPIL) (Bush et al.,

BOX 4-5
Embedded/Derived SVTs for Negative Self-Presentation

MMPI-2-RF[a]	
Infrequent Responses (F-r)	Overreporting across psychological, cognitive, and somatic dimensions (as compared with general population)
Infrequent Psychopathology Responses (Fp-r)	Overreporting of emotional distress and psychiatric illness (as compared with psychiatric populations)
Infrequent Somatic Responses (Fs)	Overreporting of somatic complaints (as compared with medical patient populations)
Symptom Validity (FBS-r)	Overreporting of somatic and cognitive complaints
Response Bias (RBS)	Overreporting of memory complaints
Henry-Heilbronner Index[b]	Physical symptom exaggeration (empirically derived from existing scales; for use with personal injury litigants and disability claimants)
Malingered Mood Disorder Scale[c]	Exaggeration of emotional disturbance (empirically derived from existing scales; for use with personal injury litigants and disability claimants)
MCMI-III[d]	
Validity (V)	Improbable symptoms; may measure confusion, difficulties reading and understanding items, or responding in a random fashion
Disclosure (X)	Acknowledgment of difficulties and willingness to present with symptoms
Debasement (Z)	Tendency to present symptoms in an accentuated fashion
PAI[e]	
Infrequency (INF)	Statistically unlikely response patterns in items that have low rates of endorsement and high rates of endorsement
Negative Impression (NIM)	Rare symptoms and those that are not reported by many respondents
Malingering Index (MAL)	Unlikely patterns; features that are more likely to be found in persons simulating mental disorders than in clinical patients
Rogers Discriminant Function (RDF)	A statistically determined method that distinguishes simulators from those who were responding honestly

[a] Ben-Porath et al., 2008.
[b] Henry et al., 2013.
[c] Henry et al., 2008.
[d] Millon et al., 2009.
[e] Morey, 2007.

BOX 4-6
Stand-Alone SVTs for Negative Self-Presentation

The 172-item *Structured Interview of Reported Symptoms* (*SIRS-2*)[a] evaluates feigning of psychiatric symptoms and deliberate distortions (e.g., exaggeration of symptom severity) in the self-report of symptoms. The inventory comprises a number of scales that produce information on how the examinee may distort his or her symptoms:

- Rare Symptoms (RS)
- Symptom Combinations (SC)
- Improbable and Absurd Symptoms (IA)
- Blatant Symptoms (BL)
- Subtle Symptoms (SU)
- Selectivity of Symptoms (SEL)
- Severity of Symptoms (SEV)
- Reported versus Observed symptoms (RO)

The 75-item *Structured Inventory of Malingered Symptomatology* (*SIMS*)[b] is a true/false screening instrument that assesses for both malingered psychopathology and neuropsychological symptoms. The inventory comprises five scale domains as well as an overall score for probable malingering (i.e., total score):

- Psychosis (P)
- Neurologic Impairment (NI)
- Amnestic Disorders (AM)
- Low Intelligence (LI)
- Affective Disorders (AF)

The 25-item *Miller Forensic Assessment of Symptoms Test* (*M-FAST*)[c] is a screening interview used to provide preliminary information regarding the possibility that an examinee is feigning psychopathology. The interview comprises seven scales corresponding to response styles and strategies related to feigning:

- Reported Versus Observed Symptoms
- Extreme Symptomatology
- Rare Combinations
- Unusual Hallucinations
- Unusual Symptom Course
- Negative Image
- Suggestibility

[a] Rogers et al., 1992.
[b] Widows and Smith, 2005.
[c] Miller, 2001.

2014), the American Academy of Clinical Neuropsychology (AACN) (Heilbronner et al., 2009), and the National Academy of Neuropsychology (NAN) (Bush et al., 2005) recommend the assessment of validity of self-report through a multimethod approach. This may include examination of consistency among self-report, test data, real-world activities, and historical records and administration of multiple SVTs throughout the evaluation. When there exists consistent evidence of invalid responding, AACN recommends that results of the inventory not be interpreted and data from other instruments without validity scales not be relied upon (Heilbronner et al., 2009, p. 1102). ASAPIL recommends reporting such concerns without "assumptions regarding examinee goals which underlie the production of invalid results" (Bush et al., 2014, p. 202). All three organizations recommend that other factors, such as culture, language, and functional limitations, also be considered when assessing validity.

Although administration of self-report measures is foundational in the field of psychology, requiring administration of SVTs in all disability claims is not a position with unequivocal supporting evidence. Administration of SVTs as part of the psychological evaluation battery can be helpful; however, interpretation of SVT data in the context of the non-cognitive testing must be undertaken carefully. Any SVT result can only be interpreted in an individual's personal context, including psychological/emotional history, level of intellectual functioning, and other factors that may affect responding. This is true for all testing and the interpretation of test results. Particular attention must be paid to the limitations of the normative and validation data available for each SVT. As such, a simple inter-individual interpretation of SVT results is not acceptable or valid. Additionally, as discussed in Chapter 3, a qualified test user is responsible for all aspects of appropriate test use; this includes understanding the normative and validation data, potential limitations, and appropriate interpretation of any SVT, whether embedded or stand-alone. Evidence of inconsistent self-report based on SVTs is cause for concern with regard to self-reported symptoms; however, it does not provide information about whether or not the individual is, in fact, disabled. As such, failure on SVTs alone is insufficient grounds for denying a disability claim.

The challenge is in determining how best to proceed when one or more SVTs indicate overreporting of symptoms on self-report measures. In such cases, self-report measures administered during the evaluation will likely yield little meaningful information; additional information will therefore be required to assess the applicant's allegation of disability. Additionally, because SVTs are used to help assess the validity of an individual's responses on standardized non-cognitive psychological measures, the administration of SVTs outside of that assessment cannot provide information about the validity of evidence already in the medical evidence record.

USE OF NON-COGNITIVE MEASURES WITH SPECIFIC POPULATIONS

As suggested above, there are a number of allegations that may warrant the administration of non-cognitive tests. Such allegations generally fall in two broad categories: mental disorders and disorders with somatic complaints that are disproportionate to demonstrable medical morbidity. Mental disorders include schizophrenic, paranoid, and other psychotic disorders; affective disorders; anxiety-related disorders; and personality disorders. It is important to note that some of these conditions may also include cognitive complaints, in which case cognitive testing (discussed in Chapter 5) may be more appropriate. Disorders with somatic complaints that are disproportionate to demonstrable medical morbidity include somatoform disorders, multisystem illnesses (e.g., chronic fatigue syndrome, repetitive strain injury, chronic Lyme disease), and chronic idiopathic pain conditions (e.g., fibromyalgia, carpal tunnel syndrome).

The committee concludes that the use of standardized non-cognitive psychological measures is essential to the determination of all cases in which an applicant's allegation of non-cognitive functional impairment meets three requirements:

- The applicant alleges a mental disorder (i.e., schizophrenic, paranoid, and other psychotic disorders; affective disorders; anxiety-related disorders; and personality disorders) unaccompanied by cognitive complaints or a disorder with somatic symptoms that are disproportionate to demonstrable medical morbidity (i.e., somatoform disorders, multisystem illnesses, and chronic idiopathic pain conditions).
- The presence and severity of impairment and associated functional limitations are based largely on applicant self-report.
- Objective medical evidence or longitudinal medical records sufficient to make a disability determination do not accompany the claim.

As noted above, when cognitive complaints accompany the applicant's allegations, cognitive testing may prove more appropriate. The committee also recognizes that some chronic conditions may generate potentially disabling, non-cognitive functional impairments but may not be accompanied by objective medical evidence (i.e., medical signs and/or laboratory or test results that constitute clear evidence of a significant mental disorder and related functional impairment of sufficient severity to make a disability determination). In such cases, the evidence provided by longitudinal medical records (i.e., a documented history of a significant mental disorder or

a chronic condition such as chronic idiopathic pain or multisystem illnesses and related functional impairment of sufficient severity and duration to make a disability determination) may be sufficient to substantiate the allegation.

When the medical evidence of record primarily relies on self-report of symptoms, a statement regarding the validity of results obtained in the assessment is essential. As noted above, a variety of methods for objectively assessing validity exist that go beyond the clinical opinion of the examiner. In addition to analysis of the results of SVTs administered at the time of the testing and analysis of internal data consistency, evidence could include a pattern of test results that is inconsistent with the alleged condition, observed behavior, documented history, and the like. It is important to note that a finding of inconsistency between the test results and the areas specified is more informative than a finding of consistency would be. Determination of the method or methods used to assess validity is best left to the discretion of a qualified evaluator.

REFERENCES

American Psychiatric Association. 2013. *The diagnostic and statistical manual of mental disorders: DSM-5*. Washington, DC: American Psychiatric Association.

APA (American Psychological Association). 2010. *Ethical principles of psychologists and code of conduct*. http://www.apa.org/ethics/code (accessed March 9, 2015).

Barsky, A. J., and J. F. Borus. 1999. Functional somatic syndromes. *Annals of Internal Medicine* 130(11):12.

Beck, A., and R. Steer. 1993. *Beck Anxiety Inventory manual*. San Antonio, TX: Harcourt Brace & Company.

Beck, A. T., R. Steer, and G. Brown. 1996. *Beck Depression Inventory*. 2nd ed. San Antonio, TX: The Psychological Corporation.

Ben-Porath, Y. S., A. Tellegen, and N. Pearson. 2008. *MMPI-2-RF: Manual for administration, scoring, and interpretation*. Minneapolis, MN: University of Minnesota Press.

Bigler, E. D. 2012. Symptom validity testing, effort, and neuropsychological assessment. *Journal of the International Neuropsychological Society* 18(4):632-642.

Bush, S. S., R. M. Ruff, A. I. Tröster, J. T. Barth, S. P. Koffler, N. H. Pliskin, C. R. Reynolds, and C. H. Silver. 2005. Symptom validity assessment: Practice issues and medical necessity. NAN policy and planning committee. *Archives of Clinical Neuropsychology* 20(4):419-426.

Bush, S. S., R. L. Heilbronner, and R. M. Ruff. 2014. Psychological assessment of symptom and performance validity, response bias, and malingering: Official position of the Association for Scientific Advancement in Psychological Injury and Law. *Psychological Injury and Law* 7(3):197-205.

Butcher, J. N., W. Dahlstrom, J. Graham, A. Tellegen, and B. Kaemmer. 1989. *MMPI-2: Manual for administration and scoring*. Minneapolis, MN: University of Minnesota Press.

Derogatis, L. 1994. SCL-90-R: Symptom Checklist-90-R. Minneapolis, MN: Pearson.

Derogatis, L. R., and P. Spencer. 1993. *Brief Symptom Inventory: BSI*. Minneapolis, MN: Pearson.

First, M. B., R. L. Spitzer, M. Gibbon, and J. B. Williams. 2012. *Structured Clinical Interview for DSM-IV axis I disorders (SCID-I), clinician version, administration booklet*. Arlington, VA: American Psychiatric Publishing.

Gibbon, M., R. L. Spitzer, and M. B. First. 1997. *User's guide for the Structured Clinical Interview for DSM-IV axis II personality disorders: SCID-II*. Arlington, VA: American Psychiatric Publishing.

Hamilton, M. 1980. Rating depressive patients. *Journal of Clinical Psychiatry* 41(12):21-24.

Hathaway, S. R., and J. C. McKinley. 1940. A multiphasic personality schedule (Minnesota): I. Construction of the schedule. *Journal of Psychology* 10:249-254.

Hathaway, S. R., and J. C. McKinley. 1943. *Manual for the Minnesota Multiphasic Personality Inventory*. New York: The Psychological Corporation.

Heilbronner, R. L., J. J. Sweet, J. E. Morgan, G. J. Larrabee, S. R. Millis, and 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.

Henningsen, P., S. Zipfel, and W. Herzog. 2007. Management of functional somatic syndromes. *Lancet* 369(9565):946-955.

Henry, G. K., R. L. Heilbronner, W. Mittenberg, C. Enders, and D. M. Roberts. 2008. Empirical derivation of a new MMPI-2 scale for identifying probable malingering in personal injury litigants and disability claimants: The 15-item Malingered Mood Disorder Scale (MMDS). *The Clinical Neuropsychologist* 22(1):158-168.

Henry, G. K., R. L. Heilbronner, J. Algina, and Y. Kaya. 2013. Derivation of the MMPI-2-RF Henry-Heilbronner Index-r (HHI-r) scale. *The Clinical Neuropsychologist* 27(3):509-515.

Larrabee, G. J. 2012. Performance validity and symptom validity in neuropsychological assessment. *Journal of the International Neuropsychological Society* 18(4):625-630.

Larrabee, G. J. 2014. *Performance and Symptom Validity*. Presentation to IOM Committee on Psychological Testing, Including Validity Testing, for Social Security Administration, June 25, 2014, Washington, DC.

Miller, H. A. 2001. *M-FAST: Miller forensic assessment of symptoms test professional manual*. Odessa, FL: Psychological Assessment Resources.

Millon, T., C. Millon, R. D. Davis, and S. Grossman. 2009. *Millon Clinical Multiaxial Inventory-III (MCMI-III) manual*. San Antonio, TX: Pearson/PsychCorp.

Morey, L. C. 2007. *Personality Assessment Inventory*. Odessa, FL: Psychological Assessment Resources.

Rogers, R., R. M. Bagby, and S. E. Dickens. 1992. *Structured Interview of Reported Symptoms: Professional manual*. Odessa, FL: Psychological Assessment Resources.

Sheehan, D., Y. Lecrubier, K. Sheehan, P. Amorim, J. Janavs, E. Weiller, T. Hergueta, R. Baker, and G. Dunbar. 1998. The Mini-International Neuropsychiatric Interview (MINI): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *Journal of Clinical Psychiatry* 59(20):22-33.

Spitzer, R. L., K. Kroenke, J. B. Williams, and P. H. Q. P. C. S. Group. 1999. Validation and utility of a self-report version of PRIME-MD: The PHQ primary care study. *JAMA* 282(18):1737-1744.

SSA (Social Security Administration). 2010. Revised medical criteria for evaluating mental disorders. *Federal Register* 75(160):34.

SSA. n.d.-a. *Disability evaluation under social security—Part III: Listing of impairments—Adult listings (Part A)—section 12.00 mental disorders*. http://www.ssa.gov/disability/professionals/bluebook/12.00-MentalDisorders-Adult.htm (accessed November 14, 2014).

SSA. n.d.-b. *Disability evaluation under Social Security: Part I—general information*. http://www.ssa.gov/disability/professionals/bluebook/general-info.htm (accessed November 14, 2014).

Tollison, D., and J. Langley. 1995. *Pain Patient Profile manual*. Minneapolis, MN: National Computer Systems.

Van Dyke, S. A., S. R. Millis, B. N. Axelrod, and R. A. Hanks. 2013. Assessing effort: Differentiating performance and symptom validity. *The Clinical Neuropsychologist* 27(8): 1234-1246.

Vranceanu, A., A. Barsky, and D. Ring. 2009. Psychosocial aspects of diabling musculoskeletal pain. *Journal of Bone and Joint Surgery* 91(8):2014-2018.

Weathers, F., B. Litz, D. Herman, J. Huska, and T. Keane. 1994. *The PTSD checklist-civilian version (PCL-C)*. Boston, MA: National Center for PTSD.

WHO (World Health Organization). 1993. *Composite International Diagnostic Interview (CIDI): Interviewer's manual*. Geneva, Switzerland: WHO.

Widows, M. R., and G. P. Smith. 2005. *Structured Inventory of Malingered Symptomatology: Professional manual*. Lutz, FL: Psychological Assessment Resources.

Wing, J. K., T. Babor, T. Brugha, J. Burke, J. Cooper, R. Giel, A. Jablenski, D. Regier, and N. Sartorius. 1990. SCAN: Schedules for Clinical Assessment in Neuropsychiatry. *Archives of General Psychiatry* 47(6):589-593.

5

Cognitive Tests and Performance Validity Tests

Disability determination is based in part on signs and symptoms of a disease, illness, or impairment. When physical symptoms are the presenting complaint, identification of signs and symptoms of illnesses are relatively concrete and easily obtained through a general medical exam. However, documentation or concrete evidence of cognitive or functional impairments, as may be claimed by many applying for disability,[1] is more difficult to obtain.

Psychological testing may help inform the evaluation of an individual's functional capacity, particularly within the domain of cognitive functioning. The term *cognitive functioning* encompasses a variety of skills and abilities, including intellectual capacity, attention and concentration, processing speed, language and communication, visual-spatial abilities, and memory. Sensorimotor and psychomotor functioning are often measured alongside neurocognitive functioning in order to clarify the brain basis of certain cognitive impairments, and are therefore considered as one of the domains that may be included within a neuropsychological or neurocognitive evaluation. These skills and abilities cannot be evaluated in any detail without formal standardized psychometric assessment.

This chapter examines cognitive testing, which relies on measures of task performance to assess cognitive functioning and establish the severity of cognitive impairments. As discussed in detail in Chapter 2, a determination of disability requires both a medically determinable impairment and

[1] As documented in Chapters 1 and 2, 57 percent of claims fall under mental disorders other than intellectual disability and/or connective tissue disorders.

evidence of functional limitations that affect an individual's ability to work. A medically determinable impairment must be substantiated by symptoms, signs, and laboratory findings (the so-called Paragraph A criteria) and the degree of functional limitations imposed by the impairment must be assessed in four broad areas: activities of daily living; social functioning; concentration, persistence, or pace; and episodes of decompensation (the so-called Paragraph B criteria). However, as discussed in Chapter 4, the U.S. Social Security Administration (SSA) is in the process of altering the functional domains, through a Notice of Proposed Rulemaking published in 2010.[2] The proposed functional domains—understand, remember, and apply information; interact with others; concentrate, persist, and maintain pace; and manage oneself—increase focus on the relation of functioning to the work setting; because of SSA's move in this direction, the committee examines the relevance of psychological testing in terms of these proposed functional domains. As will be discussed below, cognitive testing may prove beneficial to the assessment of each of these requirements.

ADMINISTRATION OF COGNITIVE AND NEUROPSYCHOLOGICAL TESTS TO EVALUATE COGNITIVE IMPAIRMENT

In contrast to testing that relies on self-report, as outlined in the preceding chapter, evaluating cognitive functioning relies on measures of task performance to establish the severity of cognitive impairments. Such tests are commonly used in clinical neuropsychological evaluations in which the goal is to identify a patient's pattern of strengths and weaknesses across a variety of cognitive domains. These performance-based measures are standardized instruments with population-based normative data that allow the examiner to compare an individual's performance with an appropriate comparison group (e.g., those of the same age group, sex, education level, and/or race/ethnicity).

Cognitive testing is the primary way to establish severity of cognitive impairment and is therefore a necessary component in a neuropsychological assessment. Clinical interviews alone are not sufficient to establish the severity of cognitive impairments, for two reasons: (1) patients are known to be poor reporters of their own cognitive functioning (Edmonds et al., 2014; Farias et al., 2005; Moritz et al., 2004; Schacter, 1990) and (2) clinicians relying solely on clinical interviews in the absence of neuropsychological test results are known to be poor judges of patients' cognitive functioning (Moritz et al., 2004). There is a long history of

[2] Public comments are currently under review and a final rule has yet to be published as of the publication of this report.

neuropsychological research linking specific cognitive impairments with specific brain lesion locations, and before the advent of neuroimaging, neuropsychological evaluation was the primary way to localize brain lesions; even today, neuropsychological evaluation is critical for identifying brain-related impairments that neuroimaging cannot identify (Lezak et al., 2012). In the context of the SSA disability determination process, cognitive testing for claimants alleging cognitive impairments could be helpful in establishing a medically determinable impairment, functional limitations, and/or residual functional capacity.

The use of standardized psychological and neuropsychological measures to assess residual cognitive functioning in individuals applying for disability will increase the credibility, reliability, and validity of determinations on the basis of these claims. A typical psychological or neuropsychological evaluation is multifaceted and may include cognitive and non-cognitive assessment tools. Evaluations typically consist of a (1) clinical interview, (2) administration of standardized cognitive or non-cognitive psychological tests, and (3) professional time for interpretation and integration of data. Some neuropsychological tests are computer administered, but the majority of tests in use today are paper-and-pencil tests.

The length of an evaluation will vary depending on the purpose of the evaluation and, more specifically, the type or degree of psychological and/or cognitive impairments that need to be evaluated. A national professional survey of 1,658 neuropsychologists from the membership of American Academy of Clinical Neuropsychology (AACN), Division 40 of American Psychological Association (APA), and the National Academy of Neuropsychologists (NAN) indicated that a typical neuropsychological evaluation takes approximately 6 hours, with a range from 0.5 to 25 hours (Sweet et al., 2011). The survey also identified a number of reasons for why the duration of an evaluation varies, including reason for referral, the type or degree of psychological and/or cognitive impairments, or factors specific to the individual.

The most important aspect of administration of cognitive and neuropsychological tests is selection of the appropriate tests to be administered. That is, selection of measures is dependent on examination of the normative data collected with each measure and consideration of the population on which the test was normed. Normative data are typically gathered on generally healthy individuals who are free from significant cognitive impairments, developmental disorders, or neurological illnesses that could compromise cognitive skills. Data are generally gathered on samples that reflect the broad demographic characteristics of the United States including factors such as age, gender, and educational status. There are some measures that also provide specific comparison data on the basis of race and ethnicity.

As discussed in detail in Chapter 3, as part of the development of any psychometrically sound measure, explicit methods and procedures by which

tasks should be administered are determined and clearly spelled out. All examiners use such methods and procedures during the process of collecting the normative data, and such procedures normally should be used in any other administration. Typical standardized administration procedures or expectations include (1) a quiet, relatively distraction-free environment; (2) precise reading of scripted instructions; and (3) provision of necessary tools or stimuli. Use of standardized administration procedures enables application of normative data to the individual being evaluated (Lezak et al., 2012). Without standardized administration, the individual's performance may not accurately reflect his or her ability. An individual's abilities may be overestimated if the examiner provides additional information or guidance than what is outlined in the test administration manual. Conversely, a claimant's abilities may be underestimated if appropriate instructions, examples, or prompts are not presented.

Cognitive Testing in Disability Evaluation

To receive benefits, claimants must have a *medically determinable physical or mental impairment*, which SSA defines as

> an impairment that results from anatomical, physiological, or psychological abnormalities which can be shown by medically acceptable clinical and laboratory diagnostic techniques ... [and] must be established by medical evidence consisting of signs, symptoms, and laboratory findings—not only by the individual's statement of symptoms. (SSA, n.d.-b)

To qualify at Step 3 in the disability evaluation process (as discussed in Chapter 2), there must be medical evidence that substantiates the existence of an impairment and associated functional limitations that meet or equal the medical criteria codified in SSA's *Listings of Impairments*. If an adult applicant's impairments do not meet or equal the medical listing, *residual functional capacity*—the most a claimant can still do despite his or her limitations—is assessed; this includes whether the applicant has the capacity for past work (Step 4) or any work in the national economy (Step 5). For child applicants, once there has been identification of a medical impairment, documentation of a "marked and severe functional limitation relative to typically developing peers" is required. Cognitive testing is valuable in both child and adult assessments in determining the existence of a medically determinable impairment and evaluating associated functional impairments and residual functional capacity.

Cognitive impairments may be the result of intrinsic factors (e.g., neurodevelopmental disorders, genetic factors) or be acquired through injury or illness (e.g., traumatic brain injury, stroke, neurological conditions) and may occur at any stage of life. Functional limitations in cognitive domains

may also result from other mental or physical disorders, such as bipolar disorder, depression, schizophrenia, psychosis, or multiple sclerosis (Etkin et al., 2013; Rao, 1986).

Cognitive Domains Relevant to SSA

SSA currently assesses mental residual functional capacity by evaluating 20 abilities in four general areas: understanding and memory, sustained concentration and persistence, social interaction, and adaptation (see Form SSA-4734-F4-SUP: Mental Residual Functional Capacity [MRFC] Assessment). Through this assessment, a claimant's ability to sustain activities that require such abilities over a normal workday or workweek is determined.

In 2009, SSA's Occupational Information Development Advisory Panel (OIDAP) created its Mental Cognitive Subcommittee "to review mental abilities that can be impaired by illness or injury, and thereby impede a person's ability to do work" (OIDAP, 2009, p. C-3). In their report, the subcommittee recommended that the conceptual model of psychological abilities required for work, as currently used by SSA through the MRFC assessment, be revised to redress shortcomings and be based on scientific evidence. The subcommittee identified four major categories of psychological functioning essential to work: neurocognitive functioning, initiative and persistence, interpersonal functioning, and self-management, recommending that "SSA adopt 15 abilities that represent specific aspects of the[se] four general categories." Within neurocognitive functioning, the testing of which is the primary focus of the current chapter, the subcommittee identified six relevant domains: general cognitive/intellectual ability, language and communication, memory acquisition, attention and distractibility, processing speed, and executive functioning; "each of the constituent abilities has been found to predict either the ability to work or level of occupational attainment among persons with various mental disorders and/or healthy adults" (OIDAP, 2009, p. C-22). Building on the subcommittee's report, the current Institute of Medicine (IOM) committee has adopted these six domains of cognitive functioning for its examination of cognitive testing in disability determinations.

Each of these functional domains would also be relevant areas of assessment in children applying for disability support. As indicated below, there are standardized measures that have been well normed and validated for pediatric populations. Interpretation of test results in children is more challenging, as it must take into account the likelihood of developmental progress and response to any interventions. Thus, the permanency of cognitive impairments identified in childhood is more difficult to ascertain in a single evaluation.

There are numerous performance-based tests that can be used to assess an individual's level of functioning within each domain identified below for both adults and children. It was beyond the scope of this committee and report to identify and describe each available standardized measure; thus, only a few commonly used tests are provided as examples for each domain. The choice of examples should not be seen as an attempt by the committee to identify or prescribe tests that should be used to assess these domains within the context of disability determinations. Rather, the committee believed that it was more appropriate to identify the most relevant domains of cognitive functioning and that it remains in the purview of the appropriately qualified psychological/neuropsychological evaluator to select the most appropriate measure for use in specific evaluations. For a more comprehensive list and review of cognitive tests, readers are referred to the comprehensive textbooks, *Neuropsychological Assessment* (Lezak et al., 2012) or *A Compendium of Neuropsychological Tests* (Strauss et al., 2006).

General Cognitive/Intellectual Ability

General cognitive/intellectual ability encompasses reasoning, problem solving, and meeting cognitive demands of varying complexity. It has been identified as "the most robust predictor of occupational attainment, and corresponds more closely to job complexity than any other ability" (OIDAP, 2009, p. C-21). Intellectual disability affects functioning in three domains: conceptual (e.g., memory, language, reading, writing, math, knowledge acquisition); social (e.g., empathy, social judgment, interpersonal skills, friendship abilities); and practical (e.g., self-management in areas such as personal care, job responsibilities, money management, recreation, organizing school and work tasks) (American Psychiatric Association, 2013, p. 37). Tests of cognitive/intellectual functioning, commonly referred to as intelligence tests, are widely accepted and used in a variety of fields, including education and neuropsychology. Prominent examples include the Wechsler Adult Intelligence Scale, fourth edition (WAIS-IV; Wechsler, 2008) and the Wechsler Intelligence Scale for Children, fourth edition (WISC-IV; Wechsler, 2003).

Language and Communication

The domain of language and communication focuses on receptive and expressive language abilities, including the ability to understand spoken or written language, communicate thoughts, and follow directions (American Psychiatric Association, 2013; OIDAP, 2009). The *International Classification of Functioning, Disability and Health* (WHO, 2001) distinguishes the two, describing language in terms of mental functioning while

describing communication in terms of activities (the execution of tasks) and participation (involvement in a life situation). The mental functions of language include reception of language (i.e., decoding messages to obtain their meaning), expression of language (i.e., production of meaningful messages), and integrative language functions (i.e., organization of semantic and symbolic meaning, grammatical structure, and ideas for the production of messages). Abilities related to communication include receiving and producing messages (spoken, nonverbal, written, or formal sign language), carrying on a conversation (starting, sustaining, and ending a conversation with one or many people) or discussion (starting, sustaining, and ending an examination of a matter, with arguments for or against, with one or more people), and use of communication devices and techniques (telecommunications devices, writing machines) (WHO, 2001). In a survey of historical governmental and scholarly data, Ruben (1999) found that communication disorders were generally associated with higher rates of unemployment, lower social class, and lower income.

A wide variety of tests are available to assess language abilities; some prominent examples include the Boston Naming Test (Kaplan et al., 2001), Controlled Oral Word Association (Benton et al., 1994a; Spreen and Strauss, 1991), the Boston Diagnostic Aphasia Examination (Goodglass and Kaplan, 1983), and for children, the Clinical Evaluation of Language Fundamentals-4 (Semel et al., 2003) or Comprehensive Assessment of Spoken Language (Carrow-Woolfolk, 1999). There are fewer formal measures of communication per se, although there are some educational measures that do assess an individual's ability to produce written language samples, for example, the Test of Written Language (Hammill and Larsen, 2009).

Learning and Memory

This domain refers to abilities to register and store new information (e.g., words, instructions, procedures) and retrieve information as needed (OIDAP, 2009; WHO, 2001). Functions of memory include "short-term and long-term memory; immediate, recent and remote memory; memory span; retrieval of memory; remembering; [and] functions used in recalling and learning" (WHO, 2001, p. 53). However, it is important to note that semantic, autobiographical, and implicit memory are generally preserved in all but the most severe forms of neurocognitive dysfunction (American Psychiatric Association, 2013; OIDAP, 2009). Impaired memory functioning can arise from a variety of internal or external factors, such as depression, stress, stroke, dementia, or traumatic brain injury (TBI), and may affect an individual's ability to sustain work, due to a lessened ability to learn and remember instructions or work-relevant material. Examples of tests for learning and memory deficits include the Wechsler Memory

Scale (Wechsler, 2009), Wide Range Assessment of Memory and Learning (Sheslow and Adams, 2003), California Verbal Learning Test (Delis, 1994; Delis et al., 2000), Hopkins Verbal Learning Test-Revised (Benedict et al., 1998; Brandt and Benedict, 2001), Brief Visuospatial Memory Test-Revised (Benedict, 1997), and the Rey-Osterrieth Complex Figure Test (Rey, 1941).

Attention and Vigilance

Attention and vigilance refers to the ability to sustain focus of attention in an environment with ordinary distractions (OIDAP, 2009). Normal functioning in this domain includes the ability to sustain, shift, divide, and share attention (WHO, 2001). Persons with impairments in this domain may have difficulty attending to complex input, holding new information in mind, and performing mental calculations. They may also exhibit increased difficulty attending in the presence of multiple stimuli, be easily distracted by external stimuli, need more time than previously to complete normal tasks, and tend to be more error prone (American Psychiatric Association, 2013). Tests for deficits in attention and vigilance include a variety of continuous performance tests (e.g., Conners Continuous Performance Test, Test of Variables of Attention), the WAIS-IV working memory index, Digit Vigilance (Lewis, 1990), and the Paced Auditory Serial Addition Test (Gronwall, 1977).

Processing Speed

Processing speed refers to the amount of time it takes to respond to questions and process information, and "has been found to account for variability in how well people perform many everyday activities, including untimed tasks" (OIDAP, 2009, p. C-23). This domain reflects mental efficiency and is central to many cognitive functions (NIH, n.d.). Tests for deficits in processing speed include the WAIS-IV processing speed index and the Trail Making Test Part A (Reitan, 1992).

Executive Functioning

Executive functioning is generally used as an overarching term encompassing many complex cognitive processes such as planning, prioritizing, organizing, decision making, task switching, responding to feedback and error correction, overriding habits and inhibition, and mental flexibility (American Psychiatric Association, 2013; Elliott, 2003; OIDAP, 2009). It has been described as "a product of the coordinated operation of various processes to accomplish a particular goal in a flexible manner" (Funahashi, 2001, p. 147). Impairments in executive functioning can lead to disjointed

and disinhibited behavior; impaired judgment, organization, planning, and decision making; and difficulty focusing on more than one task at a time (Elliott, 2003). Patients with such impairments will often have difficulty completing complex, multistage projects or resuming a task that has been interrupted (American Psychiatric Association, 2013). Because executive functioning refers to a variety of processes, it is difficult or impossible to assess executive functioning with a single measure. However, it is an important domain to consider, given the impact that impaired executive functioning can have on an individual's ability to work (OIDAP, 2009). Some tests that may assist in assessing executive functioning include the Trail Making Test Part B (Reitan, 1992), the Wisconsin Card Sorting Test (Heaton, 1993), and the Delis-Kaplan Executive Function System (Delis et al., 2001).

PSYCHOMETRICS AND TESTING NORMS FOR COGNITIVE TESTS

Once a test has been administered, assuming it has been done so according to standardized protocol, the test-taker's performance can be scored. In most instances, an individual's raw score, that is the number of items on which he or she responded correctly, is translated into a standard score based on the normative data for the specific measure. In this manner, an individual's performance can be characterized by its position on the distribution curve of normal performances.

The majority of cognitive tests have normative data from groups of people who mirror the broad demographic characteristics of the population of the United States based on census data. As a result, the normative data for most measures reflect the racial, ethnic, socioeconomic, and educational attainment of the population majorities. Unfortunately, that means that there are some individuals for whom these normative data are not clearly and specifically applicable. This does not mean that testing should not be done with these individuals, but rather that careful consideration of normative limitations should be made in interpretation of results.

Selection of appropriate measures and assessment of applicability of normative data vary depending on the purpose of the evaluation. Cognitive tests can be used to identify acquired or developmental cognitive impairment, to determine the level of functioning of an individual relative to typically functioning same-aged peers, or to assess an individual's functional capacity for everyday tasks (Freedman and Manly, 2015). Clearly, each of these purposes could be relevant for SSA disability determinations. However, each of these instances requires different interpretation and application of normative data.

When attempting to identify a change in functioning secondary to neurological injury or illness, it is most appropriate to compare an individual's postinjury performance to his or her premorbid level of functioning. Unfortunately, it is rare that an individual has a formal assessment of his or her premorbid cognitive functioning. Thus, comparison of the postinjury performance to demographically matched normative data provides the best comparison to assess a change in functioning (Freedman and Manly, 2015; Heaton et al., 2001; Manly and Echemendia, 2007). For example, assessment of a change in language functioning in a Spanish-speaking individual from Mexico who has sustained a stroke will be more accurate if the individual's performance is compared to norms collected from other Spanish-speaking individuals from Mexico rather than English speakers from the United States or even Spanish-speaking individuals from Puerto Rico. In many instances, this type of data is provided in alternative normative data sets rather than the published population-based norms provided by the test publisher.

In contrast, the population-based norms are more appropriate when the purpose of the evaluation is to describe an individual's level of functioning relative to same-aged peers (Busch, 2006; Freedman and Manly, 2015). A typical example of this would be in instances when the purpose of the evaluation is to determine an individual's overall level of intellectual (i.e., IQ) or even academic functioning. In this situation, it is more relevant to compare that individual's performance to that of the broader population in which he or she is expected to function in order to quantify his or her functional capabilities. Thus, for determination of functional disability, demographically or ethnically corrected normative data are inappropriate and may actually underestimate an individual's degree of disability (Freedman and Manly, 2015). In this situation, use of otherwise appropriate standardized and psychometrically sound performance-based or cognitive tests is appropriate.

Determination of an individual's everyday functioning or vocational capacity is perhaps the evaluation goal most relevant to the SSA disability determination process. To make this determination, the most appropriate comparison group for any individual would be other individuals who are currently completing the expected vocational tasks without limitations or disability (Freedman and Manly, 2015). Unfortunately, there are few standardized measures of skills necessary to complete specific vocational tasks and, therefore, also no vocational-specific normative data at this time. This type of functional capacity is best measured by evaluation techniques that recreate specific vocational settings and monitor an individual's completion of related tasks.

Until such specific vocational functioning measures exist and are readily available for use in disability determinations, objective assessment of cognitive skills that are presumed to underlie specific functions will be

necessary to quantify an individual's functional limitations. Despite limitations in normative data as outlined in Freedman and Manly (2015), formal psychometric assessment can be completed with individuals of various ethnic, racial, gender, educational, and functional backgrounds. However, the authors note that "limited research suggests that demographic adjustments reduce the power of cognitive test scores to predict every-day abilities" (e.g., Barrash et al., 2010; Higginson et al., 2013; Silverberg and Millis, 2009). In fact, they go on to state "the normative standard for daily functioning should not include adjustments for age, education, sex, ethnicity, or other demographic variables" (p. 9). Use of appropriate standardized measures by appropriately qualified evaluators as outlined in the following sections further mitigates the impact of normative limitations.

INTERPRETATION AND REPORTING OF TEST RESULTS

Interpretation of results is more than simply reporting the raw scores an individual achieves. Interpretation requires assigning some meaning to the standardized score within the individual context of the specific test-taker. There are several methods or levels of interpretation that can be used, and a combination of all is necessary to fully consider and understand the results of any evaluation (Lezak et al., 2012). This section is meant to provide a brief overview; although a full discussion of all approaches and nuances of interpretation is beyond the scope of this report, interested readers are referred to various textbooks (e.g., Groth-Marnat, 2009; Lezak et al, 2012).

Interindividual Differences

The most basic level of interpretation is simply to compare an individual's testing results with the normative data collected in the development of the measures administered. This level of interpretation allows the examiner to determine how typical or atypical an individual's performance is in comparison to same-aged individuals within the general population. Normative data may or may not be further specialized on the basis of race/ethnicity, gender, and educational status. There is some degree of variability in how an individual's score may be interpreted based on its deviation from the normative mean due to various schools of thought, all of which cannot be described in this text. One example of an interpretative approach would be that a performance within one standard deviation of the mean would be considered broadly average. Performances one to two standard deviations below the mean are considered mildly impaired, and those two or more standard deviations below the mean typically are interpreted as being at least moderately impaired.

Intraindividual Differences

In addition to comparing an individual's performances to that of the normative group, it also is important to compare an individual's pattern of performances across measures. This type of comparison allows for identification of a pattern of strengths and weaknesses. For example, an individual's level of intellectual functioning can be considered a benchmark to which functioning within some other domains can be compared. If all performances fall within the mildly to moderately impaired range, an interpretation of some degree of intellectual disability may be appropriate, depending on an individual's level of adaptive functioning. It is important to note that any interpretation of an individual's performance on a battery of tests must take into account that variability in performance across tasks is a normal occurrence (Binder et al., 2009) especially as the number of tests administered increases (Schretlen et al., 2008). However, if there is significant variability in performances across domains, then a specific pattern of impairment may be indicated.

Profile Analysis

When significant variability in performances across functional domains is assessed, it is necessary to consider whether or not the pattern of functioning is consistent with a known cognitive profile. That is, does the individual demonstrate a pattern of impairment that makes sense or can be reliably explained by a known neurobehavioral syndrome or neurological disorder. For example, an adult who has sustained isolated injury to the temporal lobe of the left hemisphere would be expected to demonstrate some degree of impairment on some measures of language and verbal memory, but to demonstrate relatively intact performances on measures of visual-spatial skills. This pattern of performance reflects a cognitive profile consistent with a known neurological injury. Conversely, a claimant who demonstrates impairment on all measures after sustaining a brief concussion would be demonstrating a profile of impairment that is inconsistent with research data indicating full cognitive recovery within days in most individuals who have sustained a concussion (McCrea et al., 2002, 2003).

Interpreting Poor Cognitive Test Performance

Regardless of the level of interpretation, it is important for any evaluator to keep in mind that poor performance on a set of cognitive or neuropsychological measures does not always mean that an individual is truly impaired in that area of functioning. Additionally, poor performance on a

set of cognitive or neuropsychological measures does not directly equate to functional disability.

In instances of inconsistent or unexpected profiles of performance, a thorough interpretation of the psychometric data requires use of additional information. The evaluator must consider the validity and reliability of the data acquired, such as whether or not there were errors in administration that rendered the data invalid, emotional or psychiatric factors that affected the individual's performance, or sufficient effort put forth by the individual on all measures.

To answer the latter question, administration of performance validity tests (PVTs) as part of the cognitive or neuropsychological evaluation battery can be helpful. Interpretation of PVT data must be undertaken carefully. Any PVT result can only be interpreted in an individual's personal context, including psychological/emotional history, level of intellectual functioning, and other factors that may affect performance. Particular attention must be paid to the limitations of the normative data available for each PVT to date. As such, a simple interindividual interpretation of PVT testing results is not acceptable or valid. Rather, consideration of intraindividual patterns of performance on various cognitive measures is an essential component of PVT interpretation. PVTs will be discussed in greater detail later in this chapter.

Qualifications for Administering Tests

Given the need for the use of standardized procedures, any person administering cognitive or neuropsychological measures must be well trained in standardized administration protocols. He or she should possess the interpersonal skills necessary to build rapport with the individual being tested in order to foster cooperation and maximal effort during testing. Additionally, individuals administering testing should understand important psychometric properties, including validity and reliability, as well as factors that could emerge during testing to place either at risk (as described in Chapter 3).

Many doctoral-level psychologists are well trained in test administration. In general, psychologists from clinical, counseling, school, or educational graduate psychology programs receive training in psychological test administration. However, the functional domains of emphasis in most of these programs include intellectual functioning, academic achievement, aptitude, emotional functioning, and behavioral functioning (APA, 2015). Thus, if the request for disability is based on a claim of intellectual disability or significant emotional/behavioral dysfunction, a psychologist with solid psychometric training from any of these types of graduate-level

training programs would typically be capable of completing the necessary evaluation.

For cases in which the claim is based on specific cognitive deficits, particularly those attributed to neurological disease or injury, a neuropsychologist may be needed to most accurately evaluate the claimant's functioning. *Neuropsychologists* are clinical psychologists

> trained in the science of brain-behavior relationships. The clinical neuropsychologist specializes in the application of assessment and intervention principles based on the scientific study of human behavior across the lifespan as it relates to normal and abnormal functioning of the central nervous system. (HNS, 2003)

That is, a neuropsychologist is trained to evaluate functioning within specific cognitive domains that may be affected or altered by injury to or disease of the brain or central nervous system. For example, a claimant applying for disability due to enduring attention or memory dysfunction secondary to a TBI would be most appropriately evaluated by a neuropsychologist.

The use of psychometrists or technicians in cognitive/neuropsychological test administration is a widely accepted standard of practice (Brandt and van Gorp, 1999). *Psychometrists* are often bachelor's- or master's-level individuals who have received additional specialized training in standardized test administration and test scoring. They do not practice independently, but rather work under the close supervision and direction of doctoral-level clinical psychologists.

Qualifications for Interpreting Test Results

Interpretation of testing results requires a higher degree of clinical training than administration alone. Most doctoral-level clinical psychologists who have been trained in psychometric test administration are also trained in test interpretation. As stated in the existing SSA (n.d.-a) documentation regarding evaluation of intellectual disability, the specialist completing psychological testing "must be currently licensed or certified in the state to administer, score, and interpret psychological tests and have the training and experience to perform the test." However, as mentioned above, the training received by most clinical psychologists is limited to certain domains of functioning, including measures of general intellectual functioning, academic achievement, aptitude, and psychological/emotional functioning. Again, if the request for disability is based on a claim of intellectual disability or significant emotional/behavioral dysfunction, a psychologist with solid psychometric training from any of these programs should be capable of providing appropriate interpretation of the testing

that was completed. The reason for the evaluation, or more specifically, the type of claim of impairment, may suggest a need for a specific type of qualification of the individual performing and especially interpreting the evaluation.

As stated in existing SSA (n.d.-a) documentation, individuals who administer more specific cognitive or neuropsychological evaluations "must be properly trained in this area of neuroscience." Clinical neuropsychologists, as defined above, are individuals who have been specifically trained to interpret testing results within the framework of brain-behavior relationships and who have achieved certain educational and training benchmarks as delineated by national professional organizations (AACN, 2007; NAN, 2001). More specifically, clinical neuropsychologists have been trained to interpret more complex and comprehensive cognitive or neuropsychological batteries that could include assessment of specific cognitive functions, such as attention, processing speed, executive functioning, language, visual-spatial skills, or memory. As stated above, interpretation of data involves examining patterns of individual cognitive strengths and weaknesses within the context of the individual's history including specific neurological injury or disease (i.e., claims on the basis of TBI).

ASSESSING VALIDITY OF COGNITIVE TEST PERFORMANCE

Neuropsychological tests assessing cognitive, motor, sensory, or behavioral abilities require actual performance of tasks, and they provide quantitative assessments of an individual's functioning within and across cognitive domains. The standardization of neuropsychological tests allows for comparability across test administrations. However, interpretation of an individual's performance presumes that the individual has put forth full and sustained effort while completing the tests; that is, accurate interpretation of neuropsychological performance can only proceed when the test-taker puts forth his or her best effort on the testing. If a test-taker is not able to give his or her best effort, for whatever reason, the test results cannot be interpreted as accurately reflecting the test-taker's ability level. As discussed in detail in Chapter 2, a number of studies have examined potential for malingering when there is a financial incentive for appearing impaired, suggesting anywhere from 19 to 68 percent of SSA disability applicants may be performing below their capability on cognitive tests or inaccurately reporting their symptoms (Chafetz, 2008; Chafetz et al., 2007; Griffin et al., 1996; Mittenberg et al., 2002). For a summary of reported base rates of "malingering," see Table 2-2 of this report and the ensuing discussion. However, an individual may put forth less than optimal effort due to a variety of factors other than malingering, such as pain, fatigue, medication use, and psychiatric symptomatology (Lezak et al., 2012).

For these reasons, analysis of the entire cognitive profile for consistency is generally recommended. Specific patterns that increase confidence in the validity of a test battery and overall assessment include

- Consistency between test behavior or self-reported symptoms and incidental behavior;
- Consistency between test behavior or self-reported symptoms and what is known about brain functioning and the type and severity of injury/illness claimed;
- Consistency between test behavior or self-reported symptoms and known patterns of performance (e.g., passing easy items and failing more difficult items; better performance on cued recall and recognition tests than free recall tests; intact memory requires intact attention);
- Consistency between test behavior or self-reported symptoms and reliable collateral reports or other background information, such as medical documentation;
- Consistency between self-reported history and reliable collateral history or medical documentation; and
- Consistency across tests measuring the same cognitive domain or across tests administered at different times.

Specific tests have also been designed especially to aid in the examination of performance validity. The development of and research on these PVTs has increased rapidly during the past two decades. There have been attempts to formally quantify performance validity during testing since the mid-1900s (Rey, 1964), with much of the initial focus on examining the consistency of an individual's responses across a battery of testing, with the suggestion that inconsistency may indicate variable effort. However, a significant push for specific formal measures came in response to the increased use of neuropsychological and cognitive testing in forensic contexts, including personal injury litigation, workers compensation, and criminal proceedings in the 1980s and 1990s (Bianchini et al., 2001; Larrabee, 2012a). Given the nature of these evaluations, there was often a clear incentive for an individual to exaggerate his or her impairment or to put forth less than optimal effort during testing, and neuropsychologists were being called upon to provide statements related to the validity of test results (Slick et al., 1999). Several studies documented that use of clinical judgment and interpretation of performance inconsistencies alone was an inadequate methodology for detection of poor effort or intentionally poor performance (Faust et al., 1988; Heaton et al., 1978; van Gorp et al., 1999). As such, the need for formal standardized measures of effort and means for interpretation of these measures emerged.

PVTs are measures that assess the extent to which an individual is providing valid responses during cognitive or neuropsychological testing. PVTs are typically simple tasks that are easier than they appear to be and on which an almost perfect performance is expected based on the fact that even individuals with severe brain injury have been found capable of good performance (Larrabee, 2012b). On the basis of that expectation, each measure has a performance cut-off defined by an acceptable number of errors designed to keep the false-positive rate low. Performances below these cut-off points are interpreted as demonstrating invalid test performance.

Types of PVTs

PVTs may be designed as such and embedded within other cognitive tests, later derived from standard cognitive tests, or designed as stand-alone measures. Examples of each type of measure are discussed below.

Embedded and Derived Measures

Embedded and derived PVTs are similar in that a specific score or assessment of response bias is determined from an individual's performance on an aspect of a preexisting standard cognitive measure. The primary difference is that embedded measures consist of indices specifically created to assess validity of performance in a cognitive test, whereas derived measures typically use novel calculations of performance discrepancies rather than simply examining the pattern of performance on already established indices. The rationale for this type of PVT is that it does not require administration of any additional tasks and therefore does not result in any added time or cost. Additionally, development of these types of PVTs can allow for retrospective consideration or examination of effort in batteries in which specific stand-alone measures of effort were not administered (Solomon et al., 2010).

The forced-choice condition of the California Verbal Learning Test—second edition (CVLT-II) (Delis et al., 2000) is an example of an embedded PVT. Following learning, recall, and recognition trials involving a 16-item word list, the test-taker is presented with pairs of words and asked to identify which one was on the list. More than 92 percent of the normative population, including individuals in their eighties, scored 100 percent on this test. Scores below the published cut-off are unusually low and indicative of potential noncredible performance. Scores below chance are considered to reflect purposeful noncredible performance, in that the test-taker knew the correct answer but purposely chose the wrong answer.

Reliable Digit Span, based on the Digit Span subtest of the Wechsler Adult Intelligence Scale, is an example of a measure that was derived based

on research following test publication. The Digit Span subtest requires test-takers to repeat strings of digits in forward order (forward digit span), as well as in reverse order (backward digit span). To calculate Reliable Digit Span, the maximum forward and backward span are summed, and scores below the cut-off point are associated with noncredible performance (Greiffenstein et al., 1994). A full list of embedded and derived PVTs is provided in Table 5-1.

Stand-Alone Measures

A stand-alone PVT is a measure that was developed specifically to assess a test-taker's effort or consistency of responses. That is, although the measure may appear to assess some other cognitive function (e.g., memory), it was actually developed to be so simple that even an individual with severe impairments in that function would be able to perform adequately. Such measures may be forced choice or non-forced choice (Boone and Lu, 2007; Grote and Hook, 2007).

The Test of Memory Malingering (TOMM) (Tombaugh and Tombaugh, 1996), the Word Memory Test (WMT) (Green et al., 1996), and the Rey Memory for Fifteen Items Test (RMFIT) (Rey, 1941) are examples of stand-alone measures of performance validity. As with many stand-alone measures, the TOMM, WMT, and RMFIT are memory tests that appear more difficult than they really are. The TOMM and WMT use a forced-choice method to identify noncredible performance in which the test-taker is asked to identify which of two stimuli was previously presented. Accuracy scores are compared to chance level performance (i.e., 50 percent correct), as well as performance by normative groups of head-injured and cognitively impaired individuals, with cut-offs set to minimize false-positive errors. Alternatively, the RMFIT uses a non-forced-choice method in which the test-taker is presented with a group of items and then asked to reproduce as many of the items as possible.

Forced-Choice PVTs

As noted above, some PVTs are forced-choice measures on which performance significantly below chance has been suggested to be evidence of intentionally poor performance based on application of the binomial theorem (Larrabee, 2012a). For example, if there are two choices, it would be expected that purely random guessing would result in 50 percent of items correct. Scores deviating from 50 percent in either direction indicate nonchance-level performance. The most probable explanation for substantially below-chance PVT scores is that the test-taker knew the correct answer but purposely selected the wrong answer. The Slick and colleagues

TABLE 5-1 Embedded and Derived PVTs

Test abbreviation	Test name	Source
ACS	Advanced Clinical Solutions	Holdnack and Drozdick (2009)
ACSS	Age-Corrected Scaled Score	Wechsler (1997a)
AVLT RMT	Rey Auditory Verbal Learning Test Recognition Memory Test	Binder, Villanueva, Howieson, and Moore (1993)
b-test	b-test	Boone, Lu, and Herzberg (2002[b])
BVMT-R	Brief Visuospatial Memory Test, Revised	Benedict (1997)
CVLT-II	California Verbal Learning Test, Second Edition	Delis, Kramer, Kaplan, and Ober (2000)
CVMT	Continuous Visual Memory Test	Trahan and Larrabee (1988)
DF	Discriminant Function	Mittenberg, Patton, and Legler (2003)
FTT	Finger Tapping Test	Heaton, Grant, and Matthews (1991)
HRB	Halstead-Reitan Battery	Reitan and Wolfson (1993)
LMR	Logical Memory Recognition	Killgore and DellaPietra (2000)
RAVLT	Rey Auditory Verbal Learning Test	Schmidt (1996)
RCFT	Rey Complex Figure Test	Meyers and Volbrecht (1999)
RBANS	Repeatable Battery For Assessment Of Neuropsychological Status	Randolph (1998)
RDS	Reliable Digit Span	Greiffenstein et al. (1994)
RDCT E-score	Rey Dot Counting Test	Rey (1941)
RMFIT	Rey 15-Item Memory Test	Rey (1941)
RMT	Recognition Memory Test	Warrington (1984)
ROCFT	Rey-Osterreith Complex Figure Test	Lu, Boone, Cozolino, and Mitchell (2003)
RWRT	Rey Word Recognition Test	Rey (1964)
SRT	Seashore Rhythm Test	Reitan and Wolfson (1993)
SSPT	Speech Sounds Perception Test	Reitan and Wolfson (1993)
VFDT	Visual Form Discrimination Test	Benton, de Hamsher, Varney, and Spreen (1983, 1994b)
WAIS-III	Wechsler Adult Intelligence Scale, Third Edition	Wechsler (1997)
WCST-FMS	Wisconsin Card Sorting Test, Failure-To-Maintain Set Score	Suhr and Boyer (1999)
WCT	Word Choice Test, in the WMS-IV	Wechsler (2009)
WMI	Working Memory Index	Wechsler (1997a)
WMS-III-VPA	Wechsler Memory Scale, Third Edition, Verbal Paired Associates-2 Scale Score	Wechsler (1997[b])

SOURCE: Young, 2014. Reproduced with permission.

TABLE 5-2 Forced-Choice PVTs

Test abbreviation	Test name	Source
ASTM	Amsterdam Short Term Memory Test	Jelicic, Merckelbach, Candel, and Geraets (2007)
CARB	Computerized Assessment of Response Bias Test	Allen, Conder, Green, and Cox (1997); Conder, Allen, and Cox (1992)
DMT	Digit Memory Test	Hiscock and Hiscock (1989)
FCTNA	Forced-Choice Test of Nonverbal Ability	Frederick and Foster (1991)
HDMT	Hiscock Digit Memory Test	Hiscock and Hiscock (1989)
MDMT	Multi-Digit Memory Test	Niccolls and Bolter (1991)
MPS	Malingering Probability Scale	Silverton (1999)
MSVT	Medical Symptom Validity Test	Green (2004)
NV-MSVT	Nonverbal Medical Symptom Validity Test	Green (2008)
PDRT	Portland Digit Recognition Test	Binder (1993), Binder and Willis (1991)
PDS	Paulhus Deception Scales	Paulhus (1998)
TOMM	Test of Memory Malingering	Tombaugh [and Tombaugh] (1996)
VIP	Validity Indicator Profile	Frederick (1997)
VSVT	Victoria Symptom Validity Test	Slick et al. (1997)
WMT	Word Memory Test	Green (2005)

SOURCE: Young, 2014. Reproduced with permission.

(1999) criteria for malingered neurocognitive dysfunction include below chance performance ($P < 0.05$) on one or more forced-choice measures of performance validity as indicative of malingering, and the authors state that "short of confession," below-chance performance on performance validity testing is "closest to an evidentiary 'gold standard' for malingering." Though below-chance performance on forced-choice PVTs implies intent, the committee believes it does not necessarily imply malingering, because the motivation of the performance may not be known; however, it does mean that the remainder of the test battery cannot be interpreted. A list of forced-choice PVTs can be found in Table 5-2.

Administration and Interpretation of PVTs

It is within that historical medicolegal context that clinical practice guidelines for neuropsychology emerged to emphasize the use of psychometric indicators of response validity (as opposed to clinician judgment alone) in determining the interpretability of a battery of cognitive tests

(Bianchini et al., 2001; Heilbronner et al., 2009). Moreover, it has become standard clinical practice to use multiple PVTs throughout an evaluation (Boone, 2009; Heilbronner et al., 2009). In general, multiple PVTs should be administered over the course of the evaluation because performance validity may wax and wane with increasing and decreasing fatigue, pain, motivation, or other factors that can influence effortful performance (Boone, 2009, 2014; Heilbronner et al., 2009). Some of the PVT development studies have attempted to examine these factors (i.e., effect of experimentally induced pain) and found no effect on PVT performance (Etherton et al., 2005a,b).

In clinical evaluations, most individuals will pass PVTs, and a small proportion will fail at the below-chance level. These clear passes can support the examiner's interpretation of the evaluation data being valid. Clear failures, that is below-chance performances, certainly place the validity of any other data obtained in the evaluation in question.

The risk of falsely identifying failure on one PVT as indicative of noncredible performance has resulted in the common practice of requiring failure on at least two PVTs to make any assumptions related to effort (Boone, 2009, 2014; Larrabee, 2014a). According to practice guidelines of NAN, performance slightly below the cut-off point on only one PVT cannot be construed to represent noncredible performance or biased responding; converging evidence from other indicators is needed to make a conclusion regarding performance bias (Bush et al., 2005). Similarly, AACN suggests the use of multiple validity assessments, both embedded and stand-alone, when possible, noting that effort may vary during an evaluation (Heilbronner et al., 2009). However, it should be noted that in cases where a test-taker scores significantly below chance on a single forced-choice PVT, intent to deceive may be assumed and test scores deemed invalid. It is also important to note that some situations may preclude the use of multiple validity indicators. For example, when evaluating an early school-aged child, at present, the TOMM is the only empirically established PVT (Kirkwood, 2014). In such situations, "it is the clinician's responsibility to document the reasons and explicitly note the interpretive implications" of reliance on a single PVT (Heilbronner et al., 2009).

The number of noncredible performances and the pattern of PVT failure are both considered in making a determination about whether the remainder of the neuropsychological battery can be interpreted. This consideration is particularly important in evaluations in which the test-taker's performance on cognitive measures falls below an expected level, suggesting potential cognitive impairment. That is, an individual's poor performance on cognitive measures may reflect insufficient effort to perform well, as suggested by PVT performance, rather than a true impairment. However, even in the context of PVT failure, performances that are in the average range

can be interpreted as reflecting ability that is in the average range or above, though such performances may represent an underestimate of actual level of ability. Certainly, PVT "failure" does not equate to malingering or lack of disability. However, clear PVT failures make the validity of the remainder of the cognitive battery questionable; therefore, no definitive conclusions can be drawn regarding cognitive ability (aside from interpreting normal performances as reflecting normal cognitive ability). An individual who fails PVTs may still have other evidence of disability that can be considered in making a determination; in these cases, further information would be needed to establish the case for disability.

AACN and NAN endorse the use of PVT measures in the context of any neuropsychological examination (Bush et al., 2005; Heilbronner et al., 2009). The practice standards require clinical neuropsychologists performing evaluations of cognitive functioning for diagnostic purposes to include PVTs and comment on the validity of test findings in their reports. There is no gold standard PVT, and use of multiple PVTs is recommended. A specified set of PVTs, or other cognitive measures for that matter, is not recommended due to concerns regarding test security and test-taker coaching.[3]

Caveats and Considerations in the Use of PVTs

Given the primary use of cut-off scores, even within the context of forced-choice tasks, the interpretation of PVT performance is inherently different than interpretation of performance on other standardized measures of cognitive functioning owing to the nature of the scores obtained. Unlike general cognitive measures that typically use a norm-referenced scoring paradigm assuming a normal distribution of scores, PVTs typically use a criterion-referenced scoring paradigm because of a known skewed distribution of scores (Larrabee, 2014a). That is, an individual's performance is compared to a cut-off score set to keep false-positive rates below 10 percent for determining whether or not the individual passed or failed the task.

A resulting primary critique of PVTs is that the development of the criterion or cut-off scores has not been as rigorous or systematic as is typically expected in the collection of normative data during development of a new standardized measure of cognitive functioning. In general, determination of what is an acceptable or passing performance and associated cut-off scores have been established in somewhat of a post hoc or retrospective fashion. However, there are some embedded PVTs that have been co-normed with

[3] At the committee's second meeting, Drs. Bianchini, Boone, and Larrabee all expressed great concern about the susceptibility of PVTs to coaching and stressed the importance of ensuring test security, as disclosure of test materials adversely affects the reliability and validity of psychological test results.

their "parent" tests, such as the forced-choice condition of the CVLT-II, which was normed along with the CVLT-II and thus has norms from the general population.

For most PVTs, however, rather than administering the measures to a large number of "typical" individuals of various ages, ethnicities, and even clinical diagnoses, researchers have examined the pattern of performance retrospectively in clinical samples that may have had some incentive to underperform (i.e., secondary gain), such as litigants (Roberson et al., 2013) or individuals presenting for consultative evaluations for Social Security disability determination (Chafetz, 2011; Chafetz and Underhill, 2013). An alternative methodology is to use simulation/nonsimulation samples in which one group of participants is told to perform poorly as if they had some type of impairment and the other is told to perform typically. Performances in these types of groups have then been used to establish cut-off scores via (1) identification of a fixed but arbitrary cut-off score of performance, or (2) identification of an "empirical floor" based on the lowest level of performance of a chosen clinical sample (the "known groups" approach, i.e., severely brain-injured patients) (Bianchini et al., 2001). One concern with this methodology is that data from simulators, especially data used to determine the sensitivity or specificity of a PVT, may not be applicable to real-world clinical samples (Boone et al., 2002a, 2005). In fact, few PVTs (other than some embedded PVTs such as CVLT-II Forced-Choice Recognition) have been normed on population-based samples or samples that are not biased in some way due to the method of recruitment (Freedman and Manly, 2015). Thus, the applicability or generalizability of cut-off scores to a broader (i.e., nonforensic) population is questionable.

As a result of this methodology, there are no true "traditional" normative data for many of these measures. However, the need for this type of normative data is minimal given the fact that the simple nature of tasks allows most patients with even severe brain injury, let alone "typical" individuals, to perform at near perfect levels (Larrabee, 2014a). Because of these skewed performance patterns, expectations for sensitivity and specificity for detection of poor performance have been developed rather than traditional norms (Greve and Bianchini, 2004).

Sensitivity in this context is defined as the degree to which a performance score on the measure will correctly identify an individual who is putting forth less than optimal effort. *Specificity* is the degree to which a performance score will correctly identify a person who is putting forth sufficient or optimal effort. Thus, to be most useful, ideally a PVT has high sensitivity and specificity. In general, however, most PVT cut-off scores are determined to have sensitivity within the 50–60 percent range and specificity within the 90–95 percent range. A meta-analysis of 47 studies by Sollman and Berry (2011) examined the sensitivity and specificity of five

stand-alone forced-choice PVTs, finding a mean sensitivity of 69 percent and mean specificity of 90 percent. However, the individual sensitivities and specificities of the measures varied (e.g., WMT sensitivity ranged from 49 percent to 100 percent and specificity ranged from 25 percent to 96 percent; TOMM sensitivity ranged from 34 percent to 100 percent and specificity ranged from 69 percent to 100 percent). There is general agreement among neuropsychologists that PVT specificity must be at least 90 percent for a PVT to be acceptable, in order to avoid falsely labeling valid performances as noncredible (Boone, 2007).

Sensitivity and specificity levels have been "verified" in experimental studies that employ comparison between groups that were expected to or told to perform well and those that were expected to or told to perform poorly. That is, researchers compared the performance on PVTs of groups of people "known" or expected to be performing poorly (i.e., those with clear secondary gain, those instructed to feign poor performance, or those who meet Slick and colleagues [1999] criteria for malingering) to those who perform well on PVTs or without clear secondary gain. Otherwise, studies have simply examined the pass/fail rates in clinical samples and the correlations of PVT performance with performance on the broader neuropsychological battery. There has been some comparison between the overall performance of subgroups who failed PVTs with the performance of the subgroup that did not, with the suggestion that those who fail PVTs tend to perform more poorly on testing overall. Although this methodology may appear to be more appropriate to the clinical situation, it still does not provide any indication of *why* an individual failed a PVT, which could be due to lack of effort or a variety of other factors, including true cognitive impairment (Freedman and Manly, 2015).

Although many would argue that PVT failure caused by true cognitive impairment is rare, the fact that failure could occur for valid reasons means that interpretation of PVT performances is exceptionally critical and must be done very cautiously. There are insufficient data related to the base-rate of below-chance performances on PVTs in different populations (Freedman and Manly, 2015). As Bigler (2012, 2014, 2015) points out, there are many individuals whose performances fall within a grey area, meaning they perform below the identified cut-off level but above chance. For example, individuals with multiple sclerosis, schizophrenia, TBI, or epilepsy have PVT failure rates of 11–30 percent in terms of falling below standard cut-off scores, even in the absence of known secondary gain (Hampson et al., 2013; Stevens et al., 2014; Suchy et al., 2012). Davis and Millis (2014) identified increased rates of PVT failure in individuals with lower educational status and lower functional status (i.e., independence in activities of daily living). Alternatively, others contend that concerns about grey area performance are unfounded, as the risk for false positives can be minimized, For example,

Boone (2009, 2014), Larrabee (2012, 2014a,b), and others assert that multiple PVT failures are generally required,[4] and as the number of PVT failures increase, the chance for a false positive approaches zero. Yet, it is possible that PVT failures (i.e., below cut-off score performance) in certain populations reflect legitimate cognitive impairments. For this reason, it has also been recommended that close attention be paid to the pattern of PVT performance and the potential for false positives in these at-risk populations in order to inform interpretation and reduce the chances for false positives (Larrabee, 2014a,b) and to inform future PVT research (Boone, 2007; Larrabee, 2007).

For these reasons, it is necessary to evaluate PVTs in the context of the individual disability applicant, including interpretation of the degree of PVT failure (e.g., below-chance performance versus performance slightly below cut-off score performance) and the consistency of failure across PVTs. Furthermore, careful interpretation of grey area PVT performance (significantly above chance but below standard cut-offs) is necessary, given that a significant proportion of individuals with bona fide mental or cognitive disorders may score in this "grey area." Adding to the complexity of interpreting these scores, population-based norms, and certainly norms for specific patient groups, are not available for most PVTs. Rather, owing to the process of development of these tasks, normative data exist only for select populations, typically litigants or those seeking compensation for injury. Thus, there are no norms for specific demographic groups (e.g., racial/ethnic minority groups). It has been suggested that examiners can compensate for these normative issues by using their clinical judgment to identify an alternate cut-off score for increased specificity (which will come at a cost of lower sensitivity) (Boone, 2014). For example, if an examiner identifies cultural, ethnic, and/or language factors known to affect PVT scores, the examiner should adjust his or her thresholds for identifying noncredible performance (Salazar et al., 2007).

Despite the practice standard of using multiple PVTs, there may be an increased likelihood of abnormal performances as the number of measures administered increases, a pattern that occurs in the context of standard cognitive measures (Schretlen et al., 2008). This type of analysis is beginning to be applied to PVTs specifically with inconsistent findings to date. Several studies examining PVT performance patterns in groups of clinical patients have indicated that it is very unlikely that an individual putting forth good effort on testing will fail two or more PVTs regardless of type of PVT (i.e., embedded or free-standing) (Iverson and Franzen, 1996; Larrabee, 2003). In fact, Victor and colleagues (2009) found a significant difference in the

[4] The exception being a single below-chance failure on a forced-choice PVT is sufficient to render scores invalid.

rate of failure on two or more embedded PVTs between those determined to be credible responders (5 percent failure) and noncredible responders (37 percent failure) in a clinical referral sample. Davis and Millis (2014) also found no predictive relation between the number of PVTs administered and the rate of PVT failure in a retrospective review of 158 consecutive referrals for evaluation. In contrast, others have utilized statistical modeling techniques to argue that there is an increased rate of false-positive PVT failures with increased number of PVTs administered (Berthelson et al., 2013; Bilder et al., 2014). Thus, ongoing careful interpretation of failure patterns is warranted.

Clinical use and research on PVT use in pediatric samples to date is significantly limited compared to that in adults. As such, specific pediatric criteria to determine pass/fail performances on PVTs do not exist. However, in general, the conclusion has been that children, even down to age 5 years, typically are able to pass most stand-alone measures of effort even when compared to the adult-based cut-off scores (DeRight and Carone, 2015). Despite these greater limitations in normative data, use of PVTs is becoming common practice even in pediatric patient samples. As in adults, children's performance on PVTs has been correlated with intellectual abilities (Gast and Hart, 2010; MacAllister et al., 2009), although even those with mildly impaired cognitive abilities have been able to pass stand-alone measures (Green and Flaro, 2003). Additionally, in samples of consecutive clinical referrals, failure on PVTs has not been associated with demographic, developmental disorders, or neurological status (Kirkwood et al., 2012). Even children with documented moderate to severe brain injury/dysfunction have been found to pass PVTs at the expected adult level (Carone, 2008). There are currently no studies examining PVT use with children younger than age five; however, research has shown that deception strategies at this age generally cannot be sustained and are fairly basic and obvious. As such, behavioral observations are important to assessing validity of cognitive testing with preschool-aged children (DeRight and Carone, 2015; Kirkwood, 2014).

APPLICANT POPULATIONS FOR WHOM PERFORMANCE-BASED TESTS SHOULD BE CONSIDERED OR USED

As suggested above, there are many applicants for whom administration of cognitive or neuropsychological testing would be beneficial to improve the standardization and credibility of determinations based on allegations of disability on the basis of cognitive impairment. The discussion below should not be considered all-inclusive, but rather as an attempt to highlight categories of disability applicants in which cognitive or performance-based testing would be appropriate.

Intellectual Disability

SSA has clear and appropriate standards for documentation for individuals applying for disability on the basis of intellectual disability (SSA, n.d.-a). As stated by SSA, "standardized intelligence test results are essential to the adjudication of all cases of intellectual disability" if the claimant does not clearly meet or equal the medical listing without. There are individual cases, of course, in which the claimant's level of impairment is so significant that it precludes formalized testing. For these individuals, their level of functioning and social history provides a longitudinal consistent record and documentation of impairment. For those who can complete intellectual testing and for whom their social history is inconsistent, inclusion of some documentation or assessment of effort may be warranted and would help to validate the results of intellectual and adaptive functioning assessment.

Use of PVTs is common among practitioners assessing for intellectual disability, with the TOMM being the most commonly used measure (Victor and Boone, 2007). However, caution is warranted in interpreting PVT results in individuals with intellectual disability, as IQ has consistently been correlated with PVT performance (Dean et al., 2008; Graue et al., 2007; Hurley and Deal, 2006; Shandera et al., 2010). More importantly, individuals with intellectual disability fail PVTs at a higher rate than those without (Dean et al., 2008; Salekin and Doane, 2009). In fact, Dean and colleagues (2008) found in their sample that all individuals with an IQ of less than 70 failed at least one PVT. Thus, cut-off scores for individuals with suspected intellectual disability may need to be adjusted due to a higher rate of false-positive results in this population. For example, lowering the TOMM Trial 2 and Retention Trial cut-off scores from 45 to 30 resulted in very low false-positive rates (0–4 percent) (Graue et al., 2007; Shandera et al., 2010).

Neurocognitive Impairments

There are individuals who apply for disability with primary allegations of cognitive dysfunction in one or more of the functional domains outlined above (e.g., "fuzzy" thinking, slowed thinking, poor memory, concentration difficulties). Standardized cognitive test results, as has been required for individuals claiming intellectual disability, are essential to the adjudication of such cases. These individuals may present with cognitive impairment due to a variety of reasons including, but not limited to, brain injury or disease (e.g., TBI or stroke) or neurodevelopmental disorders (e.g., learning disabilities, attention deficit hyperactivity disorder). Similarly, disability applicants may claim cognitive impairment secondary to a psychiatric disorder. For all of these claimants, documentation of impairment in functional cognitive domains with standardized cognitive tests is critically important. Within the

process of collection of test result evidence of these impairments, inclusion of some documentation or assessment of effort is warranted and would help to validate the results of intellectual and adaptive functioning assessment.

Medical Impairments Without Biological Basis

Use of PVTs is generally recommended in evaluations of individuals with medically unexplained symptoms that include cognitive impairment (e.g., cognitive symptoms related to concentration, memory, or slowed thinking in patients with fibromyalgia or other medically unexplained pain syndromes) (Greiffenstein et al., 2013; Johnson-Greene et al., 2013). The rate of PVT failure is significant in these populations. For example, Johnson-Greene and colleagues (2013) reported a 37 percent failure rate in fibromyalgia patients, regardless of disability entitlement status. Greiffenstein and colleagues (2013) reported a 74 percent failure rate in disability-seeking patients with Complex Regional Pain Syndrome Type I. Sensitivity of PVTs may vary in these populations; in one large (n = 326) study of disability claimants (mainly with musculoskeletal and other pain conditions), rates of performance below cut-off levels varied from 17 to 43 percent on three different PVTs (Gervais et al., 2004), underscoring the need for administration of multiple PVTs during the assessment session.

CONCLUSION

The results of standardized cognitive tests that are appropriately administered, interpreted, and validated can provide objective evidence to help identify and document the presence and severity of medically determinable mental impairments at Step 2 of SSA's disability determination process. In addition, such tests can provide objective evidence to help identify and assess the severity of work-related cognitive functional impairment relevant to disability evaluations at the listing level (Step 3) and to mental residual functional capacity (Steps 4 and 5).Therefore, standardized cognitive test results are essential to the determination of all cases in which an applicant's allegation of cognitive impairment is not accompanied by objective medical evidence.

The results of cognitive tests are affected by the effort put forth by the test-taker. If an individual has not given his or her best effort in taking the test, the results will not provide an accurate picture of the person's neuropsychological or cognitive functioning. Performance validity indicators, which include PVTs, analysis of internal data consistency, and other corroborative evidence, help the evaluator to interpret the validity of an individual's neuropsychological or cognitive test results. For this reason, it is important to include an assessment of performance validity at the time

cognitive testing is administered. It also is important that validity be assessed throughout the cognitive evaluation.

PVTs provide information about the validity of cognitive test results when administered as part of the test or test battery and are an important addition to the medical evidence of record for specific groups of applicants. It is important that PVTs only be administered in the context of a larger test battery and only be used to interpret information from that battery. Evidence of invalid performance based on PVT results pertains only to the cognitive test results obtained and does not provide information about whether or not the individual is, in fact, disabled. A lack of validity on PVTs alone is insufficient grounds for denying a disability claim.

REFERENCES

AACN (American Academy of Clinical Neuropsychology). 2007. AACN practice guidelines for neuropsychological assessment and consultation. *Clinical Neuropsychology* 21(2):209-231.

Allen, L. M., III, R. L. Conder, P. Green, and D. R. Cox. 1997. *CARB '97: Manual for the computerized assessment of response bias*. Durham, NC: Cognisyst.

American Psychiatric Association. 2013. *The diagnostic and statistical manual of mental disorders: DSM-5*. Washington, DC: American Psychiatric Association.

APA (American Psychological Association). 2015. *Guidelines and principles for accreditation of programs in professional psychology: Quick reference guide to doctoral programs*. http://www.apa.org/ed/accreditation/about/policies/doctoral.aspx (accessed January 20, 2015).

Barrash, J., A. Stillman, S. W. Anderson, Y. Uc, J. D. Dawson, and M. Rizzo. 2010. Predicition of driving ability with neuropsychological tests: Demographic adjustments diminish accuracy. *Journal of the International Neuropsychological Society* 16(04):679-686.

Benedict, R. H. 1997. *Brief visuospatial memory test—revised: Professional manual*. Lutz, FL: Psychological Assessment Resources.

Benedict, R. H., D. Schretlen, L. Groninger, and J. Brandt. 1998. Hopkins Verbal Learning Test–Revised: Normative data and analysis of inter-form and test-retest reliability. *The Clinical Neuropsychologist* 12(1):43-55.

Benton, A. L., K. S. de Hamsher, N. R. Varney, and O. Spreen. 1983. *Contributions to neuropsychological assessment: A clinical manual*. New York: Oxford University Press.

Benton, L., K. de Hamsher, and A. Sivan. 1994a. Controlled oral word association test. *Multilingual Aphasia Examination* 3.

Benton, A. L., K. S. de Hamsher, N. R. Varney, and O. Spreen. 1994b. *Contributions to neuropsychological assessment: A clinical manual—second edition*. New York: Oxford University Press.

Berthelson, L., S. S. Mulchan, A. P. Odland, L. J. Miller, and W. Mittenberg. 2013. False positive diagnosis of malingering due to the use of multiple effort tests. *Brain Injury* 27(7-8):909-916.

Bianchini, K. J., C. W. Mathias, and K. W. Greve. 2001. Symptom validity testing: A critical review. *The Clinical Neuropsychologist* 15(1):19-45.

Bigler, E. D. 2012. Symptom validity testing, effort, and neuropsychological assessment. *Journal of the International Neuropsychological Society* 18(4):632-642.

Bigler, E. D. 2014. Limitations with symptom validity, performance validity, and effort tests. Presentation to IOM Committee on Psychological Testing, Including Validity Testing, for Social Security Administration, June 25, 2014, Washington, DC.

Bigler, E. D. 2015. *Use of symptom validity tests and performance validity tests in disability determinations*. Paper commissioned by the IOM Committee on Psychological Testing, Including Validity Testing, for Social Security Administration Disability Determinations. http://www.iom.edu/psychtestingpaperEB (accessed April 9, 2015).

Bilder, R. M., C. A. Sugar, and G. S. Hellemann. 2014. Cumulative false positive rates given multiple performance validity tests: Commentary on Davis and Millis (2014) and Larrabee (2014). *The Clinical Neuropsychologist* 28(8):1212-1223.

Binder, L. M. 1993. *Portland Digit Recognition Test manual—second edition*. Portland, OR: Private Publication.

Binder, L. M., and S. C. Willis. 1991. Assessment of motivation after financially compensable minor head trauma. *Psychological Assessment* 3(2):175-181.

Binder, L. M., M. R. Villanueva, D. Howieson, and R. T. Moore. 1993. The Rey AVLT recognition memory task measures motivational impairment after mild head trauma. *Archives of Clinical Neuropsychology* 8:137-147.

Binder, L. M., G. L. Iverson, and B. L. Brooks. 2009. To err is human: "Abnormal" neuropsychological scores and variability are commin in healthy adults. *Archives of Clinical Neuropsychology* 24:31-46.

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

Boone, K. B. 2009. The need for continuous and comprehensive sampling of effort/response bias during neuropsychological examinations. *The Clinical Neuropsychologist* 23(4):729-741.

Boone, K. B. 2014. *Selection and use of multiple performance validity tests (PVTs)*. Presentation to IOM Committee on Psychological Testing, Including Validity Testing, for Social Security Administration, June 25, 2014, Washington, DC.

Boone, K. B. and P. Lu. 2007. Non-forced-choice effort measures. In *Assessment of malingered neurocognitive deficits*, edited by G. J. Larrabee. New York: Oxford University Press. Pp. 27-43.

Boone, K. B., P. Lu, C. Back, C. King, A. Lee, L. Philpott, E. Shamieh, and K. Warner-Chacon. 2002a. Sensitivity and specificity of the Rey Dot Counting Test in patients with suspect effort and various clinical samples. *Archives of Clinical Neuropsychology* 17(7):625-642.

Boone, K. B., P. H. Lu, and D. Herzberg. 2002b. *The B Test manual*. Los Angeles: Western Psychological Services.

Boone, K. B., P. Lu, and J. Wen. 2005. Comparison of various RAVLT scores in the detection of non-credible memory performance. *Archives of Clinical Neuropsychology* 20:301-319.

Brandt, J., and R. H. Benedict. 2001. *Hopkins Verbal Learning Test, Revised: Professional manual*. Lutz, FL: Psychological Assessment Resources.

Brandt, J., and W. van Gorp. 1999. American Academy of Clinical Neuropsychology policy on the use of non-doctoral-level personnel in conducting clinical neuropsychological evaluations. *The Clinical Neuropsychologist* 13(4):385.

Busch, R. M., G. J. Chelune, and Y. Suchy. 2006. Using norms in neuropsychological assessment of the elderly. In *Geriatric neuropsychology: Assessment and intervention*, edited by D. K. Attix and K. A. Welsh-Bohmer. New York: Guilford Press.

Bush, S. S., R. M. Ruff, A. I. Tröster, J. T. Barth, S. P. Koffler, N. H. Pliskin, C. R. Reynolds, and C. H. Silver. 2005. Symptom validity assessment: Practice issues and medical necessity. NAN policy & 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.

Carrow-Woolfolk, E. 1999. *CASL: Comprehensive Assessment of Spoken Language*. Circle Pines, MN: American Guidance Services.

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. 2011. The psychological consultative examination for Social Security disability. *Psychological Injury and Law* 4(3-4):235-244.
Chafetz, M. D., and J. Underhill. 2013. Estimated costs of malingered disability. *Archives of Clinical Neuropsychology* 28(7):633-639.
Chafetz, M. D., J. P. Abrahams, and J. Kohlmaier. 2007. Malingering on the Social Security disability consultative exam: A new rating scale. *Archives of Clinical Neuropsychology* 22(1):1-14.
Conder, R., L. Allen, and D. Cox. 1992. *Computerized Assessment of Response Bias test manual*. Durham, NC: Cognisyst.
Davis, J. J., and S. R. Millis. 2014. Examination of performance validity test failure in relation to number of tests administered. *The Clinical Neuropsychologist* 28(2):199-214.
Dean, A. C., T. L. Victor, K. B. Boone, and G. Arnold. 2008. The relationship of IQ to effort test performance. *The Clinical Neuropsychologist* 22(4):705-722.
Delis, D. C. 1994. *CVLT-C, California Verbal Learning Test: Children's version: Manual*. San Antonio, TX: The Psychological Corporation.
Delis, D. C., J. H. Kramer, and E. Kaplan. 2000. *California Verbal Learning Test: CVLT-II; adult version; manual*. San Antonio, TX: The Psychological Corporation.
Delis, D., E. Kaplan, and J. Kramer. 2001. *Delis-Kaplan executive function system*. San Antonio, TX: The Psychological Corporation.
DeRight, J., and D. A. Carone. 2015. Assessment of effort in children: A systematic review. *Child Neuropsychology* 21(1):1-24.
Edmonds, E. C., L. Delano-Wood, D. R. Galasko, D. P. Salmon, and M. W. Bondi. 2014. Subjective cognitive complaints contribute to misdiagnosis of mild cognitive impairment. *Journal of the International Neuropsychological Society* 20(8):836-847.
Elliott, R. 2003. Executive functions and their disorders. *British Medical Bulletin* 65:49-59.
Etherton, J. L., K. J. Bianchini, M. A. Ciota, and K. W. Greve. 2005a. Reliable Digit Span is unaffected by laboratory-induced pain: Implications for clinical use. *Assessment* 12(1):101-106.
Etherton, J. L., K. J. Bianchini, K. W. Greve, and M. A. Ciota. 2005b. Test of Memory Malingering performance is unaffected by laboratory-induced pain: Implications for clinical use. *Archives of Clinical Neuropsychology* 20(3):375-384.
Etkin, A., A. Gyurak, and R. O'Hara. 2013. A neurobiological approach to the cognitive deficits of psychiatric disorders. *Dialogues in Clinical Neuroscience* 15(4):419.
Farias, S. T., D. Mungas, and W. Jagust. 2005. Degree of discrepancy between self and other-reported everyday functioning by cognitive status: Dementia, mild cognitive impairment, and healthy elders. *International Journal of Geriatric Psychiatry* 20(9):827-834.
Faust, D., K. Hart, T. Guilmette, and H. Arkes. 1988. Neuropsychologists' capacity to detect adolescent malingerers. *Professional Psychology: Research and Practice* 19:508-515.
Frederick, R. I. 1997. *Validity indicator profile manual*. Minnetonka, MN: NCS Assessments.
Frederick, R. I., and H. G. Foster. 1991. Multiple measures of malingering on a forced-choice test of cognitive ability. *Psychological Assessment* 3(4):596-602.
Freedman, D., and J. Manly. 2015. *Use of normative data and measures of performance validity and symptom validity in assessment of cognitive function*. Paper commissioned by the IOM Committee on Psychological Testing, Including Validity Testing, for Social Security Administration Disability Determinations. http://www.iom.edu/psychtestingpapersDFJM (accessed April 9, 2015).
Funahashi, S. 2001. Neuronal mechanisms of executive control by the prefrontal cortex. *Neuroscience Research* 39:147-165.
Gast, J., and K. J. Hart. 2010. The performance of juvenile offenders on the Test of Memory Malingering. *Journal of Forensic Psychology Practice* 10(1):53-68.

Gervais, R. O., M. L. Rohling, P. Green, and W. Ford. 2004. A comparison of WMT, CARB, and TOMM failure rates in non-head injury disability claimants. *Archives of Clinical Neuropsychology* 19(4):475-487.

Goodglass, H., and E. Kaplan. 1983. *Boston diagnostic aphasia examination.* Philadelphia: Lea & Febiger.

Graue, L. O., D. T. Berry, J. A. Clark, M. J. Sollman, M. Cardi, J. Hopkins, and D. Werline. 2007. Identification of feigned mental retardation using the new generation of malingering detection instruments: Preliminary findings. *The Clinical Neuropsychologist* 21(6):929-942.

Green, P. 2004. *Green's Memory Complaints Inventory (MCI).* Edmonton, Alberta, Canada: Green's.

Green, P. 2005. *Green's Word Memory Test for Window's: User's manual*. Edmonton, Alberta, Canada: Green's.

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

Green, P., and L. Flaro. 2003. Word Memory Test performance in children. *Child Neuropsychology* 9(3):189-207.

Green, P., L. Allen, and K. Astner. 1996. *The Word Memory Test: A user's guide to the oral and computer-administered forms, U.S. version 1.1*. Durham, NC: CogniSyst.

Greiffenstein, M. F., W. J. Baker, and T. Gola. 1994. Validation of malingered amnesia measures with a large clinical sample. *Psychological Assessment* 6(3):218-224.

Greiffenstein, M., R. Gervais, W. J. Baker, L. Artiola, and H. Smith. 2013. Symptom validity testing in medically unexplained pain: A chronic regional pain syndrome type 1 case series. *The Clinical Neuropsychologist* 27(1):138-147.

Greve, K. W., and K. J. Bianchini. 2004. Setting empirical cutoffs on psychometric indicators of negative response bias: A methodological commentary with recommendations. *Archives of Clinical Neuropsychology* 19(4):533-541.

Griffin, G. A., J. Normington, R. May, and D. Glassmire. 1996. Assessing dissimulation among Social Security disability income claimants. *Journal of Consulting Clinical Psychology* 64(6):1425-1430.

Gronwall, D. 1977. Paced auditory serial-addition task: A measure of recovery from concussion. *Perceptual and Motor Skills* 44(2):367-373.

Grote, L. G. and J. N. Hook. 2007. Forced-choice recognition tests of malingering. In *Assessment of malingered neurocognitive deficits*, edited by G. J. Larrabee. New York: Oxford University Press. Pp. 27-43.

Groth-Marnat, G. 2009. *Handbook of psychological assessment*. Hoboken, NJ: John Wiley & Sons.

Hammill, D. D., and S. C. Larsen. 2009. *Test of written language: Examiner's manual*. 4th ed. Austin, TX: Pro-Ed.

Hampson, N. E., S. Kemp, A. K. Coughlan, C. J. Moulin, and B. B. Bhakta. 2013. Effort test performance in clinical acute brain injury, community brain injury, and epilepsy populations. *Applied Neuropsychology: Adult* (ahead-of-print):1-12.

Heaton, R. K. 1993. *Wisconsin Card Sorting Test: Computer version 2*. Odessa, FL: Psychological Assessment Resources.

Heaton, R. K., H. H. Smith, R. A. Lehman, and A. T. Vogt. 1978. Prospects for faking believable deficits on neuropsychological testing. *Journal of Consulting and Clinical Psychology* 46(5):892.

Heaton, R. K., I. Grant, and C. G. Matthews. 1991. *Comprehensive norms for an expanded Halstead-Reitan Battery: Demographic corrections, research findings, and clinical applications*. Odessa, FL: Psychological Assessment Resources.

Heaton, R. K., M. Taylor, and J. Manly. 2001. Demographic effects and demographically corrected norms with the WAIS-III and WMS-III. In *Clinical interpretations of the WAIS-II and WMS-III*, edited by D. Tulsky, R. K. Heaton, G. J. Chelune, I. Ivnik, R. A. Bornstein, A. Prifitera, and M. Ledbetter. San Diego, CA: Academic Press. Pp. 181-210.

Heilbronner, R. L., J. J. Sweet, J. E. Morgan, G. J. Larrabee, S. R. Millis, and 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.

Higginson, C. I., K. Lanni, K. A. Sigvardt, and E. A. Disbrow. 2013. The contribution of trail making to the prediction of performance-based instrumental activities of daily living in Parkinson's disease without dementia. *Journal of Clinical and Experimental Neuropsychology* 35(5):530-539.

Hiscock, M., and C. K. Hiscock. 1989. Refining the forced-choice method for the detection of malingering. *Journal of Clinical and Experimental Neuropsychology* 11(6):967-974.

HNS (Houston Neuropsychological Society). 2003. *The Houston Conference on Specialty Education and Training in Clinical Neuropsychology policy statement*. http://www.uh.edu/hns/hc.html (accessed November 25, 2014).

Holdnack, J. A., and L. W. Drozdick. 2009. *Advanced clinical solutions for WAIS-IV and WMS-IV: Clinical and interpretive manual*. San Antonio, TX: Pearson.

Hurley, K. E., and W. P. Deal. 2006. Assessment instruments measuring malingering used with individuals who have mental retardation: Potential problems and issues. *Mental Retardation* 44(2):112-119.

Iverson, G. L., and M. D. Franzen. 1996. Using multiple objective memory procedures to detect simulated malingering. *Journal of Clinical and Experimental Neuropsychology* 18(1):38-51.

Jelicic, M., H. Merckelbach, I. Candel, and E. Geraets. 2007. Detection of feigned cognitive dysfunction using special malinger tests: A simulation study in naïve and coached malingerers. *The International Journal of Neuroscience* 117(8):1185-1192.

Johnson-Greene, D., L. Brooks, and T. Ference. 2013. Relationship between performance validity testing, disability status, and somatic complaints in patients with fibromyalgia. *The Clinical Neuropsychologist* 27(1):148-158.

Kaplan, E., H. Goodglass, and S. Weintraub. 2001. *Boston Naming Test*. Austin, TX: Pro-Ed.

Killgore, W. D., and L. DellaPietra. 2000. Using the WMS-III to detect malingering: Empirical validation of the rarely missed index (RMI). *Journal of Clinical and Experimental Neuropsychology* 22:761-771.

Kirkwood, M. 2014. *Validity testing in pediatric populations*. Presentation to IOM Committee on Psychological Testing, Including Validity Testing, for Social Security Administration, June 25, 2014, Washington, DC.

Kirkwood, M. W., K. O. Yeates, C. Randolph, and J. W. Kirk. 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.

Larrabee, G. J. 2007. Introduction: Malingering, research designs, and base rates. In *Assessment of malingered neuropsychological deficits,* edited by G. J. Larrabee. New York: Oxford University Press.

Larrabee, G. J. 2012a. Assessment of malingering. In *Forensic neuropsychology: A scientific approach*, edited by G. J. Larrabee. New York: Oxford University Press.

Larrabee, G. J. 2012b. Performance validity and symptom validity in neuropsychological assessment. *Journal of the International Neuropsychological Society* 18(4):625-630.

Larrabee, G. J. 2014a. False-positive rates associated with the use of multiple performance and symptom validity tests. *Archives of Clinical Neuropsychology* 29(4):364-373.

Larrabee, G. J. 2014b. *Performance and Symptom Validity.* Presentation to IOM Committee on Psychological Testing, Including Validity Testing, for Social Security Administration, June 25, 2014, Washington, DC.

Lewis, R. F. 1990. *Digit Vigilance Test*. Lutz, FL: Psychological Assessment Resources.

Lezak, M., D. Howieson, E. Bigler, and D. Tranel. 2012. *Neuropsychological assessment*. 5th ed. New York: Oxford University Press.

Lu, P. H., K. B. Boone, L. Cozolino, and C. Mitchell. 2003. Effectiveness of the Rey-Osterrieth Complex Figure Test and the Meyers and Meyers Recognition Trial in the detection of suspect effort. *The Clinical Neuropsychologist* 17:426-440.

MacAllister, W. S., L. Nakhutina, H. A. Bender, S. Karantzoulis, and C. Carlson. 2009. Assessing effort during neuropsychological evaluation with the TOMM in children and adolescents with epilepsy. *Child Neuropsychology* 15(6):521-531.

Manly, J., and R. Echemendia. 2007. Race-specific norms: Using the model of hypertension to understand issues of race, culture, and education in neuropsychology. *Archives of Clinical Neuropsychology* 22(3):319-325.

McCrea, M., J. P. Kelly, C. Randolph, R. Cisler, and L. Berger. 2002. Immediate neurocognitive effects of concussion. *Neurosurgery* 50(5):1032-1042.

McCrea, M., K. M. Guskiewicz, S. W. Marshall, W. Barr, C. Randolph, R. C. Cantu, J. A. Onate, J. Yang, and J. P. Kelly. 2003. Acute effects and recovery time following concussion in collegiate football players: The NCAA concussion study. *JAMA* 290(19):2556-2563.

Meyers, J. E., and M. Volbrecht. 1999. Detection of malingers using the Rey Complex Figure and Recognition Trial. *Applied Neuropsychology* 6:201-207.

Mittenberg, W., C. Patton, E. M. Canyock, and D. C. Condit. 2002. Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology* 24(8):1094-1102.

Mittenberg, W., C. Patton, and W. Legler. 2003. *Identification of malingered head injury on the Wechsler Memory Scale—Third Edition*. Paper presented at the annual conference of the National Academy of Neuropsychology, Dallas, TX.

Moritz, S., S. Ferahli, and D. Naber. 2004. Memory and attention performance in psychiatric patients: Lack of correspondence between clinician-rated and patient-rated functioning with neuropsychological test results. *Journal of the International Neuropsychological Society* 10(4):623-633.

NAN (National Academy of Neuropsychology). 2001. *NAN definition of a clinical neuropsychologist: Official position of the National Academy of Neuropsychology*. https://www.nanonline.org/docs/PAIC/PDFs/NANPositionDefNeuro.pdf (accessed November 25, 2014).

Niccolls, R., and J. F. Bolter 1991. *Multi-Digit Memory Test*. San Luis Obispo, CA: Wang Neuropsychological Laboratories.

NIH (National Institutes of Health). n.d. *NIH toolbox: Processing speed*. http://www.nihtoolbox.org/WhatAndWhy/Cognition/ProcessingSpeed/Pages/default.aspx (accessed October 15, 2014).

OIDAP (Occupational Information Development Advisory Panel). 2009. *Mental cognitive subcommittee: Content model and classification recommendations*. http://www.ssa.gov/oidap/Documents/AppendixC.pdf (accessed October 6, 2014).

Paulhus, D. L. 1998. *Paulhus Deception Scales (PDS)*. Toronto: Multi-Health Systems.

Randolph, C. 1998. *Repeatable Battery for the Assessment of Neuropsychological Status (RBANS)*. San Antonio, TX: Psychological Corporation.

Rao, S. M. 1986. Neuropsychology of multiple sclerosis: A critical review. *Journal of Clinical and Experimental Neuropsychology* 8(5):503-542.

Reitan, R. M. 1992. *Trail Making Test: Manual for administration and scoring*. Mesa, AZ: Reitan Neuropsychology Laboratory.

Reitan, R. M., and D. Wolfson. 1993. *The Halstead-Reitan neuropsychological test battery: Theory and clinical interpretation—second edition*. Tucson: Neuropsychology Press.

Rey, A. 1941. L'examen psychologique dans les cas d'encéphalopathie traumatique (les problems). *Archives de Psychologie* 28:286-340.

Rey, A. 1964. *The clinical examination in psychology*. Paris, France: Presses Universitaires de France.

Roberson, C. J., K. B. Boone, H. Goldberg, D. Miora, M. Cottingham, T. Victor, E. Ziegler, M. Zeller, and M. Wright. 2013. Cross validation of the B test in a large known groups sample. *The Clinical Neuropsychologist* 27(3):495-508.

Ruben, R. J. 1999. Redefining the survival of the fittest: Communication disorders in the 21st century. *International Journal of Pediatric Otorhinolaryngology* 49:S37-S38.

Salazar, X. F., P. H. Lu, J. Wen, and K. B. Boone. 2007. The use of effort tests in ethnic minorities and in non-English-speaking and English as a second language populations. In *Assessment of feigned cognitive impairment: A neuropsychological perspective*, edited by K. B. Boone. New York: Guilford Press. Pp. 405-427.

Salekin, K. L., and B. M. Doane. 2009. Malingering intellectual disability: The value of available measures and methods. *Applied Neuropsychology* 16(2):105-113.

Schacter, D. L. 1990. Toward a cognitive neuropsychology of awareness: Implicit knowledge and anosognosia. *Journal of Clinical and Experimental Neuropsychology* 12(1):155-178.

Schmidt, M. 1996. *Rey Auditory Verbal Learning Test: RAVLT: A handbook*. Los Angeles: Western Psychological Services.

Schretlen, D. J., S. Testa, J. M. Winicki, G. D. Pearlson, and B. Gordon. 2008. Frequency and bases of abnormal performance by healthy adults on neuropsychological testing. *Journal of the International Neuropsychological Society* 14(3):436-445.

Semel, E., E. Wiig, and W. Secord. 2003. *Clinical evaluation of language fundamentals: Examiners manual*. 4th ed. San Antonio, TX: The Psychological Corporation.

Shandera, A. L., D. T. Berry, J. A. Clark, L. J. Schipper, L. O. Graue, and J. P. Harp. 2010. Detection of malingered mental retardation. *Psychological Assessment* 22(1):50-56.

Sheslow, D., and W. Adams. 2003. *Wide range assessment of memory and learning second edition administration and technical manual*. Lutz, FL: Psychological Assessment Resources.

Silverberg, N. D., and S. R. Millis. 2009. Impairment versus deficiency in neuropsychological assessment: Implications for ecological validity. *Journal of the International Neuropsychological Society* 15(1):94-102.

Silverton, L. 1999. *Malingering Probability Scale (MPS) manual*. Los Angeles, CA: Western Psychological Services.

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

Slick, D. J., E. M. S. Sherman, and G. L. Iverson. 1999. Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist (Neuropsychology, Development and Cognition: Section D)* 13(4):545-561.

Sollman, M. J., and D. T. Berry. 2011. Detection of inadequate effort on neuropsychological testing: A meta-analytic update and extension. *Archives of Clinical Neuropsychology* 26(8):774-789.

Solomon, R. E., K. B. Boone, D. Miora, S. Skidmore, M. Cottingham, T. Victor, E. Ziegler, and M. Zeller. 2010. Use of the WAIS-III picture completion subtest as an embedded measure of response bias. *The Clinical Neuropsychologist* 24(7):1243-1256.

Spreen, O., and E. Strauss. 1991. Controlled oral word association (word fluency). In *A compendium of neuropsychological tests*, edited by O. Spreen and E. Strauss. Oxford, UK: Oxford University Press. Pp. 219-227.

SSA (Social Security Administration). n.d.-a. *Disability evaluation under social security—Part III: Listing of impairments—Adult listings (Part A)—section 12.00 mental disorders*. http://www.ssa.gov/disability/professionals/bluebook/12.00-MentalDisorders-Adult.htm (accessed November 14, 2014).

SSA. n.d.-b. *Disability evaluation under Social Security: Part I—general information*. http://www.ssa.gov/disability/professionals/bluebook/general-info.htm (accessed November 14, 2014).

Stevens, A., K. Schneider, B. Liske, L. Hermle, H. Huber, and G. Hetzel. 2014. Is subnormal cognitive performance in schizophrenia due to lack of effort or to cognitive impairment? *German Journal of Psychiatry* 17(1):9.

Strauss, E., E. M. Sherman, and O. Spreen. 2006. *A compendium of neuropsychological tests: Administration, norms, and commentary.* Oxford, UK: Oxford University Press.

Suchy, Y., G. Chelune, E. I. Franchow, and S. R. Thorgusen. 2012. Confronting patients about insufficient effort: The impact on subsequent symptom validity and memory performance. *The Clinical Neuropsychologist* 26(8):1296-1311.

Suhr, J. A., and D. Boyer. 1999. Use of the Wisconsin Card Sorting Test in the detection of malingering in student simulator and patient samples. *Journal of Clinical and Experimental Neuropsychology* 21:701-708.

Sweet, J. J., D. G. Meyer, N. W. Nelson, and P. J. Moberg. 2011. The TCN/AACN 2010 "salary survey": Professional practices, beliefs, and incomes of U.S. neuropsychologists. *The Clinical Neuropsychologist* 25(1):12-61.

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

Trahan, D. E., and G. J. Larrabee. 1988. *Continuous Visual Memory Test*. Odessa, FL: Psychological Assessment Resources.

van Gorp, W. G., L. A. Humphrey, A. Kalechstein, V. L. Brumm, W. J. McMullen, M. Stoddard, and N. A. Pachana. 1999. How well do standard clinical neuropsychological tests identify malingering?: A preliminary analysis. *Journal of Clinical and Experimental Neuropsychology* 21(2):245-250.

Victor, T. L., and K. B. Boone. 2007. Identification of feigned mental retardation. In *Assessment of feigned cognitive impairment*, edited by K. Boone. New York: Guilford Press. Pp. 310-345.

Victor, T. L., K. Boone, J. G. Serpa, J. Buehler, and E. Ziegler. 2009. Interpreting the meaning of multiple symptom validity test failure. *The Clinical Neuropsychologist* 23(2):297-313.

Warrington, E. 1984. *Recognition Memory Test manual*. Windsor: Nfer-Nelson.

Wechsler, D. 1997a. *Wechsler Adult Intelligence Scale (WAIS-III): Administration and scoring manual—3rd edition*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. 1997b. *WMS-III: Wechsler Memory Scale administration and scoring manual*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. 2003. *Wechsler Intelligence Scale for Children—fourth edition (WISC-IV)*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. 2008. *Wechsler Adult Intelligence Scale—fourth edition (WAIS-IV)*. San Antonio, TX: NCS Pearson.

Wechsler, D. 2009. *WMS-IV: Wechsler Memory Scale—Administration and scoring manual*. San Antonio, TX: The Psychological Corporation.

WHO (World Health Organization). 2001. *International classification of functioning, disability, and health (ICF)*. Geneva, Switzerland: WHO.

Young, G. 2014. Resource material for ethical psychological assessment of symptom and performance validity, including malingering. *Psychological Injury and Law* 7(3):206-235.

6

Economic Considerations

This chapter discusses the possible financial impact of the committee's recommendations that the U.S. Social Security Administration (SSA) require systematic use of standardized psychological testing for a broader set of physical and mental impairments than is current practice for applicants who allege cognitive impairment or whose allegation of functional impairment is based solely on self-report. Although the committee's recommendations are based on its assessment of the scientific evidence underlying standardized psychological testing and of the contributions such testing could make to determinations regarding the extent of impairment and degree of functional capacity in those populations, it recognizes that financial considerations also are relevant to decisions regarding implementation of psychological testing. In this context, the chapter provides an initial framework for evaluating the economic costs of implementation and highlights the types of data that will be needed to accurately determine the financial impact of mandatory psychological testing as recommended by the committee for disability determinations. A more thorough assessment of the financial implications is beyond the committee's ability or charge.

The chapter begins with a discussion of the potential cost outlays associated with required psychological testing and describes how these costs vary by test type, provider, and geographical location. As a benchmark, simple cost estimates are provided, along with sensitivity analysis that illustrates the relationship between financial outlays and the size of the applicant population requiring testing. The chapter then focuses on the potential financial benefits of testing, primarily any cost savings from expanding the use of psychological testing as recommended by the committee. In this

context, the chapter discusses research arguing that requiring psychological testing, specifically symptom validity tests (SVTs) and performance validity tests (PVTs), will generate significant savings for the Social Security Disability Insurance (SSDI) and Supplemental Security Income (SSI) programs by greatly reducing the number of "false" favorable determinations (false positives). The chapter concludes with a summary of the types of data that SSA and state Disability Determination Services (DDS) offices would need to collect in order to accurately assess the net financial impact of implementation.

COSTS OF PSYCHOLOGICAL TESTING

Costs of Psychological Testing Services

As the recommendations state, the administration of psychological testing would be part of the normal disability determination process. As such, applicants could provide any required tests in their initial application for disability benefits. In these cases, required psychological testing would impose no financial costs on SSA. For applicants without such tests, SSA could gather the information as part of case development. In some cases, testing may necessitate a consultative examination. In all cases, the costs to SSA of providing testing would relate to the administration and interpretation of all required tests.[1]

To ensure that any test results are reliable, specialists appropriately trained in the administration and interpretation of standardized psychological tests would need to be used. Depending on the type of tests being given, trained providers include psychiatrists or other appropriately licensed physicians, licensed psychologists, and trained and licensed technicians.[2] One estimate of the current costs of these services comes from the Medicare reimbursement rates, which are updated yearly and are used to determine what Medicare will pay to providers treating Medicare patients. Table 6-1 reports average Medicare reimbursement rates in 2014 for psychological testing services provided outside of a facility such as a hospital.[3] These services include

[1] It is difficult to project how many applicants would respond to testing requirements by seeking testing in advance of filing an application. One way SSA could estimate this is by examining the share of applicants with intellectual disabilities who file for benefits with all required testing in the application.

[2] In some cases tests, could be administered online using computer-administered tests. These tests still require a licensed provider to interpret the results.

[3] In some cases, costs of services are significantly lower when provided inside a facility. Because most of the applicants for disability benefits live in the community rather than in an institution, the present discussion focuses on non-facility prices.

TABLE 6-1 Costs of Psychological and Neuropsychological Testing Services

	National Average Cost		Standard Deviation		
Type of Services	Weighted	Unweighted	Unweighted	Minimum	Maximum
Psychiatric diagnostic interview (90791)	$134	$136	$7.6	$124	$188
Psychological testing by psychologist/ physician (96101)	$81	$82	$4.5	$75	$115
Psychological testing by technician (96102)	$66	$67	$6.3	$51	$85
Neurobehavioral status exam (96116)	$95	$96	$5.8	$85	$129
Neuropsychological testing by psychologist/ physician (96118)	$99	$101	$6.2	$88	$134
Neuropsychological testing by technician (96119)	$81	$83	$8.1	$62	$106
Health and behavioral assessment (96150)[a]	$86	$87	$4.8	$81	$122

[a] Centers for Medicare & Medicaid Services provides pricing data for this code in 15-minute rather than hourly increments. Hence the data were transformed to hourly rates for the purpose of comparability to other codes.
SOURCE: CMS, 2015, and committee calculations.

(1) psychiatric diagnostic interview, HCPCS code 90791[4]; (2) psychological testing by a psychologist or physician, HCPCS code 96101; (3) psychological testing by a technician, HCPCS code 96102; (4) neurobehavioral status exam, HCPCS code 96116; (5) neuropsychological testing by a psychologist or physician, HCPCS code 96118; (6) neuropsychological testing by a technician, HCPCS code 96119; and (7) health and behavioral assessment, HCPCS code 96150. For purposes of comparison, the costs are shown for 1 hour of service. In practice, the time for evaluation varies with the type of testing required and the complexity of the case.[5]

The average cost of testing services varies by the type of testing, psychological versus neuropsychological, and by the type of provider, as in a psychologist or physician versus a technician.[6] For an equivalent unit of service, a psychiatric diagnostic interview is the most expensive and was reimbursed by Medicare at an average rate of $134 in 2014. Psychological testing by a technician is the least expensive, with an average reimbursement rate of $66 in 2014.

As the minimum and maximum values in the table highlight, the cost of purchasing qualified psychological testing services of any type varies considerably across states and localities (SSAB, 2012, p. 52, Figure 47). For example, in the most expensive area, 1 hour of psychiatric evaluation costs $188 compared to $124 in the least expensive area. There is also substantial variation in service costs for general psychological testing, with the variation greater among technician-provided services than services provided by psychologists or physicians. The variation in pricing is similarly large for neuropsychological testing. For physicians or psychologists providing neuropsychological testing, Medicare reimbursement rates vary from $88

[4] The codes listed reflect a sample of codes that may be used by providers.

[5] The length of an evaluation will vary depending on the purpose of the evaluation, and more specifically, the type of psychological and/or cognitive impairments being assessed. Most psychological and neuropsychological evaluations include (1) a clinical interview, (2) administration of standardized cognitive or non-cognitive psychological tests, and (3) professional time for interpretation and integration of data. The relevant CPT codes for each of these processes are generally billed in 1 hour per unit of service (the exception is 96150, which is a 15 minute/unit code). That is, an evaluation may include billing for 1 hour for clinical interview (96116), 1 hour for administration of tests (96119), and 1 hour for interpretation and integration (96118) for a total of 3 hours of clinical service. However, a more complex case likely will require additional hours of test administration and interpretation/integration in order to fully answer the clinical question. In fact, the results of a national professional survey indicate that billing for a typical neuropsychological evaluation is roughly 6 hours, with a range from 0.5 to 25 hours (Sweet et al., 2011).

[6] The table includes both weighted and unweighted averages. Weighted averages are appropriate for considering total costs to SSA since they are weighted to reflect population differences across counties in which the reimbursement rate holds. Unweighted averages provide information relevant to considering cost dispersion across states. Average prices referenced in the text reflect weighted averages.

to $134 per hour/unit billed depending on location. The variation is even larger for technicians as reflected in the larger standard deviation of reimbursement rates. In general, price variation occurs for all testing types with the exception of the health and behavioral assessment.

The variation in pricing of services by geographical area implies that the costs to SSA of requiring psychological testing will depend, in part, on the geographical location of the applicants most likely to require testing. As shown in Chapter 2, there is considerable variation in application filing rates for disability benefits across U.S. states. This variation suggests that the demand for psychological testing for disability determinations will also vary, resulting in larger outlays in some states than in others. Whether this variation in demand for testing services interacts with variation in testing prices to reduce or increase costs is something that would have to be investigated once testing is implemented.

Part of the service price variation shown in Table 6-1 owes to regional differences in overall price levels. However, differences in the availability of providers and the overall demand for psychological services in the area may also play a role. In markets where providers are limited but filing rates for SSDI or SSI are high, required use of psychological testing by SSA potentially could increase demand for testing services sufficiently to have an impact on service prices. Given the small share of disability applicants relative to the population, this seems unlikely in large metropolitan areas. However, in smaller rural areas or states with fewer providers, any increase in demand for services might affect market prices. To the extent that testing could be computer administered and scored and interpreted by a provider living outside of the applicant's geographical area, these impacts would be lessened. Determining the best method to provide testing services cost-effectively to disability benefit applicants would be an important element of implementing the recommendations in this report.

Another factor that could push up costs relative to the numbers in Table 6-1 is that providers may demand higher payments than those offered by Medicare. DDS offices are not under the Medicare reimbursement rules, and if providers asked for more to provide required psychological services presumably the offices and SSA would have to pay those rates. Finally, it is possible that the use of psychological testing by SSA could create a market for test preparation or test coaching that would in turn lead to a need for new and improved tests, and then more coaching, and so forth. Should this occur, the costs of testing by SSA could potentially rise over time. The likelihood of this type of "testing spiral" and its impact on costs is something that could be monitored and assessed in the early stages of implementation.

There are also potential cost offsets that might make testing less expensive for SSA than the Medicare reimbursement rates would suggest. For example, if SSA decides to use testing on a large scale it might be able to

purchase licenses for testing products or contract with a national provider of testing services, resulting in lower fees for service. With respect to geographic considerations, SSA might be able to rely on telemedicine for clinical interviews and/or technician administration of tests, with offsite interpretation by psychologist/neuropsychologists on large national or regional contracts. SSA could consult with the Veterans Health Administration or private disability insurers to assess the feasibility and likely cost savings of these alternatives.

Tested Populations and Estimates of Costs

The cost of requiring psychological testing depends on the price of the tests and on the number of individuals who must be tested. There is no straightforward way to map the committee's recommendations regarding who should receive psychological testing onto SSA's publicly available data to derive an accurate measure of the size of the tested population.[7] However, the data do permit the calculation of cost estimates associated with testing groups of applicants the committee judges to be most likely to fall under the recommendations in this report. The results of this exercise are provided in Table 6-2. The table shows cost computations for testing applicants who reach Step 4 or 5 of the disability determination process described in Chapter 2. These are individuals who did not qualify for benefits by meeting or equaling the medical listings but were sent along for further evaluation, rather than being denied. By definition, these are individuals for whom a determination regarding benefits requires further case development, including assessment of their ability to perform substantial gainful activity at some job in the national economy.[8] In addition to calculations for all applicants reaching this stage, the table shows cost estimates should psychological testing be required for the subset of applicants with mental impairments other than intellectual disabilities or arthritis and back disorders.

The results from this exercise demonstrate the variation in projected costs associated with factors related to implementation including which tests will be required, the qualifications mandated for testing providers, and the number of individuals who will need to be tested. For example, if SSA provided psychiatric diagnostic interviews at the average Medicare reimbursement rate for all applicants reaching Step 4 or 5, the cost would

[7] SSA collects a variety of data that it does not provide publicly and may be able to do a more accurate initial assessment of the costs associated with the recommendations. However, to fully measure the potential costs it is likely that SSA would need to pilot the use of testing and the costs associated with it.

[8] For children applying for SSI, the evaluation is based on attending school rather than working.

TABLE 6-2 Estimated Costs of Testing

Medicare Price Data (Non-Facility Rates) and Disability Application Data by Diagnostic Group (in thousand dollars)

Mental Disorders (Excluding Intellectual Disability)

	Number of Persons	Psychiatric Diagnostic Interview (90791)	Psychological Testing by Psychologist/ Physician (96101)	Psychological Testing by Technician (96102)	Neuro-behavioral Status Exam (96116)	Neuro-psychological Testing by Psychologist/ Physician (96118)	Neuro-psychological Testing by Technician (96119)	Health and Behavioral Assessment (96150)
SSDI Claimants	87,809	$11,764.65	$7,109	$5,819	$8,336	$8,713	$7,141	$1,887.02
Concurrent Claimants	124,928	$16,737.85	$10,114	$8,279	$11,859	$12,397	$10,159	$2,684.70
SSI Adult Claimants	132,163	$17,707.20	$10,700	$8,758	$12,546	$13,115	$10,747	$2,840.18
SSI Child Claimants	42,540	$5,699.51	$3,444	$2,819	$4,038	$4,221	$3,459	$914.18
Total Cost	N/A	$51,909	$31,367	$25,676	$36,780	$38,446	$31,507	$8,326

continued

TABLE 6-2 Continued

Arthritis and Back Disorders

	Number of Persons	Psychiatric Diagnostic Interview (90791)	Psychological Testing by Psychologist/ Physician (96101)	Psychological Testing by Technician (96102)	Neuro-behavioral Status Exam (96116)	Neuro-psychological Testing by Psychologist/ Physician (96118)	Neuro-psychological Testing by Technician (96119)	Health and Behavioral Assessment (96150)
SSDI Claimants	259,977	$34,831.72	$21,048	$17,229	$24,680	$25,798	$21,141	$5,586.91
Concurrent Claimants	176,617	$23,663.15	$14,299	$11,704	$16,766	$17,526	$14,362	$3,795.50
SSI Adult Claimants	106,257	$14,236.31	$8,03	$7,042	$10,087	$10,544	$8,641	$2,283.46
SSI Child Claimants	297	$39.79	$24	$20	$28	$29	$24	$6.38
Total Cost	N/A	$72,771	$43,973	$35,994	$51,561	$53,897	$44,169	$11,672

All Diagnostic Groups								
	Number of Persons	Psychiatric Diagnostic Interview (90791)	Psychological Testing by Psychologist/ Physician (96101)	Psychological Testing by Technician (96102)	Neuro-behavioral Status Exam (96116)	Neuro-psychological Testing by Psychologist/ Physician (96118)	Neuro-psychological Testing by Technician (96119)	Health and Behavioral Assessment (96150)
SSDI Claimants	584,669	$78,333.95	$47,335	$38,746	$55,503	$58,017	$47,545	$12,564.54
Concurrent Claimants	515,157	$69,020.73	$41,708	$34,139	$48,904	$51,119	$41,893	$11,070.72
SSI Adult Claimants	391,431	$52,443.93	$31,690	$25,940	$37,159	$38,842	$31,831	$8,411.85
SSI Child Claimants	921,12	$12,341.17	$7,457	$6,104	$8,744	$9,140	$7,491	$1,979.49
Total Cost	N/A	$212,140	$128,190	$104,930	$150,309	$157,118	$128,760	$34,027

NOTE: Based on 2013 application data and 2014 Medicare pricing information, geographically weighted. Values in Table 6-2 may not exactly reflect multiplication of weighted pricing data from Table 6-1 and number of persons in column one of Table 6-2 due to rounding error.
SOURCES: CMS, 2015; SSA, 2014c,d,e; and committee calculations.

be $212 million. This cost would drop to $51 million if such testing were only provided to applicants with mental disorders (excluding intellectual disabilities). Similarly, costs would be lower if other forms of psychological testing were required or if other types of service providers were used.

Importantly, the cost estimates in Table 6-2 assume that SSA will be responsible for all the costs of psychological testing. However, as noted previously, some applicants may acquire and include required tests as part of the medical records presented at application. In this case, the cost to SSA would be minimal, providing that the disability determination offices already have sufficient personnel to adequately evaluate the test findings.

Another assumption implicit in this simple cost calculation is that the psychological testing would be added to current DDS case development costs. To the extent that psychological testing replaces rather than augments existing case development modalities, the costs to SSA would be lower than the simple estimates in the table. There are good reasons to believe that this might be the case. Consultative exams are already a common component of disability determinations.[9] Some of these exams include psychological testing and it might be possible to add additional tests with limited additional costs.

Of course, the estimates in Table 6-2 could also understate the costs, especially since the calculations rely on a mapping of the recommendations to publically available data that may insufficiently capture the true number of individuals who could require testing. Accurately assessing the costs of mandatory psychological testing by SSA will require more detailed information on the parameters of implementation as well as experience in the field once testing has begun.

ASSESSING THE BENEFITS OF PSYCHOLOGICAL TESTING

Recent calls for greater use of psychological testing in SSA's disability determination process assume that the current process is making significant mistakes and allowing unqualified applicants onto the disability programs (Chafetz and Underhill, 2013; IOPC, 2013). However, the committee has been unable to uncover any evidence on either side of this claim. At present, there do not appear to be any independently conducted studies regarding the accuracy of the disability determination process as implemented by DDS offices. As such, it is difficult to assess whether greater use of psychological testing will increase, decrease, or leave unchanged the number of individuals awarded benefits. The outcome depends on how accurately DDS offices currently are in making disability determinations.

[9] On average 47 percent of disability evaluations include a consultative examination, although there is considerable variation across states (SSA, 2014a,b).

Even if DDS offices are making relatively accurate determinations in the absence of psychological testing, greater standardization could produce other benefits. A more standardized process could potentially reduce the number of applicants who appeal their decisions. For applicants who do appeal, the inclusion of psychological testing in the medical records could help reduce the burden on administrative law judges to make subjective determinations on the adequacy of the claim. Standardization might also make the process more transparent and efficient, improving public understanding and reducing the time it takes to process claims. However, none of these potential benefits can be quantified without additional research on the accuracy and efficiency of current practice. Such an assessment is an important first step in developing an implementation strategy for the committee's recommendations.

ESTIMATES OF COST SAVINGS FROM PSYCHOLOGICAL TESTING

One of the main purported benefits of mandatory psychological testing is its potential to generate significant savings for the SSDI and SSI programs. The proponents of this view argue that requiring psychological testing (SVTs and PVTs) for SSDI and SSI applicants would result in a significant reduction of the number of individuals allowed onto the benefit rolls. For example, Chafetz and Underhill (2013) estimate that requiring SVTs and PVTs in the DDS process would save approximately $12.8 billion for the SSDI system and $7.2 billion for the SSI system, or about 40 percent of total program costs (see Tables 6-3 and 6-4, reproduced from Chafetz and Underhill [2013]). The estimated savings results from the assumed reduction in the number of falsely awarded individuals coming onto the disability programs.[10]

The committee performed a critical evaluation of this estimate and concluded that it is based on several assumptions that if violated would substantially lower the projected cost savings. Most important is the assumption that the current disability determination process, as implemented by DDS offices, is unable to detect any applicants who exaggerate or fabricate their impairments and related functional limitations. Although not stated directly in the analysis, this assumption is implicit in the authors' use of base rates of malingering from populations of applicants and claimants ex ante of any disability screening. For example, the $12.8 and $7.2 billion savings computed by Chafetz and Underhill (2013) assumes that 40 percent of current SSDI

[10] Improved accuracy could also decrease the number of individuals falsely denied benefits. However, the focus of the literature has been on reducing those falsely allowed onto the program.

TABLE 6-3 Calculation of 2011 SSDI Costs for Each Level of Malingering of Mental Disorders

Level (%)	No. Disabled Workers = 2,768,928	2011 Total Cost $32,067,993,684
10	276,893	$3.207 B
20	553,786	$6.414 B
30	830,678	$9.620 B
40	**1,107,571**	**$12.827 B**
50	1,384,464	$16.034 B
60	1,661,357	$19.241 B
70	1,938,250	$22.448 B
80	2,215,142	$25.654 B
90	2,492,035	$28.861 B

NOTES: The 40 percent rate is bolded as the probable rate of malingering given in Larrabee, Millis, and Meyers (2009). For the SSDI total, the number of disabled workers is used, removing spouse and child beneficiaries. Costs were estimated by multiplying the average disability figure for each mental condition by the December 2011 number of individuals with that condition, summing over all conditions, and then multiplying by 12 for the yearly estimated amount. B = billion.
SOURCE: Chafetz and Underhill, 2013. Reproduced with permission.

TABLE 6-4 Calculation of 2011 SSI (Adult) Costs for Each Level of Malingering of Mental Disorders

Level (%)	No. of Adults less than age 65 = 2,797,743	2011 Total Cost $32,067,993,684
10	279,774	$1.799 B
20	559,549	$3.597 B
30	839,323	$5.396 B
40	**1,119,097**	**$7.195 B**
50	1,398,872	$8.994 B
60	1,678,646	$10.792 B
70	1,958,420	$12.591 B
80	2,238,194	$14.390 B
90	2,517,969	$16.189 B

NOTES: The 40 percent rate is bolded as the probable rate of malingering given in Larrabee, Millis, and Meyers (2009). The SSI figures include the number of adults (less than age 65) minus the children as of December 2011. Costs were estimated by multiplying the average disability figure for each mental condition by the December 2011 number of individuals with that condition, summing over all conditions, and then multiplying by 12 for the yearly estimated amount. B = billion.
SOURCE: Chafetz and Underhill, 2013. Reproduced with permission.

and SSI beneficiaries were falsely awarded and would have been denied benefits if given a SVT or PVT as part of the disability determination process. This assumption is synonymous with the view that DDS offices currently detect no one who exaggerates or fabricates their condition, symptoms, or functional limitations. In other words, the Chafetz and Underhill computation assumes that under current practice 40 percent of all awardees are given benefits even though they are not truly eligible. The extremeness of the Chafetz and Underhill assumption suggests that the cost savings associated with psychological testing is likely to be lower than they suggest.

The other important assumption embedded in the Chafetz and Underhill projected cost savings is that SVTs and PVTs would be retroactively applied to the population of existing beneficiaries, regardless of time on the program.[11] Should SSA choose to implement mandatory SVT and PVT testing, it would likely do so for new applicants to the disability programs, making the potential cost savings lower than that computed by Chafetz and Underhill.

Finally, the Chafetz and Underhill calculation is static. The more appropriate method of computing cost savings is to consider the present discounted value of an estimated stream of potential benefit savings, which would generate a much larger estimate.

The importance of altering the assumptions about improved accuracy of disability determinations and the size of the population exposed to testing can be seen in Table 6-5. Reflecting the mapping of the committee's recommendations for testing used in Table 6-2, cost savings are estimated for new awardees with mental impairments other than intellectual disabilities and for those with arthritis and back disorders. For completeness, the estimates are also provided for all new beneficiaries, regardless of condition and for all awardees and awardees determined eligible in Steps 4 or 5 of the disability determination process. The alternative estimates also show the sensitivity of the estimated cost savings to the assumption about the potential for mandatory SVT and PVT use to improve the accuracy of SSA disability determinations. The 40 percent test failure rate preferred by Chafetz and Underhill (2013) applies if the current SSA process detects zero percent of those who exaggerate or fabricate; the 10 percent test failure rate applies if SSA is relatively accurate, but makes some false-positive errors that would be identified through the use of SVTs and PVTs.

Several important points emerge from the computations in the table. First, the potential annual cost savings associated with mandatory SVT and PVT testing is substantially reduced when it is applied to new awardees

[11] Chafetz and Underhill (2013) limit the group to those with mental disorders, but even so this assumption greatly increases the cost savings associated with greater use of testing, because it essentially applies the 40 percent base malingering rate to all existing beneficiaries.

TABLE 6-5 Estimated Annual Savings of Testing New Disability Awardees

Average Benefit,[a] Diagnostic Distribution,[b] and Disability Applications Data[c] (in thousands of dollars)

				40 Percent Test Failure Assumed[d]		10 Percent Test Failure Rate Assumed	
		Number of Awards	Number of Awards (Step 4 or 5 of the Determination Process)	All Awardees	Awardees Awarded at Steps 4 or 5 of the Determination Process	All Awardees	Awardees Awarded at Steps 4 or 5 of the Determination Process
Mental Disorders (excluding intellectual disability)	SSDI	49,700	28,398	$236,060	$134,882	$59,015	$33,721
	Concurrent	42,041	21,430	$157,117	$80,089	$39,279	$20,022
	SSI Adults	54,639	24,225	$152,923	$67,801	$38,231	$16,950
	SSI Children	72,203	41,636	$202,081	$116,531	$50,520	$29,133
Arthritis and Back Disorders	SSDI	117,512	109,295	$671,336	$624,393	$167,834	$156,098
	Concurrent	46,459	42,098	$173,628	$157,330	$43,407	$39,332
	SSI Adults	32,649	29,677	$81,172	$73,783	$20,293	$18,466
	SSI Children	622	244	$1,546	$607	$387	$152
All Diagnostic Groups	SSDI	399,722	233,522	$2,069,914	$1,209,267	$517,479	$302,317
	Concurrent	210,812	111,331	$787,853	$416,070	$196,963	$104,017
	SSI Adults	183,930	90,792	$498,182	$245,914	$124,546	$61,479
	SSI Children	171,574	90,479	$464,716	$$245,066	$116,179	$61,267

[a] SSDI benefit data are from 2012, and SSI and concurrent benefit data are from 2013. For concurrent enrollees, there are no data available on average benefit payments by diagnosis, so the average benefit level for all persons was used for all concurrent enrollment calculations. For SSDI and SSI, the average benefit amount for mental disabilities (excluding intellectual disability) was calculated as a weighted average of the average monthly benefits awarded for mental disability diagnoses (excluding intellectual disability) using diagnostic distribution data. For musculoskeletal conditions, there are no data available specifically for back disorders and arthritis, so the average benefit for musculoskeletal disorders was used to calculate estimated savings. SSA did not have information concerning average SSI benefits by diagnosis available separately for children and adults, so a single weighted average was used for both groups using diagnostic and benefit distributions for all recipients under age 65.

[b] SSDI diagnostic distribution data are from 2012. SSI and concurrent enrolled diagnostic distribution data are from 2013.

[c] All disability application data are from 2013.

[d] Test failure rates are synonymous with what some literature refers to as malingering rates.

SOURCES: SSA, 2014c,d,e, and committee calculations.

rather than all beneficiaries on the programs. Considering only new awardees with mental impairments other than intellectual disabilities, the cost savings assuming the 40 percent malingering rate is $236 million for SSDI and $153 million for SSI, about one-fifth of the savings reported by Chafetz and Underhill (2013). Second, cost savings are also reduced when the assumption about the accuracy improvements associated with symptom and validity testing are relaxed. If SSA misses 10, rather than 40, percent of those with exaggerated or fabricated claims, the cost savings from mandatory testing on new awardees with mental impairments other than intellectual disabilities falls from $236 to $59 million for SSDI and from $153 to $38 million for SSI adults. Finally, cost savings decline if testing is required only for applicants who reach Steps 4 or 5 of the disability determination process. Although these estimates are far from exact, they suggest that caution is warranted when projecting potential cost savings from mandatory psychological testing.

As noted earlier, the static calculations in Table 6-5, although useful for comparing to Chafetz and Underhill, are not appropriate for computing the expected savings associated with implementing SVTs and PVTs in SSA's disability determination process. The expected program savings is more accurately calculated as the present discounted value of the averted payment flows associated with the denied applicants captured by psychological testing. Using the same diagnostic categories as in Table 6-5, Table 6-6 shows the present discounted value of expected savings from disallowing an unqualified applicant from each of the three disability programs. The table also shows the estimated program savings to SSA under the assumption that psychological testing as recommended would result in the denial of benefits to 10 percent of applicants who would otherwise receive them.

Two points emerge from the table. First, the expected cost savings associated with denying an applicant improperly allowed on the program can be sizeable, depending on the diagnosis and program. The estimated savings are largest for individuals with mental impairments; this reflects the earlier age of benefit receipt and longer average time on the program. Estimated savings are smallest for SSI recipients with arthritis and back pain, again largely reflecting the age at which recipients enter the program. Second, the amount of program savings that comes from implementing psychological testing depends mostly on how many additional individuals would be identified as unqualified for benefits relative to current practice. It is important to keep in mind that psychological testing as recommended may also result in the awarding of benefits to some portion of applicants who otherwise would be denied. Assuming that implementation of psychological testing reduces the number of newly awarded beneficiaries by 10 percent, the savings per cohort, while significant, still would be less than the annual savings estimated by Chafetz and Underhill.

TABLE 6-6 Estimated Lifetime Spending on an Individual Disability Awardee, 2 Percent Annual Discounting

	Individual Lifetime Savings—SSDI Average Benefit	Individual Lifetime Savings—SSI Average Benefit	Cohort Lifetime Savings—10% Test Failure Rate of New SSDI Awardees	Cohort Lifetime Savings—10% Test Failure Rate of New SSI Adult Awardees	Cohort Lifetime Savings—10% Test Failure Rate of New SSI Child Awardees
Mental Disorders (excluding intellectual disability)	$202,121	$119,101	$1,004,542,011	$650,756,461	$859,945,621
Arthritis and Back Disorders	$171,561	$74,662	$2,016,047,512	$243,763,319	$4,643,964
All Diagnostic Groups	$161,434	$84,438	$6,452,880,242	$1,553,065,482	$1,448,734,067

NOTE: SSDI benefit data from 2012, SSI from 2013. The average benefit amount for mental disabilities (excluding intellectual disability) was calculated as a weighted average of the average monthly benefits awarded for mental disability diagnoses (excluding intellectual disabilities) using diagnostic distribution data. For musculoskeletal conditions, there are no data available specifically for back disorders or arthritis, so the average benefit for musculoskeletal disorders was used to calculate estimated savings. Overall average benefit by program was used to calculate "all diagnostic groups" savings. SSA did not have information concerning average SSI benefits by diagnosis available separately for children and adults, so a single weighted average was used for both groups using diagnostic and benefit distributions for all recipients under age 65. Average time spent on disability benefits by diagnosis comes from Riley and Rupp (2014, Table 3). As Riley and Rupp do not differentiate between programs, the same value was used for all programs within a diagnosis.

FINDINGS

Understanding the financial costs and benefits of using psychological testing in the SSA disability determination process is an important, but unfinished, task. The data necessary to make accurate calculations are limited, and estimates based on available data are subject to considerable error. That said, the framework for a proper computation is well understood and can be used to guide data collection and evaluation when testing is and is not employed.

Accurate assessments of the net financial impact of mandatory psychological testing will require information on the current accuracy of DDS decisions and how the accuracy is improved, or unaffected, by the use of more

standardized testing. It will also be important to determine which types of tests should be given and to which groups in the applicant population. This information can then be used to consider the impact on the demand for testing services across the country and whether or not that demand affects service pricing. All of these components could be gathered in pilot programs that allow for experimentation and assessment prior to wider implementation. In addition, the committee found:

- The average cost of testing services varies by the type of testing (e.g., psychological, neuropsychological), by the type of provider (e.g., psychologist or physician, technician), and by geographical area. The variation in pricing implies that the expected costs to SSA of requiring psychological testing will depend on exactly which tests are required, the qualifications mandated for testing providers, and the geographical location of the providers most in demand.
- Estimating the exact cost of broad use of psychological testing by SSA will require more detailed data on the exact implementation strategy. To fully measure the potential costs, it is likely that SSA will need to pilot the use of testing and the costs associated with it.
- Some published estimates of the potential cost savings to SSA associated with the use of symptom validity testing and performance validity testing are based on assumptions that if violated would substantially lower the estimated cost savings. Potential cost savings associated with testing vary considerably based on the assumptions about who it is applied to and how many individuals it detects and thus rejects for disability benefits.
- At present, there do not appear to be any independently conducted studies regarding the accuracy of the disability determination process as implemented by DDS offices.
- A full financial cost-benefit analysis of psychological testing will require SSA to collect additional data both before and after the implementation of the recommendations of this report.

REFERENCES

Chafetz, M., and J. Underhill. 2013. Estimated costs of malingered disability. *Archives of Clinical Neuropsychology* 28(7):633-639.

CMS (Centers for Medicare & Medicaid Services). 2015. *Physician fee schedule search tool*. http://www.cms.gov/apps/physician-fee-schedule/search/search-criteria.aspx (accessed January 20, 2015).

IOPC (Inter Organizational Practice Committee). 2013. Use of symptom validity indicators in SSA psychological and neuropsychological evaluations. Letter to Senator Tom Coburn. https://www.nanonline.org/docs/PAIC/PDFs/SSA%20and%20Symptom%20Validity%20Tests%20-%20IOPC%20letter%20to%20Sen%20Coburn%20-%202-11-13.pdf (accessed February 8, 2015).

Larrabee, G. J., S. R. Millis, and J. E. Meyers. 2009. 40 plus or minus 10, a new magical number: Reply to Russell. *The Clinical Neuropsychologist* 23(5):841-849.

Riley, G. F., and K. Rupp. 2014. Cumulative expenditures under the DI, SSI, Medicare, and Medicaid programs for a cohort of disabled working-age adults. *Health Services Research* 50(2):514-536. doi: 10.1111/1475-6773.12219.

SSA (Social Security Administration). 2014a. *DDS performance management report. Disability claims data. Consultative examination rates, fiscal year 2013*. Data prepared by ORDP, ODP, and ODPMI. Submitted to the IOM Committee on Psychological Testing, Including Validity Testing, for Social Security Administration Disability Determinations by Joanna Firmin, Social Security Administration, on August 25, 2014.

SSA. 2014b. *Disability claims data (initial, reconsideration, continuing, disability review) by adjudicative level and body system. SSDI, SSI, concurrent, and total claims allowance rates for claims with consultative examinations by U.S. states, fiscal year 2013*. Data prepared by ORDP, ODP, and ODPMI. Submitted to the IOM Committee on Psychological Testing, Including Validity Testing, for Social Security Administration Disability Determinations by Joanna Firmin, Social Security Administration, on August 25, 2014.

SSA. 2014c. *National data Title II-SSDI, Title XVI-SSI, & Concurrent Title II/XVI initial disability determinations by regulation basis code (reason for decision), fiscal year 2013. All cases except mental disorders (other than intellectual disability) and arthritis and back diorders*. Data prepared by SSA, ORDP, ODP, and ODPMI. Submitted to the IOM Committee on Psychological Testing, Including Validity Testing, for Social Security Administration Disability Determinations by Joanna Firmin, Social Security Administration, on October 23, 2014.

SSA. 2014d. *National data Title II-SSDI, Title XVI-SSI, & Concurrent Title II/XVI initial disability determinations by regulation basis code (reason for decision), fiscal year 2013. Arthritis and back disorders only.* Data prepared by SSA, ORDP, ODP, and ODPMI. Submitted to the IOM Committee on Psychological Testing, Including Validity Testing, for Social Security Administration Disability Determinations by Joanna Firmin, Social Security Administration, on October 23, 2014.

SSA. 2014e. *National data Title II-SDI, Title XVI-SSI, & Concurrent Title II/XVI initial disability determinations by regulation basis code (reason for decision), fiscal year 2013. Mental disorders only (excluding intellectual disability)*. Data prepared by SSA, ORDP, ODP, and ODPMI. Submitted to the IOM Committee on Psychological Testing, Including Validity Testing, for Social Security Administration Disability Determinations by Joanna Firmin, Social Security Administration, on October 23, 2014.

SSAB (Social Security Advisory Board). 2012. *Aspects of disability decision making: Data and materials*. Washington, DC: SSAB.

Sweet, J. J., D. G. Meyer, N. W. Nelson, and P. J. Moberg. 2011. The TCN/AACN 2010 "salary survey": Professional practices, beliefs, and incomes of U.S. neuropsychologists. *The Clinical Neuropsychologist* 25(1):12-61.

7

Conclusions and Recommendations

ROLE OF PSYCHOLOGICAL TESTING IN SOCIAL SECURITY ADMINISTRATION DISABILITY PROGRAMS

The committee reached a number of general conclusions pertaining to the role of standardized psychological testing in the U.S. Social Security Administration (SSA) disability programs:

- The two largest impairment categories for Supplemental Security Income (SSI) (adults and children) and Social Security Disability Insurance (SSDI) are mental disorders (excluding intellectual disabilities) and musculoskeletal and connective tissue disorders. Within these two categories, a significant fraction of the applicants have conditions, including affective mood disorders and disorders of the back, for which the presence and severity of impairment and associated functional limitations are based largely on applicant self-report.
- SSA disability determinations are based on the medical and all relevant evidence in an applicant's case record. Physical or mental impairments must be established by objective medical evidence consisting of medical signs and laboratory findings, which may include psychological and other standardized test results. SSA establishes the presence of a medically determinable impairment in individuals with mental disorders other than intellectual disability through the use of standard diagnostic criteria, which include symptoms and signs. Evidence for claims based on mental impairment, as well as for claims for conditions in which the somatic symptoms

are disproportionate to physical findings (e.g., somatoform disorder, multisystem illness, and chronic pain), relies less on standard laboratory tests than for some other categories of impairment. The validity of the self-reported symptoms and/or impairment severity may be called into question due to the absence of objective medical evidence or biomarkers that could explain or substantiate the applicant's self-report of distress and disability.

- In some cases, SSA disability examiners must evaluate the credibility of statements by individuals about the intensity and persistence of their symptoms and the effect on the individual's ability to function and perform work-related activities. When a disability claim is based primarily on an applicant's self-report of symptoms and self-reported statements about their intensity, persistence, and limiting effects, SSA relies on an assessment of the consistency of the self-report with all of the evidence in the claimant's medical evidence record.
- There currently is great variability in allowance rates for both SSI and SSDI among states that is not fully accounted for by differences in the populations of applicants. In addition, there is great variability in the appeal rulings among administrative law judges within and across states.
- Psychological consultative examinations often consist of nonstandardized diagnostic interviews and a mental status exam, with little or no standardized psychological testing. Because clinicians generally are not as good at interpreting clinical and standardized test data as are established actuarial methods, reliance on established actuarial methods (when available) to interpret the data will improve the accuracy of diagnostic evaluations.
- Each Disability Determination Services (DDS) agency, within the confines of SSA policy, issues its own rules regarding the tests that may be purchased as part of a consultative examination. Aside from the use of intelligence tests as described in the listings for intellectual disability and certain neurological impairments, SSA does not require or specify the purchase of any type of (or individual) psychological test. SSA provides general guidance that good psychological tests are valid and reliable and have appropriate normative data. For this reason, there is variation among states about when and which standardized psychological tests can be purchased, with the exception of performance validity tests (PVTs) and symptom validity tests (SVTs), which are precluded from purchase by SSA except in rare cases such as a court order.
- The results of standardized cognitive tests and non-cognitive psychological tests that are appropriately administered, interpreted,

and validated can provide objective evidence to help identify and document the presence and severity of medically determinable mental impairments at Step 2 of SSA's disability determination process. In addition, standardized cognitive test results can provide objective evidence to help identify and assess the severity of work-related cognitive functional impairment relevant to disability evaluations at the listing level (Step 3) and to mental residual functional capacity (Steps 4 and 5).

- Current data on the prevalence of inconsistent reporting of symptoms or performing below one's capability on cognitive tests are very imprecise. In the context of SSA disability applicants, neither scenario rules out disability, but both suggest the need for additional assessment of the alleged impairment with the goal of making an accurate determination of disability.
- SVTs and PVTs provide information about the validity of standardized non-cognitive and cognitive test results when administered as part of the test or test battery and are an important addition to the medical evidence of record for specific groups of applicants. Validity tests do not provide information about whether or not the individual is, in fact, disabled.
- Because SVTs and PVTs are used to help assess the validity of an individual's standardized non-cognitive and/or cognitive psychological test results respectively, it is important that SVTs and PVTs only be administered in the context of a larger test battery and only be used to interpret information from that battery.
- Current SSA policy precludes the purchase of SVTs and PVTs to help inform determinations about the credibility of an individual's statements or about possible malingering. Specific tests outlined as examples in this policy include not only stand-alone PVTs and SVTs (e.g., Test of Memory Malingering, Validity Indicator Profile, Structured Interview of Reported Symptoms), but also psychological self-report measures that contain symptom validity scales (e.g., Minnesota Multiphasic Personality Inventory-2, Millon Clinical Multiaxial Inventory) among other scales of psychological functioning. This policy is inconsistent with the practice of other disability benefit programs, such as the Veterans Benefits Administration, private disability insurers, and some international disability programs.
- Although there currently are no data on the rates of false positives and false negatives in SSA disability determinations, systematic use of standardized psychological testing for a broader set of physical and mental impairments than is current practice is expected to improve the accuracy and consistency of disability determinations

for applicants who allege cognitive impairment or whose allegation of functional impairment is based solely on self-report.

STANDARDIZED NON-COGNITIVE PSYCHOLOGICAL MEASURES AND SYMPTOM VALIDITY TESTS

The following conclusions and recommendation pertain specifically to the use of standardized non-cognitive psychological measures and associated SVTs in SSA disability determinations:

- The use of standardized non-cognitive psychological measures is essential to the determination of all cases in which an applicant's allegation of non-cognitive functional impairment meets each of three requirements:
  1. The applicant alleges a mental disorder (i.e., schizophrenic, paranoid, and other psychotic disorders; affective disorders; anxiety-related disorders; and personality disorders) unaccompanied by cognitive complaints or a disorder with somatic symptoms that are disproportionate to demonstrable medical morbidity (i.e., somatoform disorders, multisystem illnesses, and chronic idiopathic pain conditions).
  2. The presence and severity of impairment and associated functional limitations are based largely on applicant self-report.
  3. Objective medical evidence or longitudinal medical records sufficient to make a disability determination do not accompany the claim.
- In certain instances, cognitive concerns may accompany the applicant's allegations, in which case cognitive testing, as discussed below, may be more appropriate. The committee also recognizes that there are a few chronic conditions (e.g., schizophrenia, chronic idiopathic pain, multisystem illnesses) that may generate potentially disabling, non-cognitive functional impairments but may not be accompanied by objective medical evidence. In such cases, the evidence provided by longitudinal medical records may be sufficient to substantiate the allegation.
- Assessment of symptom validity, including the use of SVTs, analysis of internal data consistency, and other corroborative evidence, helps the evaluator to interpret the accuracy of an individual's self-report of behavior, experiences, or symptoms and responses on standardized non-cognitive psychological measures. For this reason, it is important to include an assessment of symptom validity when non-cognitive psychological measures are administered.

- Evidence of inconsistent self-report based on an assessment of symptom validity is cause for concern with regard to self-reported symptoms but does not provide information about whether or not the individual is, in fact, disabled. A lack of validity on symptom validity testing alone is insufficient grounds for denying a disability claim, although additional information would be required to assess the applicants' allegation of disability.

Recommendation 1: The Social Security Administration should require the results of standardized non-cognitive psychological testing in the case record for all applicants whose claim of functional impairment relates either (1) to a mental disorder unaccompanied by cognitive complaints or (2) to a disorder in which the somatic symptoms are disproportionate to the medical findings. Testing should be required when the allegation is based primarily on applicant self-report and is not accompanied by objective medical evidence or longitudinal medical records sufficient to make a disability determination.

- All non-cognitive psychological assessments should include a statement of evidence of the validity of the results, which could include symptom validity test results, analysis of internal data consistency (e.g., item response theory), and other corroborative evidence as well as discussion of the test norms relative to the individual being assessed.
- For cases in which validation is not achieved, SSA should pursue additional evidence of the applicant's allegation.

The committee intends *standardized non-cognitive psychological tests* to include measures of behavior, affect, personality, and psychopathology. By *objective medical evidence* in this and the following recommendation, the committee means medical signs and/or laboratory or test results that constitute clear objective medical evidence of a significant mental disorder and related functional impairment of sufficient severity to make a disability determination. An example would be a severe brain injury associated with significant functional deficits (e.g., minimally conscious state). By *longitudinal medical records* the committee means a documented history of a significant mental disorder or a chronic condition such as chronic idiopathic pain or multisystem illness and related functional impairment of sufficient severity and duration to make a disability determination. An example would be a well-documented history of repeated hospitalizations and treatments for a diagnosed mental disorder, such as an affective or personality disorder.

The committee intends the "statement of evidence of the validity of the results" specified in this and the following recommendation to reflect objective evidence that goes beyond the clinical opinion of the examiner. In addition to analysis of the results of SVTs or PVTs administered at the time of the testing and analysis of internal data consistency, evidence could include a pattern of test results that is inconsistent with the alleged condition, observed behavior, documented history, and the like. It is important to note that a finding of inconsistency between the test results and the areas specified is more informative than a finding of consistency would be.

The committee's recommendation here and in the following recommendation that SSA "pursue additional evidence of the applicant's allegation" for cases in which validation is not achieved means that the test results in those cases are an insufficient basis to make a determination regarding disability status.

STANDARDIZED COGNITIVE TESTS AND PERFORMANCE VALIDITY TESTS

The following conclusions and recommendation pertain specifically to the use of standardized cognitive tests and associated PVTs in SSA disability determinations:

- Standardized cognitive test results are essential to the determination of all cases in which an applicant's allegation of cognitive impairment is not accompanied by objective medical evidence.
- The results of cognitive tests are affected by the effort put forth by the test-taker. If an individual has not given his or her best effort in taking the test, the results will not provide an accurate picture of the person's neuropsychological or cognitive functioning. Performance validity indicators, which include PVTs, analysis of internal data consistency, and other corroborative evidence, help the evaluator to interpret the validity of an individual's neuropsychological or cognitive test results. For this reason, it is important to include an assessment of performance validity at the time cognitive testing is administered. It also is important that validity be assessed throughout the cognitive evaluation.
- A PVT only provides information about the validity of an individual's cognitive test results that are obtained during the same evaluation. Evidence of invalid performance based on PVT results pertains only to the cognitive test results obtained and does not provide information about whether or not the individual is, in fact, disabled. A lack of validity on performance validity testing alone is insufficient grounds for denying a disability claim. In such cases,

additional information is required to assess the applicant's allegation of disability.

Recommendation 2: The Social Security Administration should require the results of standardized cognitive testing be included in the case record for all applicants whose allegation of cognitive impairment is not accompanied by objective medical evidence.

- All cognitive evaluations should include a statement of evidence of the validity of the results, which could include performance validity test results, analysis of internal data consistency (e.g., item response theory), and other corroborative evidence as well as discussion of the test norms relative to the individual being assessed.
- For cases in which validation is not achieved, SSA should pursue additional evidence of the applicant's allegation.

QUALIFICATIONS FOR TEST ADMINISTRATION AND INTERPRETATION

The committee reached the following conclusions and recommendation about the qualifications for the administration and interpretation of standardized psychological tests:

- Use of standardized procedures for the administration of standardized non-cognitive and cognitive psychological tests enables application of normative data to the individual being evaluated. Without standardized administration, the test-taker's performance may not accurately reflect his or her ability. It is important that any person administering cognitive or neuropsychological tests be well trained in the administration protocols for those particular tests, possess the interpersonal skills necessary to build rapport with the test-taker, and understand important psychometric properties, including validity and reliability, as well as factors that could emerge during testing to place either at risk.
- Interpretation of standardized psychological test results is more than a report of the standardized test scores; it requires assigning meaning to the scores within the individual context of the specific examinee. As such, interpretation of test results requires a higher level of clinical training than does the administration alone of some psychological tests.
- Licensed psychologists and neuropsychologists are the specialists qualified to interpret the results of most standardized psychological and neuropsychological tests. Under close supervision and direction

of licensed psychologists and neuropsychologists, it is standard practice for psychometrists or technicians with specialized training to administer and score tests. Test manuals specify the qualifications necessary for administration, scoring, and interpretation of the test or measure.

- It is important as well that the individual responsible for making the disability determination (disability examiner or administrative law judge) have the training and experience to understand and evaluate the report provided by the psychologist or neuropsychologist.

Recommendation 3: The Social Security Administration should ensure that psychological testing that is considered as part of a disability evaluation is performed by qualified specialists properly trained in the administration and interpretation of standardized psychological tests.

- "Qualified" means that the specialist must be currently licensed or certified to administer, score, and interpret psychological tests and have the training and experience to administer the test and interpret the results.
- This recommendation applies not only to standardized psychological testing that may be ordered in the course of a disability evaluation, but also to standardized psychological testing already in an applicant's medical evidence of record if the results are considered as part of the disability determination.

ECONOMIC CONSIDERATIONS

The committee concluded the following with respect to the complex economic considerations raised by increased systematic use of standardized psychological testing by SSA as recommended:

- The average cost of testing services varies by the type of testing (e.g., psychological, neuropsychological), by the type of provider (e.g., psychologist or physician, technician), and by geographical area. The variation in pricing implies that the expected costs to SSA of requiring psychological testing will depend on exactly which tests are required, the qualifications mandated for testing providers, and the geographical location of the providers most in demand.
- Estimating the exact cost of broad use of psychological testing by SSA will require more detailed data on the exact implementation strategy. To fully measure the potential costs, it is likely that SSA will need to pilot the use of testing and the costs associated with it.

- At present, there do not appear to be any independently conducted studies regarding the accuracy of the disability determination process as implemented by DDS offices. Some published estimates of billions of dollars in potential cost savings to SSA associated with the use of symptom validity testing and performance validity testing are based on assumptions that if violated would substantially lower the estimated cost savings. Potential cost savings associated with testing vary considerably based on the assumptions about who it is applied to and how many individuals it detects and thus rejects for disability benefits.
- A full financial cost-benefit analysis of psychological testing will require SSA to collect additional data both before and after the implementation of the recommendations of this report.

EVALUATION AND RESEARCH

Based on its examination of the literature and dialogues with experts in a variety of areas, including psychological and neuropsychological testing, performance validity testing and symptom validity testing, and the disability evaluation process both within SSA and in other arenas, the committee recognizes many questions remain with regard to the use of standardized psychological testing in the disability determination process.

As part of its assessment of the use of standardized psychological tests for the disability evaluation process, the committee was asked to discuss the costs and cost-effectiveness of requiring a single test or a combination of tests. This report provides an initial framework for evaluating the economic costs and highlights the types of data that will be needed to accurately determine the financial impact of implementing the committee's first two recommendations. The following conclusions and recommendation relate to this enterprise.

- Accurate assessments of the net financial impact of psychological testing as recommended by the committee will require information on the current accuracy of DDS decisions and how the accuracy is affected by the increased use of standardized psychological testing.
- The absence of data on the rates of false positives and false negatives in current SSA disability determinations precludes any assessment of their accuracy and consistency.
- There currently is great variability in allowance rates for both SSI and SSDI among states that is not fully accounted for by differences in the populations of applicants. There also is great variability in the disability determination appeal rulings among administrative

law judges within and across states. Although it is not possible to know definitively whether the large share of unexplained variation in state filing, award, and allowance rates is driven by variability in the federal disability determination process, there is some evidence that states differ in how they manage claims.

- In light of this unexplained variability, systematic use of standardized psychological testing as recommended by the committee is expected to improve the accuracy and consistency of disability determinations.

Recommendation 4: The Social Security Administration (SSA), in collaboration with other federal agencies, should establish a demonstration project(s) to investigate the accuracy and consistency of SSA's disability determinations with and without the use of recommended psychological testing.

- *Accuracy* refers to the rates of false negatives and false positives in SSA's disability determinations.
- *Consistency* means that adjudicators presented with the same evidence for comparable cases come to the same conclusion.

Recognizing that the costs and benefits of implementing the committee's recommendations go beyond the financial, the committee recommends that SSA evaluate the effect of implementing the committee's recommendations on its disability determination process using a number of different measures.

Recommendation 5: Following implementation of the committee's recommendations, the Social Security Administration should evaluate their impact on its disability determination process and end results. Measures of impact may include

- Number of backlogged cases;
- Efficiency of throughput or time to determination;
- Number of requests for appeals;
- Adherence to recommended evaluations;
- Effect on accuracy and consistency of disability determinations; and
- Effect on state-to-state variation in disability allowance rates and on appeal rulings among administrative law judges.

Over the course of the project, the committee identified two areas in particular in which it expects that the results of further research would help to inform disability determination processes as indicated in the following conclusions and recommendation.

- Additional research is needed on the use of SVTs and PVTs in populations representative of the pool of disability applicants, including in terms of gender, ethnicity, race, primary language, educational level, medical condition, and the like. In particular, additional research on the development of appropriate criterion or cut-off scores for PVTs and SVTs in these populations for the purposes of disability evaluation would be beneficial.
- The committee's task was to evaluate the value of psychological testing in the disability determination process, as reflected in the foregoing recommendations. However, the committee recognizes that just as systematic use of standardized psychological testing is expected to improve the accuracy and consistency of disability determinations for applicants who allege cognitive impairment or whose allegation of functional impairment is based solely on self-report, the use of other standardized assessment tools also may be expected to improve the accuracy of disability determinations. The value of standardized assessment tools, including psychological tests, to assessments of individuals' work-related functional capacity is an area that would benefit from further research.

Recommendation 6: The Social Security Administration and other federal agencies should support a program of research to investigate the value of standardized assessment, including psychological testing, in disability determinations. Such a program should support original research on a variety of topics, including

- The effects of standardized psychological testing on the accuracy and consistency of disability determinations;
- The use of performance validity tests and symptom validity tests with disability applicants; and
- The use of psychological tests, including performance validity tests and symptom validity tests, in different populations with regard to fairness for members of all gender, ethnic, racial, language, educational levels, and other protected groups.

A

Public Workshop Agendas

Workshop on Psychological Testing, Including Validity Testing, for Social Security Administration Disability Determinations (Workshop 1)

Hosted by the IOM Committee on Psychological Testing, Including Validity Testing, for Social Security Administration Disability Determinations

June 25, 2014
Room 106
Keck Center of the National Academies
500 Fifth Street, NW
Washington, DC

AGENDA

8:30 a.m. **Opening remarks**
Herbert Pardes, M.D., Committee Chair

8:45 a.m. **Overview of symptom validity testing and performance validity testing in the context of psychological testing**
Moderator—*Elizabeth W. Twamley, Ph.D.,* Committee Member

Performance and symptom validity
Glenn J. Larrabee, Ph.D., independent practice of clinical neuropsychology, Sarasota, Florida

Limitations with symptom validity, performance validity, and effort tests
Erin D. Bigler, Ph.D., Susa Young Gates Professor of Psychology and Neuroscience, Brigham Young University, Provo, Utah

DISCUSSION

10:15 a.m. **Break**

10:30 a.m. **An empirical approach to disability exaggeration**
Kevin J. Bianchini, Ph.D., Jefferson Neurobehavioral Group, Metairie, Louisiana

Selection and use of multiple performance validity tests (PVTs)
Kyle Brauer Boone, Ph.D., Professor, California School of Forensic Studies, Alliant International University, Torrance, California

DISCUSSION

12:00 p.m. **Break for lunch**

1:00 p.m. **Use of psychological tests, including SVTs, in select populations**
Moderator—*Lisa A. Suzuki, Ph.D.,* Committee Member

Validity testing in pediatric populations
Michael Kirkwood, Ph.D., Associate Clinical Professor, Physical Medicine and Rehabilitation, University of Colorado School of Medicine and Children's Hospital Colorado, Aurora, Colorado

Performance validity tests and symptom validity tests in culturally diverse populations
Jennifer J. Manly, Ph.D., Associate Professor of Neuropsychology, The Neurological Institute of New York, Columbia University Medical Center, New York, New York

Use of psychological tests, including PVTs and SVTs, in select populations: The U.S. military
Robert A. Seegmiller, Ph.D., Brooke Army Medical Center, Fort Sam Houston, Texas

DISCUSSION

3:00 p.m. **Break**

3:15 p.m. **Use of psychological tests in disability determinations in other systems**
Moderator—*Alan M. Jette, M.P.H., Ph.D.,* Committee Member

Veterans Affairs policies and/or practices surrounding the use of psychological tests and symptom validity tests in the disability determination process
Stacey Pollack, Ph.D., Director of Program Policy Implementation, Mental Health Services, Veterans Affairs Central Office, Washington, DC

Psychological disability evaluations under the Ontario auto insurance system and Ontario tort law
Brian Levitt, Psy.D., C.Psych., Past President, Canadian Academy of Psychologists in Disability Assessment, Ontario, Canada

Use of performance and symptom validity assessment within the independent disability insurer context
Thomas McLaren, Ph.D., Medical Consultant/Licensed Psychologist, Unum

DISCUSSION

5:10 p.m. **Closing remarks**
Herbert Pardes, M.D., Committee Chair

5:15 p.m. **Adjourn**

Workshop on Psychological Testing, Including Validity Testing, for Social Security Administration Disability Determinations (Workshop 2)

Hosted by the IOM Committee on Psychological Testing, Including Validity Testing, for Social Security Administration Disability Determinations

August 11, 2014
Room 100
Keck Center of the National Academies
500 Fifth Street, NW
Washington, DC

AGENDA

8:30 a.m. **Opening remarks**
Herbert Pardes, M.D., Committee Chair

8:40 a.m. **Discussion with the committee on the use of psychological, symptom validity, and performance validity testing in disability evaluations**
Moderator—*Peter A. Ubel, M.D.,* Committee Member

Terrence W. Dunlop, Ph.D., Chief Psychologist, Office of Medical Assistance, Social Security Administration

Robin Doyle, Medical Policy Expert, Office of Medical Policy, Social Security Administration

Michael D. Chafetz, Ph.D., Algiers Neurobehavioral Resource, LLC, New Orleans, Louisiana

Erin D. Bigler, Ph.D., Susa Young Gates Professor of Psychology and Neuroscience, Brigham Young University, Provo, Utah

10:20 a.m. **Break**

10:35 a.m. **Discussion with the committee on the use of psychological, symptom validity, and performance validity testing in disability evaluations** (continued)

11:20 a.m. **DISCUSSION**

11:45 a.m. **Break for lunch**

12:45 p.m. **Disability Determination Services panel discussion with the committee**
Moderator—*Mary C. Daly, Ph.D.,* Committee Member

Jennifer Nottingham, President, National Association of Disability Examiners; Supervisor, Ohio Disability Determination Service

Charles A. Jones, Director, Michigan Disability Determination Service

Tom A. Ward, Past President, National Association of Disability Examiners; Supervisor, Michigan Disability Determination Service

Jeffrey H. Price, President Elect, National Association of Disability Examiners; Disability Determination Specialist III, Health and Human Services Department, North Carolina

Nancy Heiser, Ph.D., Psychological Consultant, Washington, DC, Department of Disability Services

2:00 p.m. **Break**

2:15 p.m. **Disability Determination Services panel discussion with the committee** (continued)

3:30 p.m. **DISCUSSION**

3:55 p.m. **Closing remarks**
Herbert Pardes, M.D., Committee Chair

4:00 p.m. **Adjourn**

B

Biographical Sketches of Committee Members

Herbert Pardes, M.D. (*Chair*) is Executive Vice Chair of the Board of Trustees of New York-Presbyterian Hospital. He formerly served as President and Chief Executive Officer of New York-Presbyterian Hospital and the New York-Presbyterian Healthcare System. His origins are in the field of psychiatry, and he has an extensive background in health care and academic medicine. He is nationally recognized for his broad expertise in education, research, clinical care, and health policy, and as an ardent advocate of support for academic medicine. Dr. Pardes served as Director of the National Institute of Mental Health (NIMH) and U.S. Assistant Surgeon General during the Carter and Reagan administrations (1978–1984). Dr. Pardes left NIMH in 1984 to become Chair of the Department of Psychiatry at Columbia University's College of Physicians and Surgeons and in 1989 was also appointed Vice President for Health Sciences for Columbia University and Dean of the Faculty of Medicine at the College of Physicians and Surgeons. He served as President of the American Psychiatric Association (1989), as Chair of the Association of American Medical Colleges (AAMC) (1995–1996), and as Chair of the AAMC's Council of Deans (1994–1995). In addition, he served two terms as Chair of the New York Association of Medical Schools. Dr. Pardes chaired the Intramural Research Program Planning Committee of the National Institutes of Health (NIH) from 1996 to 1997, served on the Presidential Advisory Commission on Consumer Protection and Quality in the Healthcare Industry, and is President of the Scientific Council of the National Alliance for Research on Schizophrenia and Depression. He serves on numerous editorial boards, has written more

than 155 articles and chapters on mental health and academic medicine topics, and has negotiated and conducted international collaborations with a variety of countries including India, China, and the former Soviet Union. Dr. Pardes has earned numerous honors and awards, including the U.S. Army Commendation Medal (1964), the Sarnat International Prize in Mental Health (1997), election to the Institute of Medicine of the National Academy of Sciences (1997), and election to the American Academy of Arts and Sciences (2002). Dr. Pardes received his medical degree from the State University of New York-Downstate Medical Center (Brooklyn) in 1960. He received his bachelor of science degree summa cum laude from Rutgers University in 1956. He completed his internship and residency training in psychiatry at Kings County Hospital in Brooklyn and also did psychoanalytic training at the New York Psychoanalytic Institute.

Arthur J. Barsky III, M.D., is Professor of Psychiatry at Harvard Medical School and Vice Chair for Research in the Department of Psychiatry at the Brigham and Women's Hospital in Boston, Massachusetts. His major interests are hypochondriasis and somatization, the psychological factors that affect symptom reporting in the medically ill, and the cognitive and behavioral treatment of somatic symptoms. Dr. Barsky has been the principal investigator of nine National Institute of Mental Health (NIMH) and National Institutes of Health (NIH) research grants in these areas. He has authored 140 articles, 23 book chapters, and the books *Worried Sick: Our Troubled Quest for Wellness* and *Feeling Better*. Dr. Barsky received the President's Research Award from the American Psychosomatic Society. He has been a Faculty Fellow of the Mind/Brain/Behavior Interfaculty Initiative of Harvard University, and was a member of the work group to revise the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM-5*). He has been a visiting professor at the Georgetown University School of Medicine, the University of Wisconsin Medical School, the University of Illinois College of Medicine, Dartmouth Medical School, and the Allegheny University of the Health Sciences. He is a Distinguished Life Fellow of the American Psychiatric Association, a Fellow of the American College of Psychiatrists, and served on the Council of the American Psychosomatic Society. Dr. Barsky graduated from Williams College and the Columbia University College of Physicians and Surgeons. He interned at the Beth Israel Medical Center in New York City and completed a residency in psychiatry at the Massachusetts General Hospital in Boston, where he remained on the full-time faculty until 1993 when he moved to the Brigham and Women's Hospital.

Mary C. Daly, Ph.D., is Senior Vice President and Associate Director of Economic Research at the Federal Reserve Bank of San Francisco. Dr. Daly's

research spans public finance, labor, and welfare economics, and she has published widely on topics related to labor market fluctuations, public policy, income inequality, and the economic well-being of less advantaged groups. She previously served as a visiting scholar with the Congressional Budget Office, as a member of the Social Security Advisory Board's Technical Panel, and the National Academy of Social Insurance Committee on the Privatization of the Social Security Retirement Program. She has published on the economics of the Social Security system. She currently serves on the editorial board of the journal *Industrial Relations*. Dr. Daly joined the Federal Reserve as an Economist in 1996 after completing a National Institute on Aging postdoctoral fellowship at Northwestern University. Dr. Daly earned a Ph.D. in Economics from Syracuse University. She joined the Institute for the Study of Labor (IZA) as a Research Fellow in February 2014.

Kurt F. Geisinger, Ph.D., is Director of the Buros Center on Testing and WC Meierhenry Distinguished University Professor at the University of Nebraska. He previously was Professor and Chair of the Department of Psychology at Fordham University, Professor of Psychology and Dean of Arts and Sciences at the State University of New York at Oswego (SUNY-Oswego), Professor of Psychology and Academic Vice President at LeMoyne College, and Professor of Psychology and Vice President for Academic Affairs at the University of St. Thomas, in Houston, Texas. He has served the maximum two terms as council representative for the Division of Measurement, Evaluation, and Statistics in the American Psychological Association (APA), which he also represented on the International Organization for Standardization's (ISO's) International Test Standards committee. He was elected President of the Coalition for Academic, Scientific, and Applied Psychology for the 2009 year, to the board of the International Test Commission, and to the American Psychological Association's Board of Directors. He currently serves as Treasurer for the International Test Commission. His primary interests lie in validity theory, admissions testing, proper test use, test use with individuals with disabilities, the testing of language minorities, and the translation or adaptation of tests from one language and culture to another. Previously Dr. Geisinger was an APA delegate and chair of the Joint Committee on Testing Practices (1992–1996), a member of APA's Committee on Psychological Testing and Assessment, Chair of the Graduate Record Examination Board, Chair of the Technical Advisory Committee for the Graduate Record Examination, a member of the SAT Advisory Committee, a member of National Council on Measurement in Education's (NCME's) Ad Hoc Committee to Develop a Code of Ethical Standards Committee, and has served on numerous other ad hoc task forces and panels. He chaired the College Board's Research and Development Committee and is currently Chair of the Council for

the Accreditation of Educator Preparation's Research Committee, having served on their Commission on Standards and Performance Reporting. He is editor of *Applied Measurement in Education* and serves or has served on the editorial committees for the eight other journals. He has edited or co-edited the *Psychological Testing of Hispanics* and *Test Interpretation and Diversity*, both with APA Books®, as well as the 17th, 18th, and 19th Mental Measurements Yearbooks. He served as editor-in-chief for the *Handbook of Testing and Assessment in Psychology*, published by APA Books in 2013 and his vastly revised volume, *Psychological Testing of Hispanics: Clinical and Intellectual Issues* is in press, also with APA Books.

Naomi Lynn Gerber, M.D., is University Professor and Director of the Center for the Study of Chronic Illness and Disability in the College of Health and Human Services at George Mason University. She works in the areas of measurement and treatment of impairments and disability in patients with musculoskeletal deficits (including children with osteogenesis imperfecta; persons with rheumatoid arthritis and cancer). Her research investigates causes of functional loss and disability in chronic illness. Specifically, she studies human movement and the mechanisms and treatment of fatigue. Dr. Gerber is/has been a recipient of National Science Foundation, PNC Foundation, National Institute on Disability and Rehabilitation Research (NIDRR), National Institutes of Health (NIH), and Department of Defense funding administered by the Henry Jackson Foundation. She was the Chief of the Rehabilitation Medicine Department at the Clinical Center of NIH in Bethesda, Maryland, from 1975 to 2005. She has been the recipient of the Distinguished Service Award of the American Academy of Physical Medicine and Rehabilitation (AAPMR) and the Oncology Section of American Physical Therapy Association, the Distinguished Academician Award of the Association of Academic Physiatrists, the WISE/Geico award, NIH Directors Award, Surgeon General Award for Exemplary Service, and the Smith College Medal. Dr. Gerber has served on many national committees and advisory boards including Osteogenesis Imperfecta Foundation (1995–present), Kessler Medical Rehabilitation Research (2001–present), National Center for Medical Rehabilitation Research, (2007–2011), Blue Ribbon Panel Assessing Rehabilitation/Research, NIH (2011–2012). She is/has been a grant reviewer for NIDRR, NIH, National Science Foundation, and the Veterans Affairs. She served on the Board of Governors of the AAPMR (2005–2008). Dr. Gerber is a member of the Institute of Medicine of the National Academy of Sciences. In 2013 she delivered the Zeiter Lecture at the AAPMR 75th anniversary. Dr. Gerber is a graduate of Tufts University School of Medicine, diplomate of the American Board of Internal Medicine, Rheumatology sub-specialty, and the American Board of Physical Medicine and Rehabilitation.

Alan M. Jette, P.T., M.P.H., Ph.D., is Professor of Health Policy and Management at the Boston University School of Public Health. Dr. Jette is an international expert in the measurement and evaluation of functioning and health outcomes and in the measurement, epidemiology, and prevention of disability. His work has addressed the need to bring conceptual clarity to the measurement of patient-centered outcomes in a range of challenging clinical areas such as work disability, spinal cord injury, and neurologic, orthopedic, and geriatric conditions. He chaired the Institute of Medicine (IOM) panel that authored the 2007 IOM report, *The Future of Disability in America*, and currently co-chairs the IOM Forum on Aging, Disability, and Independence. Dr. Jette received a B.S. in Physical Therapy from the State University of New York at Buffalo in 1973 and his M.P.H. (1975) and Ph.D. (1979) in Public Health from the University of Michigan.

Jennifer I. Koop, Ph.D., is an Associate Professor in the Department of Neurology (Neuropsychology) at the Medical College of Wisconsin, with a secondary appointment of Associate Professor in the Department of Pediatrics. Dr. Koop specializes in the evaluation and treatment of children with neurological, behavioral, and developmental disorders. Her current research investigates the effects of early neurological injury on the development of neuropsychological functions, especially attention. She received her Ph.D. in Clinical Rehabilitation Psychology, with a specialization in neuropsychology, from Indiana University–Purdue University Indianapolis. She completed a pre-doctoral internship at Texas Children's Hospital/Baylor College of Medicine and 2-year postdoctoral fellowship in pediatric neuropsychology at the Medical College of Wisconsin. She is board certified in clinical neuropsychology by the American Board of Professional Psychology.

Lisa A. Suzuki, Ph.D., is Associate Professor in the Department of Applied Psychology at the Steinhardt School of Culture, Education, and Human Development of New York University. Prior to this, she served as a faculty member in counseling psychology at Fordham University and the University of Oregon. Dr. Suzuki received the Distinguished Contribution Award from the Asian American Psychological Association in 2006 and Visionary Leadership Award from the National Multicultural Conference and Summit in 2007. She has written extensively in the area of multicultural issues in psychological assessment, and her work appears in chapters of the *Handbook of Multicultural Counseling*, *American Psychological Association (APA) Handbook of Testing and Psychology*, *APA Handbook of Counseling Psychology*, *Handbook of Psychology*, *APA Handbook of Multicultural Psychology*, and the *Cambridge Handbook of Intelligence*. She is senior editor of the *Handbook of Multicultural Assessment* and a

co-editor of the *Handbook of Multicultural Counseling*. She is co-author of *Intelligence Testing and Minority Students* (Valencia and Suzuki, 2001). Dr. Suzuki obtained her Ph.D. from the University of Nebraska–Lincoln, in 1992.

Elizabeth W. Twamley, Ph.D., is Associate Professor of Psychiatry in Residence at the University of California, San Diego (UCSD), and Research Psychologist in the Center of Excellence for Stress and Mental Health (CESAMH) at the Veterans Affairs San Diego Healthcare System. As a licensed clinical psychologist, she specializes in neuropsychological assessment, cognitive rehabilitation, and supported employment. Dr. Twamley is particularly interested in community-based interventions that help individuals with severe mental illness or other cognitive impairments reach their highest potential social and occupational functioning. She supervises psychology interns and practicum students at UCSD Outpatient Psychiatric Services and the Veterans Affairs San Diego Healthcare System. She also conducts a neuropsychological assessment clinic at the St. Vincent De Paul Medical Clinic. Dr. Twamley's research focuses on bridging neuropsychology and interventions for individuals with severe mental illness or traumatic brain injury. Current intervention studies focus on supported employment and compensatory cognitive training. Other research interests include the neuropsychology of everyday functioning, genetic markers of cognition in schizophrenia, and cognitive impairment in posttraumatic stress disorder (PTSD). Dr. Twamley earned a B.A. in Social Ecology at University of California, Irvine, and a Ph.D. in Clinical Psychology from Arizona State University. She completed her clinical psychology internship and postdoctoral fellowship at UCSD and joined the faculty of the Department of Psychiatry in 2003.

Peter A. Ubel, M.D., is the Madge and Dennis T. McLawhorn University Professor of Business at the Fuqua School of Business and Professor of Public Policy at the Sanford School of Public Policy at Duke University. He is a physician and behavioral scientist specializing in health policy and economics, whose research and writing explores the mixture of rational and irrational forces that affect health, happiness, and the way society functions. His research explores controversial issues about the role of values and preferences in health care decision making, from decisions at the bedside to policy decisions. He uses the tools of decision psychology and behavioral economics to explore topics like informed consent, shared decision making and health care cost containment. His books include *Pricing Life: Why It's Time for Healthcare Rationing* (MIT Press, 2000) and *Free Market Madness: How Economics Is at Odds with Human Nature—and Why It Matters* (Harvard Business Press, 2009). His newest book, *Critical*

Decisions (HarperCollins, 2012) explores the challenges of shared decision making between doctors and patients. Dr. Ubel previously was Professor of Medicine and Psychology at the University of Michigan, where he taught from 2000 to 2010, and later went on to direct the Center for Behavioral and Decision Sciences in Medicine. Dr. Ubel received his B.A. from Carleton College and his M.D. from the University of Minnesota.

Jacqueline Remondet Wall, Ph.D., is Professor in the School of Psychological Sciences at the University of Indianapolis and Director of the Office of Program Consultation and Accreditation at the American Psychological Association in Washington, DC, where she is an Associate Executive Director in the Education Directorate. Her professional and research interests include assessment, selection, training, and evaluation. Dr. Wall received her Ph.D. from the University of Tulsa with a specialization in industrial and organizational psychology and a post-doctoral respecialization in clinical rehabilitation and neuropsychology at the Illinois Institute of Technology, the Medical School of the University of Mississippi, and the Rehabilitation Institute of Michigan.

C

Glossary

Activity limitations: difficulties an individual may have in executing activities (IOM, 2007; WHO, 2001)

Clinical neuropsychology: specialty in professional psychology that applies principles of assessment and intervention based on the scientific study of human behavior as it relates to normal and abnormal functioning of the central nervous system (APA, 2010)

Clinical psychology: specialty in professional psychology focused on assessment, diagnosis, prediction, prevention, and treatment of psychopathology, mental disorders, and other individual or group problems to improve behavior adjustment, adaptation, personal effectiveness, and satisfaction (APA, 2014)

Cognitive test: standardized measure of task performance used to assess cognitive functioning (e.g., intellectual capacity, attention and concentration, processing speed, language and communication, visual-spatial abilities, memory)

Disability: decrements in all three aspects of human functioning (body functions and structures, activities, and participation), which are labeled *impairments*, *activity limitations*, and *participation restrictions* (IOM, 2007;

WHO, 2001); the limitation on an individual's abilities to perform certain activities of daily life (e.g., school- or work-related, personal care, social interactions)

Disability (Social Security Administration): in adults, "the inability to engage in any substantial gainful activity … by reason of any medically determinable physical or mental impairment(s) which can be expected to result in death or which has lasted or can be expected to last for a continuous period of not less than 12 months"; in children, "a medically determinable physical or mental impairment or combination of impairments that causes marked and severe functional limitations, and that can be expected to cause death or that has lasted or can be expected to last for a continuous period of not less than 12 months" (SSA, n.d., see also 2012).

Effort: the extent to which the examinee performed to actual capacity on a test (Bush et al., 2005)

Functional limitation: a loss or restriction of an individual's ability to perform a specific physical or mental function or activity, such as walking, speaking, memory, and the like (IOM, 2007)

Impairment: problems in body function or structure such as a significant deviation or loss (IOM, 2007; WHO, 2001)

Malingering: the intentional presentation of false or exaggerated symptoms, intentionally poor performance, or a combination of the two, motivated by external incentives (American Psychiatric Association, 2013; Bush et al., 2005; Heilbronner et al., 2009)

Medically determinable impairment: "an impairment that results from anatomical, physiological, or psychological abnormalities which can be shown by medically acceptable clinical and laboratory diagnostic techniques" (SSA, n.d.)

Neuropsychological tests: performance-based tests by which various aspects of an individual's cognitive functioning can be measured (Larrabee, 2012, 2014)

Non-cognitive measure: standardized self-report measure that assesses non-cognitive psychological complaints

Participation restriction: problems an individual may experience in involvement in life situations (IOM, 2007; WHO, 2001)

Performance validity: the validity of actual ability task performance; often referred to as effort in the literature (Larrabee, 2012, 2014)

Performance validity test: stand-alone or embedded/derived measures used to assess whether an examinee is performing at a level consistent with his/her actual abilities (Larrabee, 2012, 2014)

Psychological assessment: the comprehensive integration of information from a variety of sources—including formal psychological tests, informal tests and surveys, structured clinical interviews, interviews with others, school and/or medical records, and observational data—to make inferences regarding the mental or behavioral characteristics of an individual or to predict behavior (Furr and Bacharach, 2013; Hubley and Zumbo, 2013)

Psychological testing: the use of formal, standardized procedures for sampling behavior that ensure objective evaluation of the test-taker regardless of who administers the test (Furr and Bacharach, 2013; Hubley and Zumbo, 2013). Major categories of psychological tests include (1) intelligence tests, (2) neuropsychological tests, (3) personality tests, (4) clinical or diagnostic tests (e.g., depression, anxiety), (5) achievement tests, (6) aptitude tests, and (7) occupational or interests tests

Psychometrics: the scientific study, including the development, interpretation, and evaluation, of psychological tests and measures used to assess variability in behavior and link such variability to psychological phenomena (Furr and Bacharach, 2013; Hubley and Zumbo, 2013)

Reliability: the degree to which a test produces stable and consistent results (Geisinger, 2013)

Response bias: misrepresentation of abilities in any neuropsychological domain of ability through performance, or self-report regarding performance capabilities (Heilbronner et al., 2009)

Self-report measure: standardized instruments that rely on self-report with population-based normative data that allow the examiner to compare an individual's reported behaviors or symptoms with an appropriate comparison group

Self-report of symptoms: the claimant's own description of his or her physical or mental impairment; in some cases, symptoms may be reported by a third party (e.g., children's symptoms may be reported by parent or teacher) (20 CFR § 404.1528)

Substantial gainful activity: "work that involves doing significant and productive physical or mental duties and is done (or intended) for pay or profit" (20 CFR § 416.910)

Symptom exaggeration: over-reporting of symptoms (Mittenberg et al., 2002)

Symptom validity: the accuracy of symptomatic complaint (Larrabee, 2012, 2014)

Symptom validity test: embedded or stand-alone measures used to assess whether an examinee is providing an accurate report of his or her actual symptom experience on non-cognitive psychological measures (e.g., emotional, behavioral, and personality measures) (Larrabee, 2012, 2014)

Validity: the degree to which evidence and theory support the use and interpretation of test scores (AERA et al., 2014)

REFERENCES

AERA (American Educational Research Association), APA (American Psychology Association), and NCME (National Council on Measurement in Education). 2014. *Standards for educational and psychological testing*. Washington, DC: AERA.

American Psychiatric Association. 2013. *The diagnostic and statistical manual of mental disorders: DSM-5*. Washington, DC: American Psychiatric Association.

APA (American Psychological Association). 2010. *Public description of clinical neuropsychology*. http://www.apa.org/ed/graduate/specialize/neuro.aspx (accessed June 24, 2014).

APA. 2014. *Public description of clinical psychology*. http://www.apa.org/ed/graduate/specialize/clinical.aspx (accessed June 24, 2014).

Bush, S. S., R. M. Ruff, A. I. Tröster, J. T. Barth, S. P. Koffler, N. H. Pliskin, C. R. Reynolds, and C. H. Silver. 2005. Symptom validity assessment: Practice issues and medical necessity. NAN policy & planning committee. *Archives of Clinical Neuropsychology* 20(4):419-426.

Furr, R. M., and V. R. Bacharach. 2013. *Psychometrics: An introduction.* Thousand Oaks, CA: Sage Publications, Inc.

Geisinger, K. F. 2013. Reliability. *In APA handbook of testing and assessment in psychology*. Vol. 1, edited by K. F. Geisinger (editor) and B. A. Bracken, J. F. Carlson, J. C. Hansen, N. R. Kuncel, S. P. Reise, and M. C. Rodriguez (associate editors). Washington, DC: APA.

Heilbronner, R. L., J. J. Sweet, J. E. Morgan, G. J. Larrabee, S. R. Millis, and Conference participants. 2009. American Academy of Clinical Neuropsychology consensus conference statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologis*t 23(7):1093-1129.

Hubley, A. M., and B. D. Zumbo. 2013. Psychometric characteristics of assessment procedures: An overview. In *APA handbook of testing and assessment in psychology*. 3 vols. Vol. 1, edited by K. F. Geisinger. Washington, DC: American Psychological Association.

IOM (Institute of Medicine). 2007. *The future of disability in America*. Washington, DC: The National Academies Press.

Larrabee, G. J. 2012. Performance validity and symptom validity in neuropsychological assessment. *Journal of the International Neuropsychological Society* 18(4):625-630.

Larrabee, G. J. 2014. *Performance and Symptom Validity*. Presentation to IOM Committee on Psychological Testing, Including Validity Testing, for Social Security Administration, June 25, 2014, Washington, DC.

Mittenberg, W., C. Patton, E. M. Canyock, and D. C. Condit. 2002. Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology* 24(8):1094-1102.

SSA (Social Security Administration). 2012. DI 00115.015 *Definitions of disability*. https://secure.ssa.gov/poms.nsf/lnx/0400115015 (accessed October 3, 2014).

SSA. n.d. *Disability evaluation under social security; Part I—General information*. http://www.ssa.gov/disability/professionals/bluebook/general-info.htm (accessed November 14, 2014).

WHO (World Health Organization). 2001. *International classification of functioning, disability and health (ICF)*. Geneva: WHO.