

PLEASE DO POWERPOINT FOR ASSESSMENT
AND DISCUSSION FOR CASE DESCRIPTION

INTERACTIVE ASSIGNMENT 4

Prior to beginning work on this discussion, read Chapters 8 and 9 in the textbook and carefully review Section 4 of Chapter 7 in the U.S. Department of Labor Employment and Training Administration (2006) guide, Testing and Assessment: A Guide to Good Practices for Workforce Investment Professionals (Links to an external site.). In addition, review all the required articles for this week, as well as the Case Description: Mr. I—Psychiatric Inpatient Interpretive Report (Links to an external site.) and the PSY640 Week Four Psychological Assessment Report (Links to an external site.).

For this discussion, you will take on the role of a psychologist who was recently assigned two new clients. You will examine psychological assessment information presented in two different formats: a computer generated interpretative report of personality test results for the patient Mr. I and a psychological report written by a licensed psychologist for the patient Ms. S. In your initial post, you will examine the personality assessment instruments used in each report.

Carefully review the Case Description: Mr. I—Psychiatric Inpatient Interpretive Report (Links to an external site.); this patient was referred to you after being admitted into a psychiatric inpatient facility. Write a one-paragraph summary of the computer generated MMPI-2-RF results for Mr. I based on the information in the interpretive report.

In your role as the psychologist who evaluated Ms. S., examine the personality and ability testing results in the PSY640 Week Four Psychological Assessment Report (Links to an external site.). In your next meeting with Ms. S, you will be required to give her a copy of the psychological assessment report and discuss the results with her by explaining the psychological concepts effectively observing appropriate professional standards. In order to share this discussion with your colleagues, you will create a screencast of a three- to five-minute assessment feedback session, which must walk the client through the report and summarize the most pertinent information from the psychological assessment report in language your client can understand. You may use any screen-casting software you choose. Quick-Start Guides are available for Screencast-O-Matic (Links to an external site.) for your convenience. Once you have created your screencast, include the link in your initial post.

In your initial post, provide an evaluation of the contents of both psychological evaluations in terms of ethical standards and the professionalism of the interpretation of the testing and assessment data presented. Write an analysis of the psychometric methodologies employed in the development and validation of the MMPI-2-RF personality test used with both clients. Develop a list of at least two additional tests of personality or emotional functioning to administer to the two clients that demonstrate acceptable validity. Justify your inclusion of each additional assessment measure in terms of the validity of the assessment measure and your clients' presenting concerns, diagnosis, and prognosis.

Note to Students: This assignment requires that you produce a visual presentation, supply a spoken audio narrative, and to listen the audio narrative of others. Note you are also asked to provide a transcript of your presentation. If you have a documented disability accommodation that might interfere with your ability to complete this assignment you may contact your instructor to develop a comparable alternative

assignment. If you have other issues that you feel may be a barrier to your ability to complete this course or this assignment please contact the Office of Access and Wellness at: access@ashford.edu.



Minnesota Multiphasic
Personality Inventory-2
Restructured Form®

SAMPLE REPORT

Case Description: Mr. I — Psychiatric Inpatient Interpretive Report

Mr. I is a 46-year-old, married man admitted for inpatient treatment after presenting with psychotic thinking and assaultive behavior. At intake, he described a recent pattern of decreased sleep and presented with bizarre delusional thinking, religious preoccupation, visual hallucinations, and tangential and circumstantial thinking. He had previously been diagnosed with Schizophrenia and Schizoaffective Disorder.

Case descriptions do not accompany MMPI-2-RF reports, but are provided here as background information. The following report was generated from Q-global™, Pearson's web-based scoring and reporting application, using Mr. I's responses to the MMPI-2-RF. Additional MMPI-2-RF sample reports, product offerings, training opportunities, and resources can be found at PearsonClinical.com/mmpi2rf.

Copyright © 2014 Pearson Education, Inc. or its affiliate(s). All rights reserved. Q-global, Always Learning, Pearson, design for Psi, and PsychCorp are trademarks, in the U.S. and/or other countries, of Pearson Education, Inc. or its affiliate(s). Minnesota Multiphasic Personality Inventory-2 Restructured Form and MMPI-2-RF are registered trademarks of the University of Minnesota, Minneapolis, MN. 8795-A 01/14



Minnesota Multiphasic
Personality Inventory-2
Restructured Form®

Interpretive Report: Clinical Settings

MMPI-2-RF®

Minnesota Multiphasic Personality Inventory-2-Restructured Form®

Yossef S. Ben-Porath, PhD, & Auke Tellegen, PhD

ID Number:	Mr. I
Age:	36
Gender:	Male
Marital Status:	Married
Years of Education:	Not reported
Date Assessed:	1/13/14



Copyright © 2008, 2011, 2012 by the Regents of the University of Minnesota. All rights reserved.
Distributed exclusively under license from the University of Minnesota by NCS Pearson, Inc. Portions reproduced from the MMPI-2-RF test booklet. Copyright © 2008 by the Regents of the University of Minnesota. All rights reserved. Portions excerpted from the *MMPI-2-RF Manual for Administration, Scoring, and Interpretation*. Copyright © 2008, 2011 by the Regents of the University of Minnesota. All rights reserved.
Used by permission of the University of Minnesota Press.

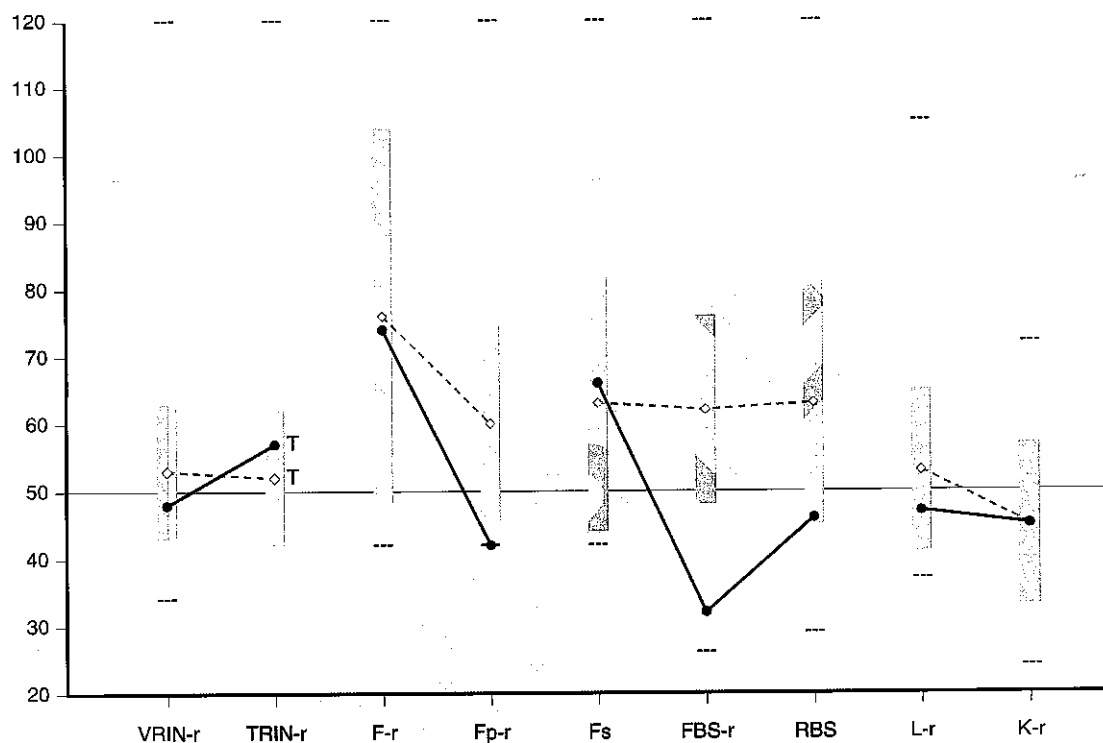
MMPI-2-RF, the MMPI-2-RF logo, and Minnesota Multiphasic Personality Inventory-2-Restructured Form are registered trademarks of the University of Minnesota. Pearson, the PSI logo, and PsychCorp are trademarks in the U.S. and/or other countries of Pearson Education, Inc., or its affiliate(s).

TRADE SECRET INFORMATION

Not for release under HIPAA or other data disclosure laws that exempt trade secrets from disclosure.

[2.2 / 1 / QG]

MMPI-2-RF Validity Scales



Raw Score:	3	12	7	0	3	2	4	2	6
T Score:	48	57 T	74	42	66	32	46	47	45
Response %:	94	100	97	95	94	87	100	100	86
Cannot Say (Raw):	17								
					Percent True (of items answered):				52%

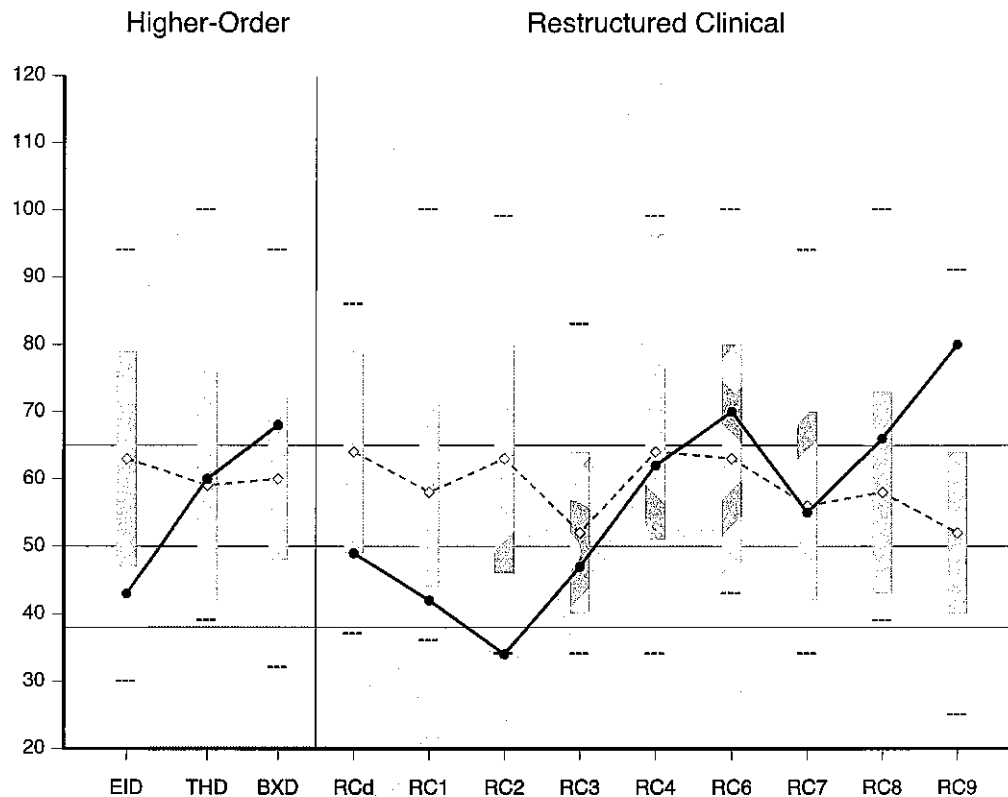
Comparison Group Data: Psychiatric Inpatient, Community Hospital (Men), N = 659

Mean Score (◇---◇):	53	52 T	76	60	63	62	63	53	45
Standard Dev (±1 sd):	10	10	28	15	19	14	18	12	12
Percent scoring at or below test taker:	45	76	58	25	73	0.9	22	45	59

The highest and lowest T scores possible on each scale are indicated by a "----"; MMPI-2-RF T scores are non-gendered.

VRIN-r	Variable Response Inconsistency	Fs	Infrequent Somatic Responses	L-r	Uncommon Virtues
TRIN-r	True Response Inconsistency	FBS-r	Symptom Validity	K-r	Adjustment Validity
F-r	Infrequent Responses	RBS	Response Bias Scale		
Fp-r	Infrequent Psychopathology Responses				

MMPI-2-RF Higher-Order (H-O) and Restructured Clinical (RC) Scales



Raw Score:	5	4	13	3	1	0	5	9	4	9	6	24
T Score:	43	60	68	49	42	34	47	62	70	55	66	80
Response %:	100	96	96	100	93	100	47	100	94	96	94	96

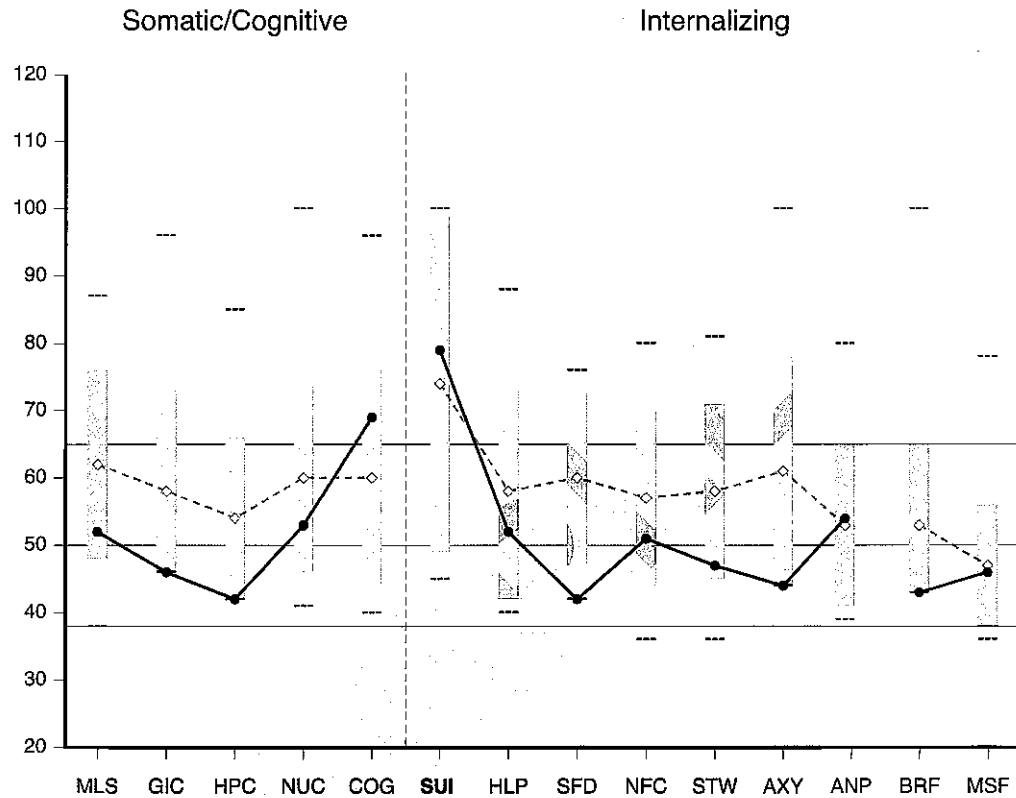
Comparison Group Data: Psychiatric Inpatient, Community Hospital (Men), N = 659

Mean Score (---◇):	63	59	60	64	58	63	52	64	63	56	58	52
Standard Dev (±1 SD):	16	17	12	15	14	17	12	13	17	14	15	12
Percent scoring at or below test taker:	14	66	75	22	19	3	44	48	75	55	76	99

The highest and lowest T scores possible on each scale are indicated by a "---"; MMPI-2-RF T scores are non-gendered.

EID	Emotional/Internalizing Dysfunction	RCd	Demoralization	RC6	Ideas of Persecution
THD	Thought Dysfunction	RC1	Somatic Complaints	RC7	Dysfunctional Negative Emotions
BXD	Behavioral/Externalizing Dysfunction	RC2	Low Positive Emotions	RC8	Aberrant Experiences
		RC3	Cynicism	RC9	Hypomanic Activation
		RC4	Antisocial Behavior		

MMPI-2-RF Somatic/Cognitive and Internalizing Scales



Raw Score:	2	0	0	1	5	2	1	0	3	2	0	3	0	2
T Score:	52	46	42	53	69	79	52	42	51	47	44	54	43	46
Response %:	100	100	100	90	100	100	100	100	100	100	100	100	100	100

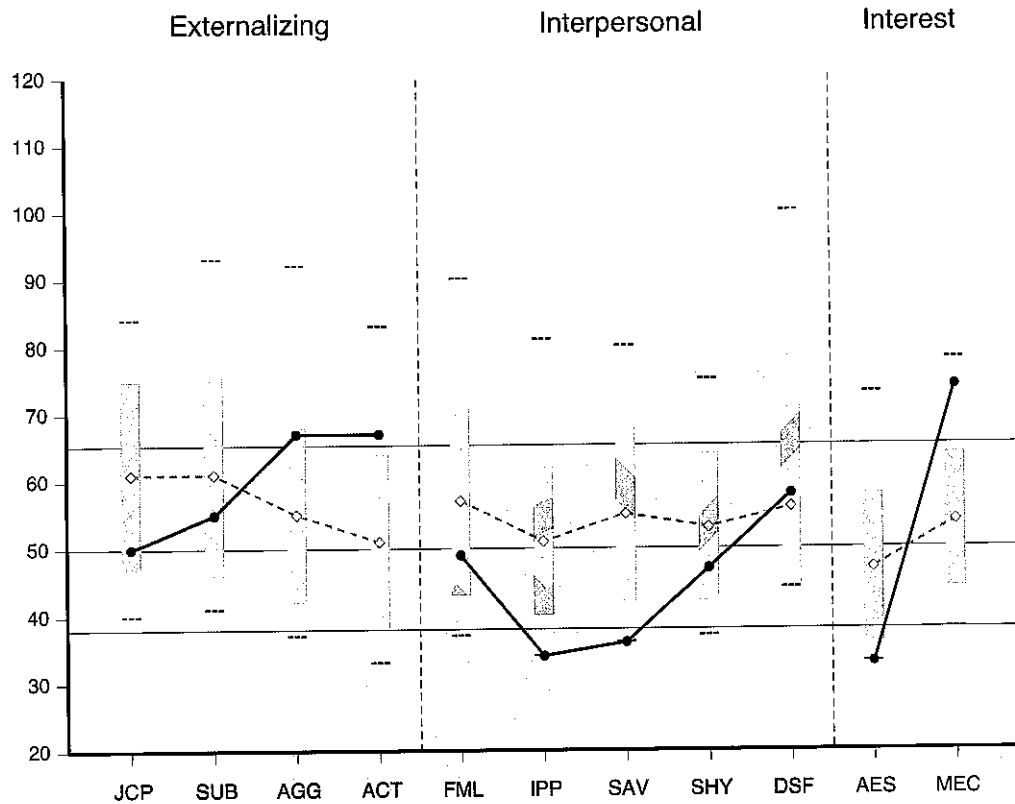
Comparison Group Data: Psychiatric Inpatient, Community Hospital (Men), N = 659

Mean Score (◇---◇):	62	58	54	60	60	74	58	60	57	58	61	53	53	47
Standard Dev (±1 SD):	14	15	12	14	16	25	16	13	13	13	17	12	12	9
Percent scoring at or below test taker:	34	53	37	44	73	64	51	25	44	31	37	67	51	55

The highest and lowest T scores possible on each scale are indicated by a "---"; MMPI-2-RF T scores are non-gendered.

MLS	Malaise	SUI	Suicidal/Death Ideation	AXY	Anxiety
GIC	Gastrointestinal Complaints	HLP	Helplessness/Hopelessness	ANP	Anger Proneness
HPC	Head Pain Complaints	SFD	Self-Doubt	BRF	Behavior-Restricting Fears
NUC	Neurological Complaints	NFC	Inefficacy	MSF	Multiple Specific Fears
COG	Cognitive Complaints	STW	Stress/Worry		

MMPI-2-RF Externalizing, Interpersonal, and Interest Scales



Raw Score:	1	2	5	6	2	0	0	2	1	0	8
T Score:	50	55	67	67	49	34	36	47	58	33	74
Response %:	100	100	100	100	100	100	100	100	100	86	100

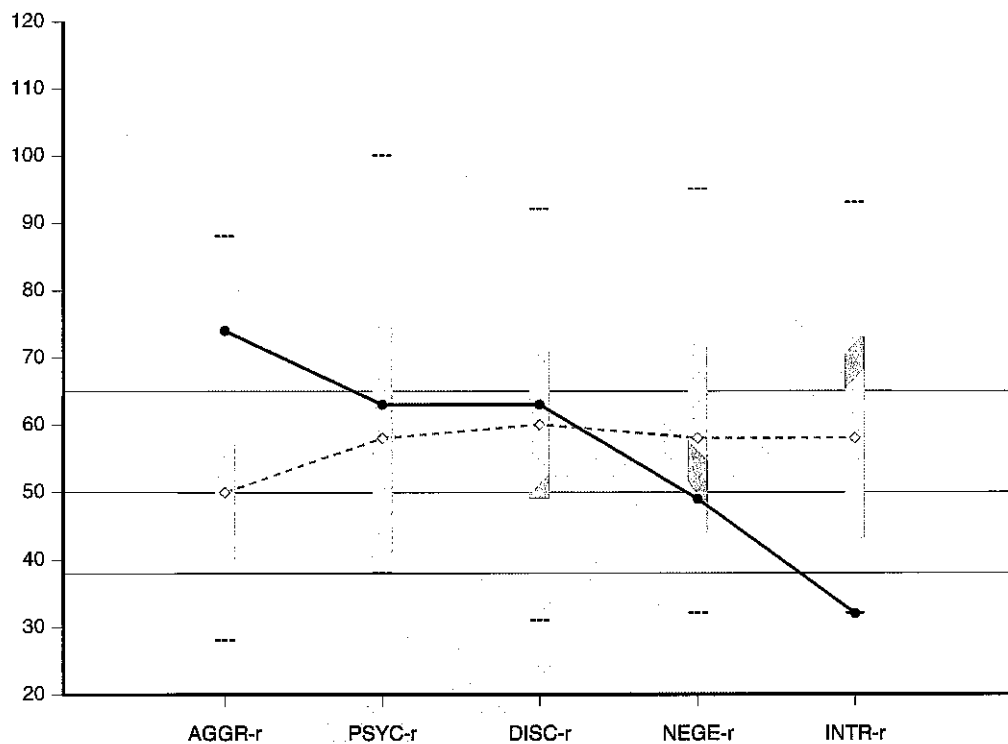
Comparison Group Data: Psychiatric Inpatient, Community Hospital (Men), N = 659

Mean Score (---):	61	61	55	51	57	51	55	53	56	47	54
Standard Dev (±1 SD):	14	15	13	13	14	11	13	11	15	11	10
Percent scoring at or below test taker:	30	50	86	91	40	7	12	39	72	16	99.2

The highest and lowest T scores possible on each scale are indicated by a "----"; MMPI-2-RF T scores are non-gendered.

JCP	Juvenile Conduct Problems	FML	Family Problems	AES	Aesthetic-Literary Interests
SUB	Substance Abuse	IPP	Interpersonal Passivity	MEC	Mechanical-Physical Interests
AGG	Aggression	SAV	Social Avoidance		
ACT	Activation	SHY	Shyness		
		DSF	Disaffiliativeness		

MMPI-2-RF PSY-5 Scales



Raw Score:	15	5	11	6	0
T Score:	74	63	63	49	32
Response %:	100	96	95	90	100

Comparison Group Data: Psychiatric Inpatient, Community Hospital (Men), N = 659

Mean Score (---):	50	58	60	58	58
Standard Dev (±1 SD):	10	17	11	14	15
Percent scoring at or below test taker:	99.1	71	64	33	2

The highest and lowest T scores possible on each scale are indicated by a "---"; MMPI-2-RF T scores are non-gendered.

AGGR-r	Aggressiveness-Revised
PSYC-r	Psychoticism-Revised
DISC-r	Disconstraint-Revised
NEGE-r	Negative Emotionality/Neuroticism-Revised
INTR-r	Introversion/Low Positive Emotionality-Revised

MMPI-2-RF T SCORES (BY DOMAIN)

PROTOCOL VALIDITY

Content Non-Responsiveness	17 CNS	48 VRIN-r	57 T TRIN-r			
Over-Reporting	74 F-r	42 Fp-r		66 Fs	32* FBS-r	46 RBS
Under-Reporting	47 L-r	45* K-r				

SUBSTANTIVE SCALES

Somatic/Cognitive Dysfunction		42 RC1	52 MLS	46 GIC	42 HPC	53 NUC	69 COG	
Emotional Dysfunction	43 EID	49 RCd	79 SUI	52 HLP	42 SFD	51 NFC		
		34 RC2	32 INTR-r					
		55 RC7	47 STW	44 AXY	54 ANP	43 BRF	46 MSF	49 NEGE-r
Thought Dysfunction	60 THD	70 RC6						
		66 RC8						
		63 PSYC-r						
Behavioral Dysfunction	68 BXD	62 RC4	50 JCP	55 SUB				
		80 RC9	67 AGG	67 ACT	74 AGGR-r	63 DISC-r		
Interpersonal Functioning		49 FML	47* RC3	34 IPP	36 SAV	47 SHY	58 DSF	
Interests		33* AES	74 MEC					

*The test taker provided scorable responses to less than 90% of the items scored on this scale. See the relevant profile page for the specific percentage.

Note. This information is provided to facilitate interpretation following the recommended structure for MMPI-2-RF interpretation in Chapter 5 of the *MMPI-2-RF Manual for Administration, Scoring, and Interpretation*, which provides details in the text and an outline in Table 5-1.

This interpretive report is intended for use by a professional qualified to interpret the MMPI-2-RF. The information it contains should be considered in the context of the test taker's background, the circumstances of the assessment, and other available information.

SYNOPSIS

Scores on the MMPI-2-RF validity scales raise concerns about the possible impact of unscorable responses on the validity of this protocol. With that caution noted, scores on the substantive scales indicate cognitive complaints and emotional, thought, behavioral, and interpersonal dysfunction. Cognitive complaints include difficulties in memory and concentration. Emotional-internalizing findings relate to **suicidal ideation**. Dysfunctional thinking includes ideas of persecution and aberrant perceptions and thoughts. Behavioral-externalizing problems include aggression and excessive activation. Interpersonal difficulties relate to over-assertiveness.

PROTOCOL VALIDITY

Content Non-Responsiveness

Unscorable Responses

The test taker answered less than 90% of the items on the following scales. The resulting scores may therefore be artificially lowered. In particular, the absence of elevation on these scales is not interpretable¹. A list of all items for which the test taker provided unscorable responses appears under the heading "Item-Level Information."

Symptom Validity (FBS-r): 87%
Adjustment Validity (K-r): 86%
Cynicism (RC3): 47%
Aesthetic-Literary Interests (AES): 86%

Inconsistent Responding

The test taker responded to the items in a consistent manner, indicating that he responded relevantly.

Over-Reporting

There are no indications of over-reporting in this protocol.

Under-Reporting

There are no indications of under-reporting in this protocol.

SUBSTANTIVE SCALE INTERPRETATION

Clinical symptoms, personality characteristics, and behavioral tendencies of the test taker are described in this section and organized according to an empirically guided framework. Statements containing the word "reports" are based on the item content of MMPI-2-RF scales, whereas statements that include the word "likely" are based on empirical correlates of scale scores. Specific sources for each statement can be viewed with the annotation features of this report.

The following interpretation needs to be considered in light of cautions noted about the possible impact of unscorable responses on the validity of this protocol.

Somatic/Cognitive Dysfunction

The test taker reports a diffuse pattern of cognitive difficulties². He is likely to complain about memory problems³, to have low tolerance for frustration⁴, not to cope well with stress⁴, and to experience difficulties in concentration⁵.

Emotional Dysfunction

The test taker reports a history of suicidal ideation and/or attempts⁶. He is likely to be preoccupied with suicide and death⁷ and to be at risk for current suicidal ideation and attempts⁷. This risk is exacerbated by poor impulse control⁸.

Thought Dysfunction

The test taker reports significant persecutory ideation such as believing that others seek to harm him⁹. He is likely to be suspicious of and alienated from others¹⁰, to experience interpersonal difficulties as a result of suspiciousness¹¹, and to lack insight¹¹.

He reports unusual thought processes¹². He is likely to experience thought disorganization¹³, to engage in unrealistic thinking¹⁴, and to believe he has unusual sensory-perceptual abilities¹⁵.

Behavioral Dysfunction

The test taker's responses indicate significant externalizing, acting-out behavior, which is likely to have gotten him into difficulties¹⁶. More specifically, he is very likely to be restless and become bored¹⁷ and to be acutely over-activated as manifested in aggression¹⁸, mood instability¹⁹, euphoria¹⁷, excitability²⁰, and sensation-seeking, risk-taking, or other forms of under-controlled, irresponsible behavior²¹. He reports episodes of heightened excitation and energy level²² and may have a history of symptoms associated with manic or hypomanic episodes²³. He also reports engaging in physically aggressive, violent behavior and losing control²⁴, and is indeed likely to have a history of violent behavior toward others²⁵.

Interpersonal Functioning Scales

The test taker describes himself as having strong opinions, as standing up for himself, as assertive and direct, and able to lead others²⁶. He is likely to believe he has leadership capabilities, but to be viewed by others as domineering, self-centered, and possibly grandiose²⁷. He also reports enjoying social situations and events²⁸, and is likely to be perceived as outgoing and gregarious²⁹.

Interest Scales

The test taker reports an above average number of interests in activities or occupations of a mechanical or physical nature (e.g., fixing and building things, the outdoors, sports)³⁰. Individuals who respond in this manner are likely to be adventure- and sensation-seeking³¹. The extent to which he lacks aesthetic or literary interests cannot be accurately gauged because of unscorable responses. There is possible evidence that he indicates little or no interest in activities or occupations of an aesthetic or literary nature (e.g., writing, music, the theater)³².

DIAGNOSTIC CONSIDERATIONS

This section provides recommendations for psychodiagnostic assessment based on the test taker's MMPI-2-RF results. It is recommended that he be evaluated for the following:

Emotional-Internalizing Disorders

- Cycling mood disorder³³

Thought Disorders

- Disorders involving persecutory ideation³⁴
- Disorders manifesting psychotic symptoms³⁵
- Personality disorders manifesting unusual thoughts and perceptions³⁶
- Schizoaffective disorder³⁷

Behavioral-Externalizing Disorders

- Manic or hypomanic episode or other conditions associated with excessive energy and activation³⁸
- Disorders associated with interpersonally aggressive behavior such as intermittent explosive disorder³⁹

TREATMENT CONSIDERATIONS

This section provides inferential treatment-related recommendations based on the test taker's MMPI-2-RF scores.

Areas for Further Evaluation

- **Risk for suicide should be assessed immediately**⁴⁰.
- May require inpatient treatment due to hypomania⁴¹.
- Need for mood-stabilizing medication⁴².
- Origin of cognitive complaints⁴³. May require a neuropsychological evaluation.

Psychotherapy Process Issues

- Persecutory ideation may interfere with forming a therapeutic relationship and treatment compliance⁴⁴.
- Impaired thinking may disrupt treatment³⁶.
- Unlikely to be internally motivated for treatment⁴⁵.
- At significant risk for treatment non-compliance⁴⁵.
- Excessive behavioral activation may interfere with treatment⁴².

Possible Targets for Treatment

- Mood stabilization in initial stages of treatment⁴¹
- Persecutory ideation⁴⁴
- Inadequate self-control⁴⁵
- Reduction in interpersonally aggressive behavior³⁹

ITEM-LEVEL INFORMATION

Unscorable Responses

Following is a list of items to which the test taker did not provide scorable responses. Unanswered or double answered (both True and False) items are unscorable. The scales on which the items appear are in parentheses following the item content.

- 9. Item Content Omitted. (RC7, NEGE-r)
- 15. Item Content Omitted. (Fs, FBS-r, RC1)
- 36. Item Content Omitted. (FBS-r, K-r, RC3)
- 55. Item Content Omitted. (VRIN-r, FBS-r, RC3)
- 99. Item Content Omitted. (VRIN-r, FBS-r, K-r, RC3)
- 107. Item Content Omitted. (BXD, RC9, DISC-r)
- 121. Item Content Omitted. (RC3)
- 185. Item Content Omitted. (RC3)
- 191. Item Content Omitted. (Fp-r)
- 194. Item Content Omitted. (VRIN-r, RC6)
- 203. Item Content Omitted. (F-r, THD, RC8, PSYC-r)
- 209. Item Content Omitted. (NEGE-r)
- 238. Item Content Omitted. (RC3)
- 296. Item Content Omitted. (AES)
- 304. Item Content Omitted. (RC3)
- 313. Item Content Omitted. (RC1, NUC)
- 326. Item Content Omitted. (RC3)



Special Note:

The content of the test items is included in the actual reports. To protect the integrity of the test, the item content does not appear in this sample report.

Critical Responses

Seven MMPI-2-RF scales--Suicidal/Death Ideation (SUI), Helplessness/Hopelessness (HLP), Anxiety (AXY), Ideas of Persecution (RC6), Aberrant Experiences (RC8), Substance Abuse (SUB), and Aggression (AGG)--have been designated by the test authors as having critical item content that may require immediate attention and follow-up. Items answered by the individual in the keyed direction (True or False) on a critical scale are listed below if his T score on that scale is 65 or higher. The percentage of the MMPI-2-RF normative sample (NS) and of the Psychiatric Inpatient, Community Hospital (Men) comparison group (CG) that answered each item in the keyed direction are provided in parentheses following the item content.

Suicidal/Death Ideation (SUI, T Score = 79)

251. Item Content Omitted. (True; NS 3.0%, CG 20.8%)

334. Item Content Omitted. (True; NS 13.5%, CG 35.5%)

Ideas of Persecution (RC6, T Score = 70)

14. Item Content Omitted. (True; NS 2.9%, CG 8.5%)

34. Item Content Omitted. (True; NS 10.6%, CG 27.3%)

71. Item Content Omitted. (True; NS 2.0%, CG 17.3%)

110. Item Content Omitted. (True; NS 9.9%, CG 32.5%)



Special Note:

The content of the test items is included in the actual reports. To protect the integrity of the test, the item content does not appear in this sample report.

Aberrant Experiences (RC8, T Score = 66)

32. Item Content Omitted. (True; NS 21.1%, CG 51.0%)

85. Item Content Omitted. (False; NS 17.1%, CG 35.2%)

106. Item Content Omitted. (True; NS 8.7%, CG 31.7%)

159. Item Content Omitted. (True; NS 6.0%, CG 27.0%)

240. Item Content Omitted. (True; NS 8.8%, CG 23.2%)

257. Item Content Omitted. (True; NS 12.4%, CG 37.0%)

Aggression (AGG, T Score = 67)

23. Item Content Omitted. (True; NS 39.0%, CG 46.3%)

312. Item Content Omitted. (True; NS 5.5%, CG 25.8%)

316. Item Content Omitted. (True; NS 45.1%, CG 50.5%)

329. Item Content Omitted. (True; NS 12.7%, CG 29.3%)

337. Item Content Omitted. (True; NS 50.2%, CG 52.2%)

User-Designated Item-Level Information

The following item-level information is based on the report user's selection of additional scales, and/or of lower cutoffs for the critical scales from the previous section. Items answered by the test taker in the keyed direction (True or False) on a selected scale are listed below if his T score on that scale is at the user-designated cutoff score or higher. The percentage of the MMPI-2-RF normative sample (NS) and of the Psychiatric Inpatient, Community Hospital (Men) comparison group (CG) that answered each item in the keyed direction are provided in parentheses following the item content.

Hypomanic Activation (RC9, T Score = 80)

- 13. Item Content Omitted. (True; NS 40.9%, CG 43.4%)
- 39. Item Content Omitted. (True; NS 51.0%, CG 53.3%)
- 47. Item Content Omitted. (True; NS 42.7%, CG 45.7%)
- 61. Item Content Omitted. (False; NS 61.6%, CG 73.4%)
- 72. Item Content Omitted. (True; NS 81.5%, CG 69.3%)
- 97. Item Content Omitted. (True; NS 50.5%, CG 45.2%)
- 118. Item Content Omitted. (True; NS 57.4%, CG 61.3%)
- 131. Item Content Omitted. (True; NS 43.3%, CG 47.0%)
- 143. Item Content Omitted. (True; NS 27.5%, CG 32.3%)
- 155. Item Content Omitted. (True; NS 41.6%, CG 37.9%)
- 166. Item Content Omitted. (True; NS 38.9%, CG 31.7%)
- 181. Item Content Omitted. (True; NS 35.3%, CG 36.7%)
- 193. Item Content Omitted. (True; NS 32.8%, CG 38.2%)
- 207. Item Content Omitted. (True; NS 66.9%, CG 47.3%)
- 219. Item Content Omitted. (True; NS 51.5%, CG 54.9%)
- 244. Item Content Omitted. (True; NS 56.9%, CG 64.5%)
- 248. Item Content Omitted. (True; NS 16.1%, CG 25.6%)
- 256. Item Content Omitted. (True; NS 65.7%, CG 58.1%)
- 267. Item Content Omitted. (True; NS 12.9%, CG 32.0%)
- 292. Item Content Omitted. (True; NS 26.1%, CG 30.3%)
- 305. Item Content Omitted. (True; NS 37.6%, CG 47.2%)
- 316. Item Content Omitted. (True; NS 45.1%, CG 50.5%)
- 327. Item Content Omitted. (True; NS 41.7%, CG 46.4%)
- 337. Item Content Omitted. (True; NS 50.2%, CG 52.2%)



Special Note:

The content of the test items is included in the actual reports. To protect the integrity of the test, the item content does not appear in this sample report.

Activation (ACT, T Score = 67)

- 72. Item Content Omitted. (True; NS 81.5%, CG 69.3%)
- 166. Item Content Omitted. (True; NS 38.9%, CG 31.7%)
- 181. Item Content Omitted. (True; NS 35.3%, CG 36.7%)
- 207. Item Content Omitted. (True; NS 66.9%, CG 47.3%)
- 219. Item Content Omitted. (True; NS 51.5%, CG 54.9%)
- 267. Item Content Omitted. (True; NS 12.9%, CG 32.0%)

ENDNOTES

This section lists for each statement in the report the MMPI-2-RF score(s) that triggered it. In addition, each statement is identified as a Test Response, if based on item content, a Correlate, if based on empirical correlates, or an Inference, if based on the report authors' judgment. (This information can also be accessed on-screen by placing the cursor on a given statement.) For correlate-based statements, research references (Ref. No.) are provided, keyed to the consecutively numbered reference list following the endnotes.

- ¹ Correlate: Response % < 90, Ref. 5
- ² Test Response: COG=69
- ³ Correlate: COG=69, Ref. 3, 10, 21
- ⁴ Correlate: COG=69, Ref. 21
- ⁵ Correlate: COG=69, Ref. 3, 21
- ⁶ Test Response: SUI=79
- ⁷ Correlate: SUI=79, Ref. 21
- ⁸ Inference: BXD=68; RC9=80
- ⁹ Test Response: RC6=70
- ¹⁰ Correlate: RC6=70, Ref. 1, 3, 4, 11, 15, 19, 21
- ¹¹ Correlate: RC6=70, Ref. 21
- ¹² Test Response: RC8=66
- ¹³ Correlate: RC8=66, Ref. 11, 21
- ¹⁴ Correlate: RC8=66, Ref. 3, 6, 7, 9, 21
- ¹⁵ Correlate: RC8=66, Ref. 6, 7, 9, 20, 21
- ¹⁶ Correlate: BXD=68, Ref. 13, 21
- ¹⁷ Correlate: RC9=80, Ref. 21
- ¹⁸ Correlate: RC9=80, Ref. 8, 15, 17, 18, 20, 21
- ¹⁹ Correlate: RC9=80, Ref. 3, 19, 21
- ²⁰ Correlate: RC9=80, Ref. 3, 11, 16, 21
- ²¹ Correlate: RC9=80, Ref. 16, 21
- ²² Test Response: ACT=67
- ²³ Correlate: RC9=80, Ref. 19, 21; ACT=67, Ref. 21, 23
- ²⁴ Test Response: AGG=67
- ²⁵ Correlate: RC9=80, Ref. 8, 15, 17, 18, 20, 21; AGG=67, Ref. 21
- ²⁶ Test Response: IPP=34
- ²⁷ Correlate: IPP=34, Ref. 2, 12, 21; AGGR-r=74, Ref. 21
- ²⁸ Test Response: SAV=36
- ²⁹ Correlate: SAV=36, Ref. 2, 21; INTR-r=32, Ref. 21
- ³⁰ Test Response: MEC=74
- ³¹ Correlate: MEC=74, Ref. 21
- ³² Test Response: AES=33
- ³³ Correlate: ACT=67, Ref. 23
- ³⁴ Correlate: RC6=70, Ref. 14, 22
- ³⁵ Correlate: RC8=66, Ref. 21
- ³⁶ Inference: RC8=66

³⁷ Inference: RC6=70; RC9=80

³⁸ Correlate: ACT=67, Ref. 14, 23

³⁹ Inference: AGG=67

⁴⁰ Inference: SUI=79

⁴¹ Inference: RC9=80

⁴² Inference: RC9=80; ACT=67

⁴³ Inference: COG=69

⁴⁴ Inference: RC6=70

⁴⁵ Inference: BXD=68

RESEARCH REFERENCE LIST

1. Arbisi, P. A., Sellbom, M., & Ben-Porath, Y. S. (2008). Empirical correlates of the MMPI-2 Restructured Clinical (RC) Scales in psychiatric inpatients. *Journal of Personality Assessment*, 90, 122-128. doi: [10.1080/00223890701845146](https://doi.org/10.1080/00223890701845146)
2. Ayearst, L. E., Sellbom, M., Trobst, K. K., & Bagby, R. M. (2013). Evaluating the interpersonal content of the MMPI-2-RF Interpersonal Scales. *Journal of Personality Assessment*, 95, 187-196. doi: [10.1080/00223891.2012.730085](https://doi.org/10.1080/00223891.2012.730085)
3. Burchett, D. L., & Ben-Porath, Y. S. (2010). The impact of over-reporting on MMPI-2-RF substantive scale score validity. *Assessment*, 17, 497-516. doi: [10.1177/1073191110378972](https://doi.org/10.1177/1073191110378972)
4. Cox, A., Pant, H., Gilson, A. N., Rodriguez, J. L., Young, K. R., Kwon, S., & Weed, N. C., (2012). Effects of augmenting response options on MMPI-2 RC Scale psychometrics. *Journal of Personality Assessment*, 94, 613-619. doi: [10.1080/00223891.2012.700464](https://doi.org/10.1080/00223891.2012.700464)
5. Dragon, W. R., Ben-Porath, Y. S., & Handel, R. H. (2012). Examining the impact of unscorable item responses on the validity and interpretability of MMPI-2/MMPI-2-RF Restructured Clinical (RC) Scale scores. *Assessment*, 19, 101-113. doi: [10.1177/1073191111415362](https://doi.org/10.1177/1073191111415362)
6. Forbey, J. D., Arbisi, P. A., & Ben-Porath, Y. S. (2012). The MMPI-2 computer adaptive version (MMPI-2-CA) in a VA medical outpatient facility. *Psychological Assessment*, 24, 628-639. doi: [10.1037/a0026509](https://doi.org/10.1037/a0026509)
7. Forbey, J. D., & Ben-Porath, Y. S. (2007). A comparison of the MMPI-2 Restructured Clinical (RC) and Clinical Scales in a substance abuse treatment sample. *Psychological Services*, 4, 46-58. doi: [10.1037/1541-1559.4.1.46](https://doi.org/10.1037/1541-1559.4.1.46)
8. Forbey, J. D., & Ben-Porath, Y. S. (2008). Empirical correlates of the MMPI-2 Restructured Clinical (RC) Scales in a non-clinical setting. *Journal of Personality Assessment*, 90, 136-141. doi: [10.1080/00223890701845161](https://doi.org/10.1080/00223890701845161)
9. Forbey, J. D., Ben-Porath, Y. S., & Gartland, D. (2009). Validation of the MMPI-2 Computerized Adaptive Version (MMPI-2-CA) in a correctional intake facility. *Psychological Services*, 6, 279-292. doi: [10.1037/a0016195](https://doi.org/10.1037/a0016195)
10. Gervais, R. O., Ben-Porath, Y. S., & Wygant, D. B. (2009). Empirical correlates and interpretation of the MMPI-2-RF Cognitive Complaints scale. *The Clinical Neuropsychologist*, 23, 996-1015. doi: [10.1080/13854040902748249](https://doi.org/10.1080/13854040902748249)
11. Handel, R. W., & Archer, R. P. (2008). An investigation of the psychometric properties of the MMPI-2 Restructured Clinical (RC) Scales with mental health inpatients. *Journal of Personality Assessment*, 90, 239-249. doi: [10.1080/00223890701884954](https://doi.org/10.1080/00223890701884954)
12. Kastner, R. M., Sellbom, M., & Lilienfeld, S. O. (2012). A comparison of the psychometric properties of the Psychopathic Personality Inventory full-length and short-form versions. *Psychological Assessment*, 24, 261-267. doi: [10.1037/a0025832](https://doi.org/10.1037/a0025832)

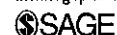
13. Lanyon, R. I., & Thomas, M. L. (2013). Assessment of global psychiatric categories: The PSI/PSI-2 and the MMPI-2-RF. *Psychological Assessment*, 25, 227-232. doi: [10.1037/a0030313](https://doi.org/10.1037/a0030313)
14. Sellbom, M., Bagby, R. M., Kushner, S., Quilty, L. C., & Ayearst, L. E. (2011). Diagnostic construct validity of the MMPI-2 Restructured Form (MMPI-2-RF) scale scores. *Assessment*, 19, 176-186. doi: [10.1177/1073191111428763](https://doi.org/10.1177/1073191111428763)
15. Sellbom, M., & Ben-Porath, Y. S. (2005). Mapping the MMPI-2 Restructured Clinical (RC) Scales onto normal personality traits: Evidence of construct validity. *Journal of Personality Assessment*, 85, 179-187. doi: [10.1207/s15327752jpa8502_10](https://doi.org/10.1207/s15327752jpa8502_10)
16. Sellbom, M., Ben-Porath, Y. S., & Bagby, R. M. (2008). Personality and psychopathology: Mapping the MMPI-2 Restructured Clinical (RC) Scales onto the five factor model of personality. *Journal of Personality Disorders*, 22, 291-312. doi: [10.1521/pedi.2008.22.3.291](https://doi.org/10.1521/pedi.2008.22.3.291)
17. Sellbom, M., Ben-Porath, Y. S., Baum, L. J., Erez, E., & Gregory, C. (2008). Predictive validity of the MMPI-2 Restructured Clinical (RC) Scales in a batterers' intervention program. *Journal of Personality Assessment*, 90, 129-135. doi: [10.1080/00223890701845153](https://doi.org/10.1080/00223890701845153)
18. Sellbom, M., Ben-Porath, Y. S., & Graham, J. R. (2006). Correlates of the MMPI-2 Restructured Clinical (RC) Scales in a college counseling setting. *Journal of Personality Assessment*, 86, 89-99. doi: [10.1207/s15327752jpa8601_10](https://doi.org/10.1207/s15327752jpa8601_10)
19. Sellbom, M., Graham, J. R., & Schenk, P. (2006). Incremental validity of the MMPI-2 Restructured Clinical (RC) Scales in a private practice sample. *Journal of Personality Assessment*, 86, 196-205. doi: [10.1207/s15327752jpa8602_09](https://doi.org/10.1207/s15327752jpa8602_09)
20. Simms, L. J., Casillas, A., Clark, L. A., Watson, D., & Doebbeling, B. I. (2005). Psychometric evaluation of the Restructured Clinical Scales of the MMPI-2. *Psychological Assessment*, 17, 345-358. doi: [10.1037/1040-3590.17.3.345](https://doi.org/10.1037/1040-3590.17.3.345)
21. Tellegen, A., & Ben-Porath, Y. S. (2008/2011). *The Minnesota Multiphasic Personality Inventory-2-Restructured Form (MMPI-2-RF): Technical manual*. Minneapolis: University of Minnesota Press.
22. Van der Heijden, P. T., Egger, J. I. M., Rossi, G., Grundel, G., & Derksen, J. J. L. (2012). The MMPI-2 Restructured Form and the standard MMPI-2 Clinical Scales in relation to DSM-IV. *European Journal of Psychological Assessment*. doi: [10.1027/1015-5759/a000140](https://doi.org/10.1027/1015-5759/a000140)
23. Watson, C., Quilty, L. C., & Bagby, R. M. (2011). Differentiating bipolar disorder from major depressive disorder using the MMPI-2-RF: A receiver operating characteristics (ROC) analysis. *Journal of Psychopathology and Behavioral Assessment*, 33, 368-374. doi: [10.1007/s10862-010-9212-7](https://doi.org/10.1007/s10862-010-9212-7)

End of Report

This and previous pages of this report contain trade secrets and are not to be released in response to requests under HIPAA (or any other data disclosure law that exempts trade secret information from release). Further, release in response to litigation discovery demands should be made only in accordance with your profession's ethical guidelines and under an appropriate protective order.

Assessing DSM-5 Section III Personality Traits and Disorders With the MMPI-2-RF

Assessment
20(6) 709–722
© The Author(s) 2013
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1073191113508808
asm.sagepub.com



Martin Sellbom¹, Jaime L. Anderson², and R. Michael Bagby³

Abstract

An alternative model for diagnosing personality disorders (PDs) appears in *DSM-5* Section III. This model includes a set of dimensional personality traits, which along with impairment in personality functioning can be configured to represent one of six PDs. Although specific assessment instruments for these personality traits have already been developed (e.g., the Personality Inventory for DSM-5 [PID-5]), clinicians will likely continue to use omnibus measures of psychopathology that are familiar to them to inform diagnostic decision making. One such measure, the Minnesota Multiphasic Personality Inventory–2–Restructured Form (MMPI-2-RF), will likely remain in the test armamentarium of many practitioners and be employed to assess the *DSM-5* dimensional traits. In the current investigation, we examined the associations between MMPI-2-RF scale scores and the PID-5 trait scores and *DSM-5* Section III PDs in a combined sample of university students ($n = 668$) from the United States and Canada. Our results indicated that the MMPI-2-RF scale scores mostly converge with PID-5 dimensional traits as well as the Section III PDs in a conceptually expected manner. As such, we conclude that the MMPI-2-RF is a potentially useful instrument in assessing personality psychopathology as conceptualized in *DSM-5* Section III.

Keywords

MMPI-2-RF, PID-5, *DSM-5*, personality traits

The *Diagnostic and Statistical Manual for Mental Disorders*, text revision (*DSM-IV-TR*; American Psychiatric Association [APA], 2000) provides diagnostic criteria for each of the 10 personality disorders (PDs) that appear in the main text of the manual. The method of diagnosis uses a polythetic, categorical approach in which a specified number of pathological personality symptoms or traits from a larger list of symptoms must be present to indicate a PD diagnosis. For example, to confer the diagnosis of Borderline PD, at least five symptoms/traits from a list of nine must be present. The problems and rather significant shortfalls of this PD diagnostic system are well documented and widely recognized (see, e.g., Clark, 2007; Widiger & Mullins-Sweatt, 2010); problems include (but are not limited to) excessively high rates of comorbidity with other PDs, inadequate coverage of personality pathology, extensive within-diagnosis heterogeneity, poor convergent and discriminant validity of measurement tools designed to assess PDs, marked temporal instability, and poor scientific research base for many disorders (e.g., Clark, 2007; Skodol et al., 2011; Widiger & Simonsen, 2005).

To address these concerns, many scholars have argued for the adoption of an alternative model and approach to conceptualize, assess, and diagnose PDs and a number of different models have been proposed. The proponents of most of these models argue compellingly for the use of dimensional personality traits to characterize PDs (see, e.g.,

Widiger, Simonsen, Sirovatka, & Regier, 2006). Along these lines, the *DSM-5* Personality and Personality Disorders (P&PD) workgroup, charged with the task of revising or developing a new system for conceptualizing and diagnosing PDs, proposed 25 “lower order” trait-facets, which, in turn, combine into five “higher order” dimensional trait domains (negative affectivity, detachment, antagonism, disinhibition, and psychoticism). These 25 facet traits form the basis for the conceptualizing and assessing personality pathology in Section III (e.g., APA, 2011, 2012; Krueger, Derringer, Markon, Watson, & Skodol, 2012; Skodol, 2012). The P&PD workgroup also proposed six PDs, each based on a specific and mostly unique trait-facet profile that results in or “co-occurs” with noted impairment in self and interpersonal functioning (Skodol et al., 2011). The six PDs are Antisocial, Avoidant, Borderline, Narcissistic, Obsessive-Compulsive, and Schizotypal. A seventh PD diagnosis is also possible—Personality Disorder–Trait Specified (PD-TS)—in which

¹The Australian National University, Canberra, ACT, Australia

²The University of Alabama, Tuscaloosa, AL, USA

³University of Toronto, Toronto, Ontario, Canada

Corresponding Author:

Martin Sellbom, Research School of Psychology, Building 39, The Australian National University, Canberra, ACT 0200, Australia.
Email: martin.sellbom@anu.edu.au

the presence of dysfunctional traits and associated impairment are noted, but none of the specific profiles that characterize the six PDs are present. The APA Board of Trustees decided not to incorporate the proposal into Section II of *DSM-5* (Essential Elements: Diagnostic Criteria and Codes) that lists disorders formally recognized in *DSM-5*, but rather placed it in Section III (Emerging Measures and Models) for further study until more empirical support for the proposed P&PD model and diagnostic system is generated.

The accurate assessment of *DSM-5* personality traits is critical if the proposed model and system of diagnosis is eventually to be implemented. The *DSM-5* P&PD workgroup and their associates developed the Personality Inventory for the *DSM-5* (PID-5; Krueger, Eaton, et al., 2012), an informant version of this measure (Markon, Quilty, Bagby, & Krueger, 2013), and a clinician rating form (APA, 2011) for direct assessment of this trait system. Notwithstanding these developments, we believe that many clinicians will likely continue to rely on well-known and widely used omnibus test instruments that are already part of their test armamentarium rather than include additional lengthy tests. This will especially be the case if it can be demonstrated that such omnibus instruments can indeed capture these *DSM-5* constructs. The Minnesota Multiphasic Personality Inventory-2 (MMPI-2; Butcher et al., 2011) has been the most widely and frequently used personality instrument in clinical practice for several decades (e.g., Archer, Buffington-Vollum, Stredny, & Handel, 2006; Camara, Nathan, & Puente, 2000), and the MMPI-2 Restructured Form (MMPI-2-RF; Ben Porath & Tellegen, 2008/2011) is the most recent version of the MMPI instruments. The MMPI-2-RF consists of 338 items derived from the MMPI-2 item pool, which aggregate into psychometrically sound scales that map well onto contemporary models of personality and psychopathology (e.g., Bagby et al., in press; Sellbom, Ben-Porath, & Bagby, 2008; Tellegen & Ben-Porath, 2008). The MMPI-2-RF scales are arranged in a hierarchical fashion, with three Higher-Order (Emotional-Internalizing, Thought Dysfunction, and Behavioral-Externalizing Dysfunction) scales at the top of the hierarchy; the nine Restructured Clinical (RC) scales at the mid-level; and the 23 Specific Problems (SP) scales occupying the lowest level. In parallel, a dimensional model of personality is also assessed via the Personality Psychopathology Five (PSY-5) scales (see Harkness, Finn, McNulty, & Shields, 2012).

Anderson et al. (2013) recently demonstrated meaningful convergence between the MMPI-2-RF PSY-5 and *DSM-5* Section III trait domains, as indexed by the MMPI-2-RF and PID-5, respectively. Unlike the PID-5 domains, however, the MMPI-2-RF PSY-5 domain scales do not have facets.¹ This limits the utility of these scales in providing information needed for a Section III PD diagnosis, as the personality trait profiles for these PDs are based on facet

traits. We do think, however, that an examination of the entire array of MMPI-2-RF scales located across all levels of the hierarchy can provide sufficient information to capture both the domain and facet traits of the PID-5 used to diagnose *DSM-5*, Section III PD types.

The Current Study

The goal of the current investigation was to examine the associations between the *DSM-5* Section III personality traits (as indexed via the PID-5) and MMPI-2-RF scale scores. We also investigated the degree to which a set of MMPI-2-RF scales conceptually and thematically similar to the PID-5 facet traits could account for significant amounts of variance in, and contribute uniquely to, the prediction of *DSM-5* Section III PDs (i.e., Antisocial, Avoidant, Borderline, Narcissistic, Obsessive-Compulsive, and Schizotypal) as defined by a composite of their respective facets. Although these latter analyses are of tentative utility since these PD versions will not formally be used in *DSM-5*, it is likely that similar variants will appear in future versions of the manual, and the analyses we conducted here provide a "first look" at whether the MMPI-2-RF could account for variance in such trait configurations.

We generated a set of hypotheses (see Table 1) for MMPI-2-RF scale and *DSM-5* Section III trait associations based on our conceptual understanding of these constructs, as well as empirical research that have examined both associations between MMPI-2-RF scale scores and personality traits from other models (Sellbom & Ben-Porath, 2005; Sellbom et al., 2008; Simms, Casillas, Clark, Watson, & Doebbeling, 2005; Tellegen & Ben-Porath, 2008) and the PID-5 and such traits (e.g., De Fruyt et al., 2013; Thomas et al., 2013). These hypotheses were stated at the domain level, as there rarely was a clear one-to-one correspondence between an MMPI-2-RF scale and a *DSM-5* trait-facet. Furthermore, given the personality trait-facets that will comprise each *DSM-5* Section III PD, we also hypothesized certain MMPI-2-RF scales to be particularly relevant in capturing variance in these disorders based on conceptually derived expectations and our knowledge of the constructs underlying the PID-5 and MMPI-2-RF scales; these are also listed in Table 1.

Method

Participants and Procedures

We used two separate university samples—one from the United States—the University of Alabama (UA)—and the other from Canada—the University of Toronto (UT). This combination created a larger and more diverse sample, conferring us greater statistical power and generalizability. The UA sample comprised 252 male and 211 female undergraduate

Table 1. MMPI-2-RF Scale Information and Hypotheses.

MMPI-2-RF Scale	Abbreviation	<i>M</i>	<i>SD</i>	α	Most relevant DSM-5 trait domain	Personality disorder hypotheses
Higher-order scales						
Emotional/Internalizing Dysfunction	EID	52.29	10.51	.93	Negative Affectivity; Detachment	Avoidant; Borderline; OCPD; Schizotypal
Thought Dysfunction	THD	53.82	10.57	.85	Psychoticism	Schizotypal
Behavioral/Externalizing Dysfunction	BXD	52.00	9.97	.81	Antagonism; Disinhibition	Antisocial; Narcissistic
Restructured clinical scales						
Demoralization	RCd	55.60	10.15	.93	Negative Affectivity; Detachment	Avoidant; Borderline; OCPD
Somatic Complaints	RC1	56.95	9.89	.83	Negative Affectivity	Avoidant; OCPD; Schizotypal
Low Positive Emotions	RC2	48.85	11.29	.84	Detachment	
Cynicism	RC3	55.22	9.78	.79	Antagonism; Negative Affectivity	
Antisocial Behavior	RC4	51.86	9.59	.81	Antagonism; Disinhibition	Antisocial; Narcissistic; low OCPD
Ideas of Persecution	RC6	56.27	10.50	.81	Psychoticism	Schizotypal
Dysfunctional Negative Emotions	RC7	54.13	10.67	.87	Negative Affectivity	Avoidant; Borderline; Schizotypal
Aberrant Experiences	RC8	54.75	11.62	.83	Psychoticism	Schizotypal
Hypomanic Activation	RC9	54.63	11.94	.83	Antagonism; Disinhibition	Antisocial; Narcissistic; low OCPD
Somatic/cognitive specific problem scales						
Malaise	MLS	52.42	9.90	.76	Negative Affectivity	Schizotypal OCPD; Schizotypal
Gastrointestinal Complaints	GIC	53.73	12.33	.72	None	
Head Pain Complaints	HPC	53.19	11.00	.71	None	
Neurological Complaints	NUC	57.06	10.94	.65	Psychoticism	
Cognitive Complaints	COG	57.26	12.03	.77	Negative Affectivity; Psychoticism	
Internalizing specific problem scales						
Suicide/Death Ideation	SUI	49.00	9.86	.75	Negative Affectivity; Detachment	Borderline Avoidant; Borderline Avoidant; Borderline; OCPD Borderline Borderline Antisocial; Borderline
Helplessness/Hopelessness	HLP	50.44	10.85	.67	Negative Affectivity	
Self-Doubt	SFD	53.47	12.04	.81	Negative Affectivity	
Inefficacy	NFC	56.32	11.40	.76	Negative Affectivity	
Stress/Worry	STW	55.77	10.60	.69	Negative Affectivity	
Anxiety	AXY	55.41	13.21	.69	Negative Affectivity	
Anger Proneness	ANP	52.84	10.70	.75	Antagonism; Negative Affectivity	
Behavior-Restricting Fears	BRF	54.69	11.96	.62	Negative Affectivity	
Multiple Specific Fears	MSF	48.49	8.32	.68	Negative Affectivity	
Externalizing specific problem scales						
Juvenile Conduct Problems	JCP	49.31	9.48	.68	Disinhibition	Antisocial
Substance Abuse	SUB	51.90	12.02	.76	Disinhibition	Antisocial
Aggression	AGG	51.78	11.55	.71	Antagonism	Antisocial
Activation	ACT	55.51	12.66	.71	Antagonism	Narcissistic
Interpersonal specific problem scales						
Family Problems	FML	52.32	10.81	.74	Negative Affectivity; Detachment	low Narcissistic Avoidant; OCPD; Schizotypal Avoidant; Schizotypal
Interpersonal Passivity	IPP	46.72	8.86	.71	Detachment	
Social Avoidance	SAV	46.59	10.81	.87	Detachment	
Shyness	SHY	49.26	9.51	.75	Negative Affectivity; Detachment	
Disaffiliativeness	DSF	52.76	10.61	.51	Detachment	

Note. OCPD = obsessive-compulsive personality disorder; MMPI-2-RF = Minnesota Multiphasic Personality Inventory-2-Restructured Form.

students who participated in the study for course credit in introductory psychology courses. Participants completed the study measures in groups of up to 10 under the supervision of a trained research assistant. Participants had a mean age of 19.54 years ($SD = 1.67$) and a mean education of 13.4 years. The sample was primarily Caucasian (Euro-American; 76%); the largest racial/ethnic minority group was African American (15%), and "other" racial/ethnic groups accounted for the remaining 9%. Sixty-six participants were removed from the analyses due to invalid MMPI-2-RF protocols (here defined as exhibiting any of the following: Cannot Say Scale > 18, Variable Response Inconsistency > 75T, True Response Inconsistency > 75T, Infrequent Responses > 100T, and Infrequency Psychopathology Responses > 90T). This left a total of 397 participants from this sample for analyses.² There were no significant differences between included and excluded participants on any demographic variables.

The UT sample consisted of 111 male and 246 female undergraduate students who participated in the study in which they had the option to receive either partial course credit or have their name entered into a drawing for a chance to win a cash prize of \$150.00 if they completed the study. Participants completed the study measures in groups of up to five under the supervision of a trained research assistant. Participants had a mean age of 20.43 years ($SD = 4.18$) and a mean education of 14.2 years. Consistent with the racial/ethnic composition of the suburban campus from which most of these data were collected, the sample was primarily of Asian (e.g., East Asian, South Asian, Southeast Asian) descent or origin (60%), with 18% reportedly being White (Euro-Canadian), and "other" racial/ethnic groups accounting for the remaining 22% (e.g., Indo-Caribbean, Latino, and Black Canadians of Caribbean descent). Eighty-six of the participants were excluded from analyses due to invalid MMPI-2-RF protocols using the same criteria as for the UA sample. This left a total of 271 participants for analyses. There were no significant differences between included and excluded participants on any demographic characteristic. In sum, all analyses were performed on a total sample of 628 participant test protocols.³

Measures

MMPI-2 Restructured Form (MMPI-2-RF). The MMPI-2-RF (Ben-Porath & Tellegen, 2008/2011) is a revised version of the MMPI-2 consisting of 338 true or false items. This measure includes 9 validity scales, 3 Higher-Order scales (H-O), 9 Restructured Clinical (RC) scales, 23 Specific Problems (SP) scales, 2 Interest scales, and 5 Personality Psychopathology Five (PSY-5) scales. Given that the PSY-5 scales were examined in a different study (Anderson et al., 2013), we focused exclusively on the H-O, RC, and SP scales in the current investigation. A full listing of MMPI-2-RF scales and psychometric statistics associated with these scales from the combined sample are displayed in Table 1.

Personality Inventory for DSM-5 (PID-5). The PID-5 (Krueger, Eaton, et al., 2012) is a 220-item self-report inventory developed to index the five proposed *DSM-5* personality domains and their respective facets. Item responses are based on a 4-point Likert-type scale. A complete list of facets for each domain is displayed in Table 2. Internal consistencies (Cronbach's alpha) in the current sample for the PID-5 domain scale scores were .88 (Disinhibition), .93 (Detachment), .94 (Antagonism and Negative Affectivity), and .95 (Psychoticism). At the facet level, Cronbach's alphas were also acceptable, ranging from .71 (Irresponsibility) to .95 (Depressivity).⁴ Furthermore, we indexed *DSM-5* Section III PD scores based on personality trait profiles via a sum of the facets proposed for each PD (see, e.g., Samuel, Hopwood, Krueger, Thomas, & Ruggero, 2013). These scores were internally consistent, with coefficient alphas ranging from .49 (OCPD) to .82 (Antisocial PD) and mean interitem correlations from .19 (OCPD) to .49 (Narcissistic PD).

Results

We examined associations between MMPI-2-RF scales and PID-5 domain and facet scales via zero-order correlations. Because of the sheer number of correlations reported, we used a highly conservative alpha level of .0001; as such, given the current sample size, an absolute correlation of .16 was considered statistically significant. Because of shared method variance possibly artificially inflating effect size magnitudes, we focused our interpretation on correlations of a moderate (i.e., $r = .30-.49$) or large ($r \geq .50$; see Cohen, 1992) magnitude.

DSM-5 Personality Traits

The correlations between all MMPI-2-RF H-O, RC, and SP scales and PID-5 domain and facet scores are displayed in Tables 2 to 4. We organize the results per the prediction of each *DSM-5* personality trait domain and its associated facets.⁵

Negative Affectivity. The pattern of associations between PID-5 Negative Affectivity and the MMPI-2-RF scales was quite clear; the internalizing scales capture this domain quite well. In particular, the EID scale, along with RCd and RC7 were associated with the largest effect sizes. Most of the internalizing SP scales also exhibited large correlations with this domain. In terms of the PID-5 facet scores, EID, RCd, and RC7 were moderately to strongly correlated with all of them. The SFD, NFC, STW, and AXY scales exhibited large correlations with PID-5 Anxiousness, whereas COG and NFC were, as expected, the best and strongest predictors of PID-5 Perseveration. There were no distinct predictors of PID-5 Emotional Dysregulation or Separation Insecurity; RC7 was the only MMPI-2-RF scale that

Table 2. Correlations Between MMPI-2-RF Higher Order and Restructured Clinical Scales and PID-5 Personality Trait Scales.

	EID	THD	BXD	RCd	RC1	RC2	RC3	RC4	RC6	RC7	RC8	RC9
Negative Affectivity	.68	.29	.18	.67	.39	.34	.36	.22	.34	.71	.30	.28
Emotional Dysregulation	.47	.25	.13	.46	.38	.19	.24	.15	.25	.53	.28	.24
Perseveration	.51	.36	.20	.56	.30	.26	.35	.23	.35	.62	.40	.31
Anxiousness	.65	.23	.04	.61	.39	.37	.31	.11	.30	.64	.19	.13
Separation Insecurity	.39	.19	.17	.40	.23	.13	.27	.21	.19	.47	.21	.25
Hostility	.42	.29	.37	.38	.24	.16	.37	.31	.30	.53	.27	.43
Submissiveness	.38	.00	-.04	.38	.04	.26	.06	.03	.04	.36	.04	.03
Suspiciousness	.43	.40	.28	.44	.33	.20	.45	.28	.45	.48	.34	.31
Detachment	.63	.21	.11	.61	.23	.55	.29	.17	.25	.40	.19	.02
Restricted Affectivity	.15	.13	.17	.18	-.02	.17	.18	.16	.10	.10	.12	.10
Depressivity	.76	.13	.10	.76	.27	.61	.22	.18	.21	.45	.12	.03
Anhedonia	.68	.08	.04	.64	.20	.62	.21	.12	.13	.35	.07	-.05
Withdrawal	.51	.09	-.05	.42	.16	.51	.17	.02	.11	.29	.07	-.13
Intimacy Avoidance	.22	.09	-.07	.23	.08	.25	.05	.00	.09	.11	.10	-.11
Antagonism	.15	.35	.53	.24	.15	-.11	.40	.42	.30	.37	.40	.60
Manipulativeness	-.03	.26	.43	.09	.08	-.21	.32	.32	.19	.18	.31	.51
Deceitfulness	.22	.27	.47	.31	.13	.03	.36	.42	.22	.34	.33	.47
Callousness	.17	.35	.49	.23	.11	.06	.36	.40	.30	.25	.35	.44
Grandiosity	-.05	.25	.28	.00	.06	-.23	.21	.20	.20	.18	.27	.37
Attention Seeking	.01	.22	.41	.13	.07	-.22	.23	.31	.20	.25	.30	.49
Disinhibition	.05	.30	.51	.19	.04	-.10	.29	.48	.21	.25	.42	.50
Impulsivity	.20	.35	.49	.31	.16	.00	.29	.44	.28	.34	.43	.46
Irresponsibility	.37	.20	.36	.45	.18	.27	.25	.41	.20	.35	.27	.27
Risk Taking	-.16	.21	.54	-.02	-.03	-.23	.16	.43	.14	-.03	.27	.45
Distractibility	.46	.30	.27	.56	.26	.26	.29	.28	.28	.49	.36	.28
Rigid Perfectionism	-.25	.06	.17	-.19	-.15	-.26	.08	.16	-.02	-.05	.14	.22
Psychoticism	.34	.54	.40	.46	.32	.15	.34	.37	.44	.48	.63	.43
Eccentricity	.33	.35	.34	.42	.22	.19	.25	.33	.33	.41	.45	.34
Unusual Beliefs & Thoughts	.18	.56	.36	.30	.27	.01	.30	.28	.41	.33	.61	.40
Cognitive & Perceptual Dysregulation	.37	.54	.34	.49	.38	.16	.34	.35	.44	.52	.60	.39
Antisocial PD	.23	.39	.64	.35	.18	.00	.43	.55	.33	.40	.46	.63
Avoidant PD	.71	.18	-.00	.65	.30	.58	.26	.09	.23	.50	.16	-.04
Borderline PD	.61	.37	.39	.64	.38	.27	.41	.39	.37	.67	.39	.44
Narcissistic PD	-.02	.27	.41	.09	.07	-.26	.25	.30	.23	.25	.33	.50
Obsessive Compulsive PD	.41	.44	-.03	.39	.21	.37	.11	-.01	.17	.27	.07	-.08
Schizotypal PD	.46	.46	.34	.53	.30	.31	.39	.33	.42	.49	.50	.32

Note. Underlined correlations are of moderate effect size; bolded correlations are of large effect size. EID = Emotional/Internalizing Dysfunction; THD = Thought Dysfunction; BXD = Behavioral/External Dysfunction; RCd = Demoralization; RC1 = Somatic Complaints; RC2 = Low Positive Emotions; RC3 = Cynicism; RC4 = Antisocial Behavior; RC6 = Ideas of Persecution; RC7 = Dysfunctional Negative Emotions; RC8 = Aberrant Experiences; RC9 = Hypomanic Activation; PD = personality disorder; MMPI-2-RF = Minnesota Multiphasic Personality Inventory-2-Restructured Form; PID-5 = Personality Inventory for DSM-5.

exhibited a large correlation with these facets. Furthermore, along with RC7, RC3 and RC6 were the best predictors of PID-5 Suspiciousness, whereas RCd, RC7 and NFC had the largest correlations with PID-5 Submissiveness. Finally, ANP was the best predictor of PID-5 Hostility.

Detachment. The MMPI-2-RF scales EID, RCd, RC2, and several of the interpersonal SP scales were associated with

the largest effect sizes in the prediction of the PID-5 Detachment domain, outcomes which are in line with our a priori expectations. The EID, RCd, RC2, MLS, HLP, and SFD scales were the best predictors of PID-5 Depressivity, with RC7 and several other internalizing SP scales also correlating with this facet in the moderate to large range. These findings were not unexpected given that PID-5 Depressivity cross-loads on both the Detachment and Negative Affectivity

Table 3. Correlations Between MMPI-2-RF Somatic/Cognitive and Internalizing Specific Problem Scales and PID-5 Personality Trait Scales.

	MLS	GIC	HPC	NUC	COG	SUI	HLP	SFD	NFC	STW	AXY	ANP	BRF	MSF
Negative Affectivity	<u>.45</u>	.25	<u>.32</u>	.28	<u>.46</u>	.24	<u>.46</u>	.57	.61	.60	<u>.49</u>	<u>.49</u>	<u>.31</u>	.20
Emotional Dysregulation	<u>.33</u>	.22	<u>.31</u>	.26	<u>.37</u>	.18	<u>.32</u>	<u>.38</u>	<u>.41</u>	<u>.42</u>	<u>.39</u>	<u>.43</u>	<u>.29</u>	.23
Perseveration	<u>.34</u>	.16	<u>.20</u>	<u>.31</u>	.54	.15	<u>.38</u>	<u>.45</u>	.55	<u>.46</u>	<u>.37</u>	<u>.37</u>	<u>.24</u>	.09
Anxiousness	<u>.44</u>	.26	<u>.32</u>	<u>.22</u>	<u>.36</u>	.20	<u>.42</u>	.53	.56	.64	<u>.49</u>	<u>.43</u>	<u>.35</u>	.26
Separation Insecurity	<u>.24</u>	.17	<u>.21</u>	.15	<u>.30</u>	.13	<u>.25</u>	<u>.36</u>	<u>.41</u>	<u>.40</u>	<u>.36</u>	<u>.26</u>	<u>.22</u>	.12
Hostility	.25	.10	.22	.20	<u>.29</u>	.14	.27	<u>.27</u>	<u>.31</u>	<u>.36</u>	<u>.29</u>	.71	.17	.14
Submissiveness	.22	.13	.05	.05	.20	.13	.28	<u>.37</u>	<u>.40</u>	<u>.30</u>	.21	.10	.09	.02
Suspiciousness	.28	.09	.24	<u>.31</u>	<u>.32</u>	.16	.25	<u>.34</u>	<u>.35</u>	<u>.39</u>	<u>.33</u>	<u>.34</u>	.14	.07
Detachment	<u>.44</u>	.11	.16	<u>.19</u>	<u>.29</u>	.26	<u>.45</u>	<u>.49</u>	<u>.39</u>	<u>.37</u>	<u>.19</u>	.20	.05	-.02
Restricted Affectivity	<u>.08</u>	-.05	-.03	.06	.10	.08	.14	.11	.07	.06	-.01	.08	-.10	-.14
Depressivity	.54	.19	.21	.17	<u>.37</u>	<u>.38</u>	.58	.68	<u>.46</u>	<u>.44</u>	<u>.30</u>	.22	.12	.01
Anhedonia	.53	.15	.15	.12	.26	<u>.28</u>	<u>.47</u>	<u>.49</u>	<u>.37</u>	<u>.36</u>	<u>.17</u>	.18	.01	-.02
Withdrawal	<u>.33</u>	.08	.14	.09	.14	.19	<u>.32</u>	<u>.35</u>	<u>.30</u>	.25	.10	.10	.05	.01
Intimacy Avoidance	<u>.17</u>	.03	.02	.09	.11	.05	.17	.17	.15	.12	-.01	-.02	.05	.02
Antagonism	.05	.05	.09	.22	<u>.30</u>	.11	.18	.13	.20	.20	.21	<u>.45</u>	.04	.00
Manipulativeness	-.01	.05	.04	.17	.18	.08	.08	.02	.09	.09	.09	.26	-.02	-.03
Deceitfulness	.13	.06	.07	.22	<u>.34</u>	.13	.26	.23	.23	.19	.20	<u>.30</u>	.02	-.04
Callousness	.08	.01	.06	.21	.26	.12	.22	.13	.07	.06	.13	<u>.31</u>	-.07	-.13
Grandiosity	-.10	.02	.03	.11	.07	.00	.02	-.08	.04	.05	.06	.21	-.03	.03
Attention Seeking	-.05	.00	.02	.15	.26	.04	.06	.05	.16	.15	.18	.28	.06	.00
Disinhibition	-.01	.02	.00	.21	<u>.46</u>	.08	.03	.11	.11	.09	.21	.19	-.06	-.19
Impulsivity	.10	.08	.07	.27	<u>.48</u>	.12	.21	.22	.22	.18	.26	<u>.34</u>	.11	-.02
Irresponsibility	.26	.06	.09	.25	<u>.44</u>	.13	<u>.34</u>	<u>.34</u>	<u>.33</u>	.26	.22	.23	.11	-.04
Risk Taking	-.14	-.06	-.03	.09	.16	.00	-.11	-.10	-.15	-.08	.02	.13	-.18	-.27
Distractibility	<u>.33</u>	.18	.17	.27	.68	.16	<u>.38</u>	<u>.43</u>	.51	<u>.42</u>	<u>.31</u>	<u>.32</u>	.21	.08
Rigid Perfectionism	-.21	-.08	-.11	-.01	.05	-.03	-.26	-.16	-.17	-.16	.01	-.08	-.20	-.21
Psychoticism	.24	.11	.22	<u>.34</u>	.55	.13	<u>.33</u>	<u>.34</u>	<u>.36</u>	<u>.31</u>	<u>.32</u>	.28	.09	-.01
Eccentricity	.22	.07	.13	.26	<u>.49</u>	.13	<u>.30</u>	<u>.33</u>	<u>.30</u>	<u>.27</u>	<u>.25</u>	.22	.02	-.08
Unusual Beliefs & Thoughts	.13	.08	.19	.29	<u>.39</u>	.08	.24	.19	.25	.21	.22	.22	.05	.01
Cognitive & Perceptual Dysregulation	.28	.15	.29	<u>.36</u>	.56	.12	<u>.32</u>	<u>.35</u>	<u>.40</u>	<u>.32</u>	<u>.39</u>	<u>.30</u>	.21	.07
Antisocial PD	.13	.07	.11	.28	<u>.43</u>	.15	.25	.22	.22	.22	.25	<u>.47</u>	.04	-.07
Avoidant PD	.50	.19	.23	.18	<u>.31</u>	.25	<u>.47</u>	.53	<u>.49</u>	.50	.29	.26	.18	.12
Borderline PD	<u>.40</u>	.22	<u>.30</u>	<u>.31</u>	.52	.25	<u>.43</u>	.52	.51	.54	<u>.49</u>	.57	.26	.13
Narcissistic PD	-.09	.01	.02	.15	.20	.03	.05	-.00	.12	.12	.15	.29	.03	.02
Obsessive Compulsive PD	<u>.31</u>	.09	.13	.13	.16	.09	<u>.38</u>	<u>.32</u>	<u>.33</u>	<u>.30</u>	.09	.18	.20	.14
Schizotypal PD	<u>.30</u>	.09	.21	<u>.31</u>	<u>.46</u>	.18	<u>.37</u>	<u>.39</u>	<u>.39</u>	<u>.35</u>	.29	.29	.07	-.03

Note. Underlined correlations are of moderate effect size; bolded correlations are of large effect size. MLS = Malaise; GIC = Gastrointestinal Complaints; HPC = Head Pain Complaints; NUC = Neurological Complaints; COG = Cognitive Complaints; SUI = Suicide/Death Ideation; HLP = Hopelessness/Helplessness; SFD = Self-Doubt; NFC = Inefficacy; STW = Stress/Worry; AXY = Anxiety; ANP = Anger Proneness; BRF = Behavior-Restricting Fears; MSF = Multiple Specific Fears; PD = personality disorder; MMPI-2-RF = Minnesota Multiphasic Personality Inventory-2-Restructured Form; PID-5 = Personality Inventory for DSM-5.

domains (Krueger, Eaton, et al., 2012). EID, RCd, RC2, MLS, and SAV were the most strongly correlated scales with PID-5 Anhedonia, whereas RC2, SAV, and SHY were clearly the best indicators of PID-5 Withdrawal. The MMPI-2-RF scales did not predict PID-5 Restricted Affectivity or Intimacy Avoidance to the same degree as other facets; it is

noteworthy that DSF was associated with the largest correlation for both of these facets (.23 and .33, respectively), which is conceptually intuitive.

Antagonism. As expected, the best predictors of PID-5 Antagonism were the BXD, RC9, and AGG scales at each

Table 4. Correlations Between MMPI-2-RF Externalizing and Interpersonal Specific Problem Scales and PID-5 Personality Trait Scales.

	JCP	SUB	AGG	ACT	FML	IPP	SAV	SHY	DSF
Negative Affectivity	.15	.13	.29	.23	<u>.42</u>	.03	.08	<u>.33</u>	.12
Emotional Dysregulation	.12	.06	.21	<u>.30</u>	<u>.34</u>	-.01	.03	<u>.18</u>	.02
Perseveration	.16	.14	.29	<u>.28</u>	<u>.39</u>	-.01	.08	<u>.30</u>	.14
Anxiousness	.07	.03	.18	.12	<u>.40</u>	.12	.16	<u>.35</u>	.16
Separation Insecurity	.08	.20	.17	.21	<u>.21</u>	-.03	-.03	<u>.20</u>	.05
Hostility	.24	.19	.53	.18	<u>.36</u>	-.21	.00	.19	.09
Submissiveness	-.02	.07	-.00	.04	<u>.13</u>	.18	.08	.25	.11
Suspiciousness	.21	.17	<u>.34</u>	.20	.29	-.02	.10	.23	.14
Detachment	.15	.04	<u>.21</u>	-.08	<u>.35</u>	.24	<u>.48</u>	<u>.44</u>	<u>.37</u>
Restricted Affectivity	.11	.14	.18	-.04	<u>.11</u>	.07	<u>.21</u>	<u>.16</u>	<u>.23</u>
Depressivity	.16	.05	.19	-.08	<u>.39</u>	.24	<u>.31</u>	<u>.34</u>	.24
Anhedonia	.10	.01	.16	-.18	<u>.31</u>	.24	<u>.48</u>	<u>.43</u>	.29
Withdrawal	.06	-.11	.09	-.16	<u>.25</u>	.29	.63	.51	<u>.32</u>
Intimacy Avoidance	.03	-.07	-.02	-.04	.16	.20	.26	.19	<u>.33</u>
Antagonism	.29	<u>.36</u>	.52	<u>.31</u>	.27	-.36	-.22	-.05	.03
Manipulativeness	.21	<u>.30</u>	<u>.36</u>	<u>.28</u>	.14	-.36	-.27	-.15	-.02
Deceitfulness	<u>.30</u>	<u>.34</u>	<u>.41</u>	.23	.24	-.19	-.10	.04	.09
Callousness	.29	.26	.55	.12	.28	-.18	.00	.06	.17
Grandiosity	.11	.23	.29	.22	.08	-.30	-.14	-.04	-.01
Attention Seeking	.21	.29	<u>.31</u>	<u>.32</u>	.17	-.33	-.38	-.22	-.10
Disinhibition	.26	<u>.45</u>	<u>.34</u>	<u>.33</u>	.11	-.23	-.27	-.08	-.01
Impulsivity	<u>.35</u>	<u>.31</u>	<u>.40</u>	<u>.30</u>	.26	-.17	-.20	.00	.01
Irresponsibility	<u>.30</u>	<u>.31</u>	<u>.31</u>	.12	<u>.30</u>	.04	.01	.16	.08
Risk Taking	<u>.28</u>	<u>.37</u>	<u>.32</u>	.22	<u>.06</u>	-.28	-.34	-.26	-.11
Distractibility	.22	.19	.28	.22	<u>.31</u>	.00	-.00	.22	.11
Rigid Perfectionism	-.02	.26	.04	.18	-.18	-.20	-.22	-.16	-.05
Psychoticism	<u>.31</u>	.20	<u>.40</u>	<u>.33</u>	<u>.39</u>	-.10	.03	.13	.12
Eccentricity	.26	.18	<u>.31</u>	.25	<u>.34</u>	-.04	.07	.13	.14
Unusual Beliefs and Thoughts	.27	.13	<u>.37</u>	.28	<u>.29</u>	-.16	-.02	.04	.06
Cognitive and Perceptual Dysregulation	.27	.21	<u>.37</u>	<u>.35</u>	<u>.39</u>	-.06	.02	.16	.11
Antisocial PD	.40	<u>.42</u>	.58	<u>.31</u>	<u>.32</u>	-.29	-.20	-.01	.05
Avoidant PD	.09	-.04	.15	-.07	<u>.39</u>	.27	<u>.49</u>	.50	<u>.36</u>
Borderline PD	.28	.26	<u>.43</u>	.29	<u>.45</u>	-.08	-.02	.23	<u>.10</u>
Narcissistic PD	.19	.31	<u>.34</u>	<u>.31</u>	<u>.15</u>	-.37	-.32	-.16	-.07
Obsessive Compulsive PD	.09	-.13	.10	-.07	<u>.31</u>	.21	<u>.30</u>	.29	.22
Schizotypal PD	.28	.17	<u>.38</u>	.19	<u>.39</u>	.03	.26	<u>.30</u>	.25

Note. Underlined correlations are of moderate effect size; bolded correlations are of large effect size. JCP = Juvenile Conduct Problems; SUB = Substance Abuse; AGG = Aggression; ACT = Activation; FML = Family Problems; IPP = Interpersonal Passivity; SAV = Social Avoidance; SHY = Shyness; DSF = Disaffiliativeness; PD = personality disorder; MMPI-2-RF = Minnesota Multiphasic Personality Inventory-2-Restructured Form; PID-5 = Personality Inventory for DSM-5.

consecutive level of the MMPI-2-RF hierarchy; these scales were the only ones associated with a large effect size. Several other MMPI-2-RF scales were associated with moderate effect sizes, most notably RC4, ANP, SUB, and low IPP. With respect to Antagonism facets, RC9 was associated with the largest effect size in the prediction of PID-5 Manipulativeness, Deceitfulness, Grandiosity, and Attention Seeking.

AGG was the only MMPI-2-RF scale that reached a large effect size in the prediction of PID-5 Callousness, and ANP was clearly associated with the most substantial prediction of PID-5 Hostility. The RC4 and BXD scales were also consistent predictors of all PID-5 Antagonism facets, except Grandiosity. Low scores on IPP were specifically associated with high scores on PID-5 Manipulativeness, Grandiosity, and

Attention Seeking, reflecting its low social potency interpersonal properties (e.g., Ayearst, Sellbom, Trobst, & Bagby, 2013), whereas the SAV scale was specifically negatively correlated with PID-5 Attention Seeking.

Disinhibition. The MMPI-2-RF scales reflective of externalizing symptoms and traits were the best predictors of the PID-5 Disinhibition domain. These scales included BXD, RC4, RC9, SUB, AGG, and ACT, but also the COG scale. In terms of the prediction of specific PID-5 facets, the aforementioned scales were also, as expected, associated with the largest correlations with PID-5 Impulsivity and Risk Taking. The pattern was less clear, however, with respect to PID-5 Irresponsibility, as both externalizing and internalizing scales were equally strong predictors of this facet. Indeed, the six internalizing and externalizing SP scales that were moderately correlated with PID-5 Irresponsibility were also significant and approximately equal predictors in a regression equation (β s = .10-.21, p s < .01), with no pattern favoring one set over the other. The prediction of PID-5 Distractibility, however, was associated with a very different pattern. The COG scale was the best predictor of this Disinhibition facet, and several internalizing scales (EID, RCd, RC7, NFC) were also strongly correlated with this facet, and to a substantially greater magnitude than the externalizing scales. Finally, none of the MMPI-2-RF scales reached a moderate correlation with PID-5 (lack of) Rigid Perfectionism.

Psychoticism. The MMPI-2-RF scales associated with large effect size predictions of the PID-5 Psychoticism domain were THD, RC8, and COG at each respective level of the MMPI-2-RF hierarchy. These MMPI-2-RF scales were also the best predictors of two of the three PID-5 facets of Cognitive and Perceptual Dysregulation and Unusual Thoughts and Beliefs. The pattern of correlations between MMPI-2-RF scales and PID-5 Eccentricity was more diffuse than those for the other two facets, with all three H-O scales and several RC scales exhibiting very similar correlations with this PID-5 facet. To clarify the picture of MMPI-2-RF associations with PID-5 Psychoticism in general, and PID-5 Eccentricity in particular, we regressed each of these PID-5 scales onto the full set of RC scales. These results indicated that RC8 contributed most substantially to the prediction of both PID-5 Psychoticism ($\beta = .47, p < .001$) and Eccentricity ($\beta = .30, p < .001$), with the second largest contributors being RCd ($\beta = .15, p = .003$) and RC9 ($\beta = .14, p = .007$), respectively. Thus, MMPI-2-RF indicators of thought dysfunction are clearly the best unique predictors of this PID-5 domain and its facets.

DSM-5 Personality Disorders

Our second major goal in this study was to examine the best MMPI-2-RF predictors of the personality trait profiles characterizing each of the six DSM-5 Section III PDs. We

examined zero-order correlations between all MMPI-2-RF scale scores and DSM-5 Section III PDs, interpreting medium or large effect sizes as meaningful (Cohen, 1992). Next, we estimated a series of regression models in which a set of conceptually relevant MMPI-2-RF scales were entered as predictors for each PD. MMPI-2-RF scales that exhibited at least a large correlation with at least one of the facets proposed for a particular PD were also included even if such a scale was not initially indicated from a conceptual perspective. Three series of regression equations were calculated for each PD; two separate ones for the RC and SP scales and one overall equation to determine which scales contributed uniquely when all MMPI-2-RF scales were considered simultaneously.⁶ These regression results are shown in Table 5.

Antisocial. The best predictors of Antisocial PD were BXD, RC4, RC9, and AGG, as expected. These scales were also among the best individual predictors of the specific traits associated with this disorder. For the regression models, we also considered RC3, JCP, SUB, ANP, and IPP because these scales are conceptually relevant to (or substantive predictors of) many of the specific traits that define Antisocial PD. As evident from Table 5, among the RC scales, all three contributed uniquely to the prediction, but RC4 and RC9 accounted almost exclusively for the predicted variance. Among the SP scales, all four (JCP, SUB, AGG, and ANP) contributed uniquely to the prediction of Antisocial PD, with AGG accounting for the most unique variance. When all scales were combined in one regression equation, which accounted for 51% of variance in Antisocial PD scores, RC3 and JCP were no longer significant, and RC4 and RC9 continued to be associated with the most substantial prediction of this PD.

Avoidant. Avoidant PD was most strongly associated with EID, RCd, RC2, RC7, SFD, NFC, STW, SAV, and SHY, which was consistent with conceptual expectations. These scales were also considered in the regression models along with MLS (due to a large zero-order effect size) and DSF (for conceptual relevance as a predictor of Intimacy Avoidance). The RC scale regression equation indicated that RCd, RC2, and RC7 all contributed meaningfully to the prediction of Avoidant PD scores. Among the SP scales, all scales except SHY contributed meaningfully to this prediction. In the full regression model, which accounted for 60% of the variance in Avoidant PD scores, RCd, RC2, RC7, STW, SAV, and DSF were significant unique contributors.

Borderline. DSM-5 Section III Borderline PD was moderately to strongly correlated with a wide range of MMPI-2-RF scales, but most notably EID, RCd, RC7, COG, SFD, NFC, STW, AXY, and ANP. This combination of scales reflects a wide and diverse range of negative emotional

Table 5. Multiple Regression Analyses Predicting DSM-5 Section III Personality Disorders.

	R^2	β	p	Full model		
				R^2	β	p
Antisocial PD				.51		
RC Scales	.48					
RC3		.09	.006		.06	.084
RC4		.32	<.001		.20	.002
RC9		.43	<.001		.29	<.001
SP Scales	.46					
ANP		.24	<.001		.17	<.001
JCP		.14	<.001		.03	.539
SUB		.25	<.001		.09	.040
AGG		.33	<.001		.13	.003
Avoidant PD				.60		
RC Scales	.52					
RCd		.29	<.001		.28	<.001
RC2		.37	<.001		.10	.030
RC7		.23	<.001		.13	.001
SP Scales	.57					
MLS		.17	<.001		.06	.081
SFD		.16	<.001		-.03	.473
NFC		.13	<.001		.03	.483
STW		.22	<.001		.14	<.001
SAV		.29	<.001		.28	<.001
SHY		.07	.033		.03	.309
DSF		.15	<.001		.13	<.001
Borderline PD				.60		
RC Scales	.56					
RCd		.35	<.001		.21	<.001
RC4		.14	<.001		.13	<.001
RC7		.33	<.001		.11	.033
RC9		.15	<.001		.10	.002
SP Scales	.56					
COG		.15	<.001		.02	.660
HLP		.07	.021		.05	.129
SFD		.19	<.001		.08	.088
NFC		.04	.225		.01	.833
STW		.17	<.001		.14	<.001
AXY		.11	.001		.05	.142
ANP		.33	<.001		.22	<.001
Narcissistic PD				.28		
RC Scales	.26					
RC4		.09	.022		.09	.020
RC9		.46	<.001		.39	<.001
SP Scales	.20					
ACT		.18	<.001		-.05	.267
IPP		-.24	<.001		-.10	.014
SAV		-.16	<.001		-.11	.003
OC PD				.23		
RC Scales	.20					
RCd		.34	<.001		.25	<.001
RC2		.16	.002		.07	.209
RC4		-.10	.040		-.04	.373
RC9		-.06	.175		-.05	.265

(continued)

Table 5. (continued)

	R^2	β	p	Full model		
				R^2	β	p
SP Scales	.19					
COG		.02	.616		-.04	.414
NFC		.27	<.001		.15	.001
SAV		.25	<.001		.13	.005
DSF		.09	.011		.07	.051
Schizotypal PD				.47		
RC Scales	.44					
RCd		.17	.001		.20	<.001
RC2		.24	<.001		.07	.139
RC6		.11	.002		.12	.002
RC7		.09	.052		.06	.170
RC8		.37	<.001		.35	<.001
SP Scales	.31					
NUC		.14	<.001		.04	.253
COG		.38	<.001		.02	.701
SAV		.21	<.001		.20	<.001
SHY		.07	.061		.01	.825
DSF		.12	.001		.06	.035

Note. PD = Personality Disorder; RC = Restructured Clinical; SP = Specific Problems; RCd = Demoralization; RC2 = Low Positive Emotions; RC3 = Cynicism; RC4 = Antisocial Behavior; RC6 = Ideas of Persecution; RC7 = Dysfunctional Negative Emotions; RC8 = Aberrant Experiences; RC9 = Hypomanic Activation; MLS = Malaise; NUC = Neurological Complaints; COG = Cognitive Complaints; HLP = Hopelessness/Helplessness; SFD = Self-Doubt; NFC = Inefficacy; STW = Stress/Worry; AXY = Anxiety; ANP = Anger Proneness; JCP = Juvenile Conduct Problems; SUB = Substance Abuse; AGG = Aggression; ACT = Activation; IPP = Interpersonal Passivity; SAV = Social Avoidance; SHY = Shyness; DSF = Disaffiliativeness; DSM = *Diagnostic and Statistical Manual for Mental Disorders*.

experiences, including emotional dysregulation, a "cardinal" feature this PD. The regression models included RCd and RC7 as both conceptually and empirically indicated, but also RC4 and RC9 given the prominence of disinhibition traits associated with the disorder. Indeed, the regression equation for the RC scales not only indicated substantial contributions by RCd and RC7 but also secondary unique contributions of RC4 and RC9. In the SP scale regression model, all scales except NFC contributed uniquely and meaningfully to the prediction of Borderline PD. For the overall model, which explained 60% of variance in Borderline PD scores, only RCd, RC4, RC9, STW, and ANP accounted for unique variance, with the latter two scales likely reflecting stress reactivity and intense anger dysregulation more specifically than RC7.

Narcissistic. Only one MMPI-2-RF scale (RC9) exhibited a large correlation with Narcissistic PD scores. BXD, RC8, SUB, AGG, low IPP, and low SAV were associated with this PD at a moderate level, which was generally expected. In the regression model, RC4 and RC9 were entered as predictors, and while both contributed significantly to the prediction of Narcissistic PD, only RC9 was associated with a meaningful magnitude. In terms of the SP scales, ACT, IPP, and SAV were entered on conceptual grounds given their

link to attention seeking, and to a much lesser degree, grandiosity; all of these scales contributed significantly to the prediction of this PD. The overall regression equation accounted for 28% of variance in Narcissistic PD type scores, with RC9, low IPP, and low SAV being unique contributors.

Obsessive-Compulsive. None of the MMPI-2-RF scales exhibited a large correlation with OCPD scores. EID, RCd, RC2, MLS, HLP, SFD, NFC, STW, and SAV all evidenced moderate correlations with this PD, most likely owing to their prediction of PID-5 Perseveration. Except for DSF, none of the MMPI-2-RF scales was moderately associated with the other facets that comprise OCPD (i.e., PID-5 Rigid Perfectionism, Intimacy Avoidance, and Restricted Affectivity). In the RC scale regression equation, we entered RCd and RC2 as negative affectivity and detachment markers, respectively, and RC4 and RC9 as disinhibition markers. RCd, RC2, and RC4 significantly contributed to the prediction of OCPD scores in their expected directions, but RCd was associated with the largest amount of predicted variance. In the SP scale equation, four markers of perseveration, intimacy avoidance, and restricted affectivity (COG, NFC, SAV, and DSF) were entered, and NFC and SAV were associated with a meaningful unique prediction of OCPD.

In the overall regression equation, which explained 23% of variance, only RCd, SAV, and NFC remained significant.

Schizotypal. The MMPI-2-RF RCd and RC8 were the only scales associated with large correlations with Schizotypal PD, which reflects both the negative affectivity and thought dysfunction processes that define this disorder. Similarly, correlations with EID and THD were in the moderate to large range. Several other scales, such as RC6, RC7, and COG were also notable predictors, primarily as they captured individual psychoticism and negative affectivity facets that define Schizotypal PD. In the RC scale regression model, we entered RCd, RC2, RC6, RC7, and RC8 to capture all elements of psychoticism, detachment, and negative affectivity. The results indicated that RCd, RC2, RC6, and RC8 were distinct predictors, with RC8 accounting for the most unique variance in Schizotypal PD. At the SP scale level, NUC, COG, SAV, SHY, and DSF were entered primarily as markers of psychoticism and detachment, and results showed that all scales except SHY contributed significantly to the prediction, with COG accounting for the largest proportion of unique variance. The overall regression explained 47% of variance in Schizotypal PD scores, with RCd, RC6, RC8, SAV, and DSF contributing uniquely to this prediction.

Discussion

In the current investigation, we examined the associations between the MMPI-2-RF and *DSM-5* Section III personality traits and disorders. We found that MMPI-2-RF scale scores were able to account for a substantial proportion of variance in *DSM-5* traits, particularly at the domain level. Moreover, the MMPI-2-RF scales generally predicted *DSM-5* Section III PDs in a conceptually expected manner. A substantial proportion of variance (47% to 60%) was explained in Antisocial, Avoidant, Borderline, and Schizotypal PDs, with smaller (albeit still large) amounts in Narcissistic and Obsessive-Compulsive PDs.

The MMPI-2-RF scales, especially those located in a higher level of the hierarchy of scales (i.e., H-O and RC), accounted for a substantial amount of variance in the PID-5 domain scales in a manner consistent with the prediction of traits from other dimensional personality models (e.g., Sellbom & Ben-Porath, 2005; Sellbom et al., 2008; Simms et al., 2005; Tellegen & Ben-Porath, 2008). The three H-O scales (EID, THD, and BXD) are anchored at a broader level of the personality hierarchy, reflecting propensities toward internalizing (combination of negative affect and detachment), thought dysfunction (psychoticism), and externalizing (combination of antagonism and disinhibition), which have been supported in various hierarchical analyses of personality trait models (e.g., Bagby et al., in press; Watson, 2005; Wright et al., 2012). The RC scales

further map onto the PID-5 domains in mostly expected ways, and largely consistent with research on how other omnibus measures, such as the PAI (Hopwood et al., 2013), NEO PI-R (De Fruyt et al., 2013; Markon et al., 2013; Thomas et al., 2013) and SNAP-2 (Watson, Stasik, Ro, & Clark, 2013), align with these constructs. More specifically, RCd indexes demoralization, which is a nonspecific distress domain that conjoins negative affectivity and low positive affectivity; as such, RCd is associated with both negative affectivity and detachment in the *DSM-5* trait model, whereas RC7 and RC2 show a clear convergence and divergence with these respective domains (Sellbom & Ban-Porath, 2005).

The RC4 and RC9 scales showed a less distinctive, but nonetheless conceptually expected pattern of associations with disinhibition and antagonism, respectively. As expected, RC9 was the best RC scale predictor of antagonism and was most strongly associated with this domain, whereas RC4 exhibited a larger correlation with the disinhibition domain relative to all others. However, the correlations for RC4 and RC9 with disinhibition were roughly equivalent, with the most notable difference being RC4's larger association with the irresponsibility facet. These findings mirror those of Sellbom et al. (2008) quite well, which showed a very similar pattern of associations with the NEO PI-R Agreeableness and Conscientiousness domains. Furthermore, RC3 was also most strongly associated with antagonism relative to other domains, particularly due to its association with the hostility facet.

The RC6 and RC8 scales, as expected, were both more strongly associated with psychoticism than any other domain. RC8, however, was more strongly associated with this domain than RC6, likely attributable to the low coverage of paranoid ideation in the three originally posited facets of the psychoticism domain. Such content is included in the Suspiciousness facet scale, and MMPI-2-RF scale RC6 (i.e., Ideas of Persecution) has the highest correlations with PID-5 Suspiciousness, which is posited to load on the Negative Affectivity domain of the PID-5. However, this placement may not be accurate. For instance, several studies have found that this facet loads appreciably on multiple domains, often including psychoticism (e.g., De Fruyt et al., 2013; Markon et al., 2013; Quilty, Ayeast, Chmielewski, Pollock, & Bagby, 2013; Watson et al., 2013), Anderson et al. (2013) reported that it correlated equally with PSY-5 Neuroticism/Negative Emotionality and Psychoticism, and Hopwood et al. (2013) found it correlated most strongly with *DSM-IV* Schizotypal PD, indicating relations with psychoticism and social anxiety (a mixture of Negative Affectivity and Detachment), whereas Anderson et al. (2013) and Thomas et al. (2013) found it did not load greater than .30 on any of the five factors. Thus, the exact placement of PID-5 Suspiciousness in the *DSM-5* model remains unclear. Furthermore, in a normal range model of personality indexed

via the MPQ, RC8 is also the best indicator of absorption (Sellbom & Ben-Porath, 2005), but in other models, which do not feature a substantial amount of psychoticism variance (e.g., FMM, SNAP), the RC8 scale demonstrates good discriminant validity (Sellbom et al., 2008; Simms et al., 2005). It should be noted, however, that RC6 and RC8 are better indicators of positive psychotic symptoms than personality traits (Ben-Porath, 2012); as such, when these scales are elevated it is important to consider such symptoms prior to elaborating on potential PD diagnoses.

In the current study, we further demonstrated that more specific MMPI-2-RF scales can also capture the majority of the 25 trait-facets proposed for *DSM-5* Section III. Although there was rarely a one-to-one correspondence, MMPI-2-RF scales and PID-5 facets generally converged as expected. For instance, the AGG scale was clearly a good marker for PID-5 Callousness and ANP for PID-5 Hostility within the Antagonism domain; COG for PID-5 Distractibility within the Disinhibition domain; NFC for PID-5 Perseveration and PID-5 STW for Anxiousness within the Negative Affectivity domain; RC2 for PID-5 Anhedonia, SAV for PID-5 Withdrawal; and DSF for PID-5 Intimacy Avoidance within the Detachment domain. Because the MMPI-2-RF scales do not have narrowband assessment of psychoticism facets, these could not be differentiated; however, RC8 was the best predictor of all three facets. It is noteworthy though, that MMPI-2-RF scales accounted for more variance in trait domains relative to facets, and several PID-5 facets were not captured well by the MMPI-2-RF scales. In particular, the PID-5 (lack of) Rigid Perfectionism and Restricted Affectivity facets did not correlate at a moderate level with any MMPI-2-RF scale. PID-5 Intimacy Avoidance only reached a moderate correlation with DSF. It was not surprising, therefore, that the MMPI-2-RF accounted for the least amount of variance in OCPD, which is comprised of these three facets (in addition to perseveration).

The MMPI-2-RF appears to be in a good position to provide information directly relevant to *DSM-5* Section III PD trait constellations. The great majority of MMPI-2-RF scales hypothesized for each PD contributed unique variance in their prediction; this was particularly the case when scales were examined at each level of the hierarchy (i.e., RC and SP scales). Of note, these findings also suggest that the RC scales follow well replicated patterns of associations with personality psychopathology as observed with the original MMPI Clinical scales, insofar as the 49/94 code type representing Antisocial, 27/72 reflecting Avoidant, and 68/86 indicating Schizotypal PDs (see, e.g., Morey & Smith, 1988), but with better psychometric efficiency and specificity. Indeed, one of the benefits with the Restructured Clinical scales over these original Clinical scales is that these profile configurations might emerge with clearer interpretative meaning now that demoralization and other systematic error variance has been removed (Tellegen et al.,

2006). As such, once these findings have been replicated in other samples, and the *DSM-5* PD model currently placed in Section III is implemented, clinicians who use the MMPI-2-RF will be able to generate hypotheses about various PDs from MMPI-2-RF information. For instance, an individual with elevated scores on RC4 and RC9 coupled with ANP and AGG would likely have characteristics associated with impulsivity, irresponsibility, risk taking, hostility, manipulateness, and deceitfulness, which are reflective of Antisocial PD. On the other hand, a client with a more specific set of an elevated RC9 scale with low scores on IPP and SAV could be a better candidate for a Narcissistic PD diagnosis. Furthermore, elevations on broader internalizing scales, such as RCd and RC7 reflecting nonspecific demoralization and negative affectivity, could be differentiated based on more specific scale elevations. A client with Avoidant PD would potentially be differentiated based on high RC2, SFD, SAV, and DSF scores, whereas the Borderline PD client would not elevate the detachment-relevant scales but rather exhibit a range of elevations across internalizing scales as well as some externalizing scales. Schizotypal PD clients are differentiated from Avoidant PD clients particularly via RC8 and COG, whereas the OCPD patient will likely elevate NFC in addition to the negative affect and detachment markers. Of course, these findings are subject to replication, and it will be particularly important in future to determine the degree to which these PDs (and thus, these proposed MMPI-2-RF scale patterns) are truly more independent (i.e., reduced comorbidity) relative to their original counterparts given the serious limitations associated with these categories in general. In the end, further work along these lines will likely answer the question of whether the attempt to re-invent the problematic *DSM-IV* categories using trait-facets is really the optimal strategy over straightforward descriptions of trait profiles using measures such as the PID-5 or MMPI-2-RF.

Our general conclusions must be considered in light of several limitations that are also grounds for future research directions in this area. First, the configurations of *DSM-5* Section III PDs were completely focused on the Criterion B trait model and did not include Criterion A information (i.e., impairment in personality functioning). Unfortunately, there is currently no standardized assessment for Criterion A available. The current findings can therefore be best conceptualized as assessing the trait profiles associated with *DSM-5* Section III PDs, and future research should examine MMPI-2-RF associations with PDs reflecting the full Section III criterion set. Second, although we used a combined university sample derived from two separate countries, with significant ethnic diversity, affording some degree of enhanced generalizability over a single sample from a single country, the sample is composed of student research participants and most likely limited in the overall range of dysfunction. This truncated range of impairment

could potentially affect both effect size estimates and patterns of associations. Replication in mental health and other settings in which PDs are frequently diagnosed is warranted. Furthermore, although beyond the scope of the current study, it will also be important for future research to examine measurement invariance in *DSM-5* Section III personality traits and disorders across these national and ethnic groups. A final limitation is that both the MMPI-2-RF and PID-5 are self-report inventories, which introduces shared method variance that might artificially inflate correlation magnitudes. It is likely that we partially overcame this limitation by focusing our interpretation of correlations that reached at least a moderate effect size; nevertheless, future studies should examine other measurement modalities of the *DSM-5* Section III trait model, such as the Other-Report Form (Markon et al., 2013) or the APA Trait Rating Form (APA, 2011). Furthermore, future directions of this research should also compare the relative utility of the PID-5, MMPI-2-RF, and other relevant multiscale inventories in predicting clinician ratings of *DSM-5* Section III traits. Such information will be useful to determine the degree to which the MMPI-2-RF provides sufficient coverage relative to the PID-5 in measuring these traits.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Drs. Sellbom and Bagby have both (independently) received grants from the University of Minnesota Press, publisher of the MMPI-2-RF used in this study, to support the research presented in the article.

Notes

1. Attempts have been made to develop PSY-5 scale facets (e.g., Arnau, Handel, & Archer, 2005), but these do not hold up structurally (Quilty & Bagby, 2007).
2. This proportion of invalid protocols is consistent with other MMPI-2 and MMPI-2-RF research studies using university samples (see, e.g., Avdeyeva, Tellegen, & Ben-Porath, 2011; Sellbom, Ben-Porath, Lilienfeld, Patrick, & Graham, 2005).
3. Overall, there were very few differences on MMPI-2-RF and PID-5 scale scores across the two groups, and separate analyses for each sample would by and large have yielded the same conclusions. As such, we combined the samples for ease of presentation and interpretation. Detailed analyses separated for the two university samples are available on request.
4. Full descriptive statistics for PID-5 scale scores are available on request.
5. A thoughtful reviewer recommended using latent covariance modeling (e.g., exploratory structural equation modeling [ESEM]) to determine whether MMPI-2-RF and PID-5 scale

scores converged in expected patterns in latent multivariate space. Because we were interested in both facet- and domain-level relations, we retained the current analyses, as the alternative would only provide information about MMPI-2-RF and PID-5 scale alignment at the domain level. Nevertheless, we did also conduct ESEM analyses, which showed that MMPI-2-RF scale scores did indeed converge with PID-5 scores as expected. These are available from the first author on request.

6. We recognize that item overlap complicates this latter regression equation. We also reestimated these regression equations by removing overlapping items, and this procedure would not have altered the general conclusions. These results are available on request.

References

- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (text revision). Washington, DC: Author.
- American Psychiatric Association. (2011). *American Psychiatric Association DSM-5 development: Personality disorders*. Retrieved from <http://www.dsm5.org/ProposedRevision/Pages/PersonalityDisorders.aspx>
- American Psychiatric Association. (2012). *Rationale for the proposed changes to the personality disorders classification in DSM-5: Part 1*. Retrieved from <http://www.dsm5.org/ProposedRevision/Pages/PersonalityDisorders.aspx>
- Anderson, J. L., Sellbom, M., Bagby, R. M., Quilty, L. C., Veltri, C. O., Markon, K. E., & Krueger, R. F. (2013). On the convergence between PSY-5 domains and PID-5 domains and facets implications for assessment of DSM-5 personality traits. *Assessment, 20*, 286-294.
- Avdeyeva, T. V., Tellegen, A., & Ben-Porath, Y. S. (2012). Empirical correlates of low scores on MMPI-2/MMPI-2-RF restructured clinical scales in a sample of university students. *Assessment, 19*, 388-393.
- Archer, R. P., Buffington-Vollum, J. K., Stredny, R. V., & Handel, R. W. (2006). A survey of psychological tests use patterns among forensic psychologists. *Journal of Personality Assessment, 87*, 85-95.
- Arnau, R. C., Handel, R. W., & Archer, R. P. (2005). Principal components analyses of the MMPI-2 PSY-5 scales: Identification of facet subscales. *Assessment, 12*, 186-198.
- Ayearst, L. E., Sellbom, M., Trobst, K. K., & Bagby, R. M. (2013). Evaluating the interpersonal content of the MMPI-2-RF interpersonal scales. *Journal of Personality Assessment, 95*, 187-196.
- Bagby, R. M., Sellbom, M., Ayearst, L. E., Chmielewski, M. S., Anderson, J. L., & Quilty, L. C. (in press). Exploring the hierarchical structure of the MMPI-2-RF personality psychopathology five in psychiatric patient and university student samples. *Journal of Personality Assessment*.
- Ben-Porath, Y. S. (2012). *Interpreting the MMPI-2-RF*. Minneapolis: University of Minnesota Press.
- Ben-Porath, Y. S., & Tellegen, A. (2008/2011). *Minnesota Multiphasic Personality Inventory-2 Restructured Form: Manual for administration, scoring, and interpretation*. Minneapolis: University of Minnesota Press.

- Butcher, J. N., Graham, J. R., Ben-Porath, Y. S., Tellegen, A., Dahlstrom, W. G., & Kaemmer, B. (2001). *MMPI-2: Manual for administration and scoring* (Revised ed.). Minneapolis: University of Minnesota Press.
- Camara, W. J., Nathan, J. S., & Puente, A. E. (2000). Psychological test usage: Implications in professional psychology. *Professional Psychology: Research and Practice, 31*, 141-154.
- Clark, L. (2007). Assessment and diagnosis of personality disorder: Perennial issues and an emerging reconceptualization. *Annual Review of Psychology, 58*, 227-257.
- Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159.
- De Fruyt, F., De Clercq, B., De Bolle, M., Wille, B., Markon, K., & Krueger, R. F. (2013). General and maladaptive traits in a five-factor framework for DSM-5 in a university student sample. *Assessment, 20*, 295-307.
- Harkness, A. R., Finn, J. A., McNulty, J. L., & Shields, S. M. (2012). The Personality Psychopathology-Five (PSY-5): Recent constructive replication and assessment literature review. *Psychological Assessment, 24*, 432-443.
- Hopwood, C. J., Wright, A. G. C., Krueger, R. F., Schade, N., Markon, K. E., & Morey, L. C. (2013). DSM-5 pathological personality traits and the Personality Assessment Inventory. *Assessment, 20*, 269-285.
- Krueger, R. F., Derringer, J., Markon, K. E., Watson, D., & Skodol, A. V. (2012). Initial construction of a maladaptive personality trait model and inventory for DSM-5. *Psychological Medicine, 42*, 1879-1890.
- Krueger, R. F., Eaton, N. R., Clark, L. E., Watson, D., Markon, K. E., Derringer, J., . . . Livesley, W. J. (2012). Deriving an empirical structure of personality pathology for DSM-5. *Journal of Personality Disorders, 25*, 170-191.
- Markon, K. E., Quilty, L. C., Bagby, R. M., & Krueger, R. F. (2013). The development and psychometric properties of an informant-report form of the PID-5. *Assessment, 20*, 370-383.
- Morey, L. C., & Smith, M. R. (1988). Personality disorders. In R. Greene (Ed.), *The MMPI: Use with specific populations* (pp. 110-158). Needham Heights, MA: Allyn & Bacon.
- Quilty, L. C., Ayeart, L., Chmielewski, M., Pollock, B. G., & Bagby, R. (2013). The psychometric properties of the Personality Inventory for DSM-5 in an APA DSM-5 Field Trial sample. *Assessment, 20*, 362-369.
- Quilty, L. C., & Bagby, R. (2007). Psychometric and structural analysis of the MMPI-2 Personality Psychopathology Five (PSY-5) facet subscales. *Assessment, 14*, 375-384.
- Samuel, D. B., Hopwood, C. J., Krueger, R. F., Thomas, K. M., & Ruggero, C. (2013). Comparing methods for scoring personality disorder types using maladaptive traits in DSM-5. *Assessment, 20*, 353-361.
- Sellbom, M., & Ben-Porath, Y. S. (2005). Mapping the MMPI-2 restructured clinical scales onto normal personality traits: Evidence of construct validity. *Journal of Personality Assessment, 85*, 179-187.
- Sellbom, M., Ben-Porath, Y. S., & Bagby, R. M. (2008). Personality and psychopathology: Mapping the MMPI-2 Restructured Clinical (RC) scales onto the five factor model of personality. *Journal of Personality Disorders, 22*, 291-312.
- Sellbom, M., Ben-Porath, Y. S., Lilienfeld, S. O., Patrick, C. J., & Graham, J. R. (2005). Assessing psychopathic personality traits with the MMPI-2. *Journal of Personality Assessment, 85*, 334-343.
- Simms, L. J., Casillas, A., Clark, L., Watson, D., & Doebbeling, B. N. (2005). Psychometric evaluation of the restructured clinical scales of the MMPI-2. *Psychological Assessment, 17*, 345-358.
- Skodol, A. E. (2012). Personality disorders in DSM-5. *Annual Review of Clinical Psychology, 83*, 317-344.
- Skodol, A. E., Bender, D. S., Morey, L. C., Clark, L. A., Oldham, J. M., Alarcon, R. D., . . . Siever, L. J. (2011). Personality disorder types proposed for DSM-5. *Journal of Personality Disorders, 25*, 136-169.
- Tellegen, A., & Ben-Porath, Y. S. (2008). *The Minnesota Multiphasic Personality Inventory-2 Restructured form: Technical manual*. Minneapolis: University of Minnesota Press.
- Tellegen, A., Ben-Porath, Y. S., Sellbom, M., Arbisi, P. A., McNulty, J. L., & Graham, J. R. (2006). Further evidence on the validity of the MMPI-2 Restructured Clinical (RC) scales: Addressing questions raised by Rogers et al. and Nichols. *Journal of Personality Assessment, 87*, 148-171.
- Thomas, K. M., Yalch, M. M., Krueger, R. F., Wright, A. G. C., Markon, K. E., & Hopwood, C. J. (2013). The convergent structure of DSM-5 personality trait facets and Five Factor Model trait domains. *Assessment, 20*, 308-311.
- Watson, D. (2005). Delineating the structure of normal and abnormal personality: An integrative hierarchical approach. *Journal of Personality and Social Psychology, 88*, 139-157.
- Watson, D., Stasik, S. M., Ro, E., & Clark, L. A. (2013). Integrating normal and pathological personality: Relating the DSM-5 trait dimensional model to general traits of personality. *Assessment, 20*, 312-326.
- Widiger, T. A., & Mullins-Sweatt, S. N. (2010). Clinical utility of a dimensional model of personality. *Professional Psychology: Research and Practice, 41*, 488-494.
- Widiger, T. A., & Simonsen, E. (2005). Alternative dimensional models of personality disorder: Finding a common ground. *Journal of Personality Disorders, 19*, 110-130.
- Widiger, T. A., Simonsen, E., Sirovatka, P. J., & Regier, D. A. (Eds.). (2006). *Dimensional models of personality disorders: Refining the research agenda for DSM-V*. Arlington, VA: American Psychiatric Association.
- Wright, A. G. C., Thomas, K. M., Hopwood, C. J., Markon, K. E., Pincus, A. L., & Krueger, R. F. (2012). The hierarchical structure of DSM-5 pathological personality traits. *Journal of Abnormal Psychology, 121*, 951-957.

PSY640 Week Four Psychological Assessment Report

Patient's Name: Ms. S.
Date of Birth: 01/01/1985
Education: 12 years
Current Medications: None

Date of Evaluation: 07/01/2014
Age: 29 years
Occupation: Student
Handedness: Right

Evaluation Completed by: Dr. K., Licensed Psychologist
Evaluation Time: 1 hour diagnostic interview (90791); 7 hours test administration, scoring, interpretation, and report (96118 x 7)

REASON FOR REFERRAL: Ms. S. was referred by Dr. R.N. for concerns about attentional functioning.

HISTORY OF CURRENT SYMPTOMS: The symptom description and history were obtained from an interview with Ms. S. and a review of her available medical records.

Ms. S. reported a longstanding history of anxiety and depression since high school but stated her symptoms have worsened over the past year; she eventually sought treatment. She reported her anxiety continues to be moderate but is slightly improved, and her depression symptoms have improved significantly with medication. However, she stated she has also experienced problems in attention and concentration in the past several years, and these have not improved despite the noted improvements in her mood symptoms. She reported being referred for a psychiatric evaluation while in the U.S. Army due to her reports to her supervisor that she was experiencing symptoms of acute stress after hearing a gunshot that led to her discovering one of her platoon mates had committed suicide.

Summary of Previous Investigations and Findings: No previous neurological or neuropsychological evaluations.

PAST MEDICAL, NEUROLOGICAL, PSYCHIATRIC, SUBSTANCE USE HISTORY: *(Inclusive review of symptoms and disorders; only positive features listed)* Medical history is significant for reconstructive surgery for a bile duct cyst in 2009 (involving multiple surgeries), activity induced asthma, and irregular menstruation (currently treated with medication). Previous psychiatric history is reported above. Ms. S. stated she does not drink alcohol and has never used tobacco or recreational drugs. Ms. S. stated that she gained over 200 pounds after her discharge from the Army and has attempted to obtain a referral from her physician for bariatric surgery; however, reportedly, her physician has not been willing to recommend her.

BIRTH, DEVELOPMENTAL, OCCUPATIONAL HISTORY: *(Review of perinatal factors, early childhood development and milestones, academic history and achievement, employment)*. Ms. S. denied any problems with her birth or development. She stated math skills were always a relative weakness for her in school, but she was never diagnosed with a learning disability or attention deficit hyperactivity disorder (ADHD). She completed high school and started college immediately after high school but quit after one year due to a lack of focus and financial strain. Subsequently, she enlisted in the army and received an honorable discharge after 9 months due to being psychologically incapable of performing her assigned duties. She worked full time as an inside salesperson until 2012, when she began working as a fitness instructor. She is currently working part time and has been taking classes at a local college; she stated she plans to enroll at a university full time in the fall semester for a bachelor's degree in social work. She has received accommodations (e.g., additional time for examinations, taking tests in a distraction free environment, etc.) at a local school this semester due to her diagnosis of generalized anxiety disorder.

FAMILY HISTORY: *(First degree relatives; only pertinent features reported)*. No significant family history reported.

PSYCHOSOCIAL HISTORY AND CURRENT ADAPTATION: *(Current living situation, social relationships, activities of daily living)* Ms. S. is married and lives with her husband. She has no children. She remains fully independent in all activities of daily living. She stated her hobbies tend to focus on

CONFIDENTIAL

Patient's Name: Ms. S.
Date of Evaluation: 07/01/2014

Page: 2

physically demanding activities such as running or working out. She also stated she enjoys riding her horse and has a large network of social support.

CURRENT EXAMINATION: *Review of records; Clinical Interview; Cognitive Assessment: Wechsler Adult Intelligence Scale-IV (WAIS-IV); Achievement Assessment: Nelson-Denny Reading Test (Form G), Wide Range Achievement Test-4 (WRAT-4); Information Processing: Lexical Fluency, Semantic Fluency, Digit Span, Ruff 2 & 7 Selective Attention Test, Trail Making Test, California Verbal Learning Test-II (CVLT-II), Stroop Color Word Test, Wisconsin Card Sorting Test; Personality Assessment: Minnesota Multiphasic Personality Inventory-2-Restructured Form (MMPI-2-RF); Mood: Beck Depression Inventory (BDI-II), Beck Anxiety Inventory (BAI)*

BEHAVIORAL OBSERVATIONS:

Ms. S. arrived on time for her appointment and was unaccompanied. She was casually dressed, neatly groomed, and her social skills were appropriate. She was fully cooperative throughout the evaluation. Frustration tolerance and task persistence during testing were preserved. There were no behavioral indications of a depression, and a full range of affect was demonstrated. However, Ms. S. appeared highly anxious at the onset of the evaluation and throughout the testing on specific measures (e.g., mental arithmetic) that she perceived as difficult for her. Her anxiety negatively impacted her performance on some measures.

The results of this evaluation are considered reliable and valid for interpretation.

SUMMARY OF FINDINGS:

Raw test scores and standard scores for all measures are listed at the end of the report.

1. **Cognitive Ability:** Ms. S.'s cognitive functioning is within at least the average range based on her performance on the majority of subtests in both verbal and non-verbal (performance) areas on the WAIS-IV. A mild relative weakness was noted on several subtests where she performed in the low average range, but this was most likely secondary to anxiety and therefore her cognitive ability scores are likely an underestimation of her true level of functioning.
2. **Achievement:** With the exception of math computation, which was an area of significant weakness for her, Ms. S. performed in the expected range on other tests of achievement including reading, spelling, and sentence comprehension. On the Nelson Denny Timed Reading Test, she demonstrated a slight but significant benefit from extended time administration (i.e., the total score improved from 37th to 47th percentile). However, her performance on the reading comprehension subtest in both conditions was still below expectation based on her educational level and was likely compromised by her anxiety causing reduced attention.
3. **Information Processing:**
 - a. **Attention:** Ms. S.'s scores on measures of simple attention and working memory were variable, but generally within at least the average range. Her performance was reduced on the arithmetic subtest of the WAIS-IV, but this is likely related to her significant weakness in calculations (as it was also seen on a written calculation test), rather than a working memory impairment. On measure of sustained attention (Ruff 2 & 7 Selective Attention Test), her performance in the areas of speed and accuracy were consistent with that of an individual with ADHD at the 0.01 significance level.
 - b. **Language:** Speech was fluent and adequately articulated, and there was no indication of any type impairment in auditory communication or expressive language.
 - c. **Visuospatial abilities:** No evidence of hemispatial neglect, object agnosia, or other visuospatial deficits.

CONFIDENTIAL

Patient's Name: Ms. S.
Date of Evaluation: 07/01/2014

Page: 3

d. Memory: There was no evidence of a primary retentive memory problem, but Ms. S. demonstrated a mild impairment in initially learning an attention-demanding word list. However, she retained all of the information she had encoded after a delay, and her overall performance was intact. Similarly there was no indication of a retentive memory disturbance for narrative story or figural information on the WMS-IV.

e. Executive functions: Reasoning, planning, and response inhibition were all generally intact. She showed a mild impairment in problem solving on the WCST due to loss of set errors; however, Ms. S. was highly anxious during this test, and that was the likely reason for her poor performance.

4. Personality and Mood: Ms. S. completed the MMPI-2-RF and obtained a valid profile, although her responses on the latter part of the test suggested an exaggerated pattern of reporting; this may be related to her overall level of anxiety and her consequently reduced attention span. Her responses on the basic clinical scales indicate she is experiencing a high degree of psychological distress at this time, including symptoms of tension, depression, and agitation over problems in her environment. She may have some strained interpersonal relationships, and individuals with this profile often feel alienated from others. Her responses suggest she has a somewhat pessimistic outlook on life at this time and may be angry about her personal situation, and at times she may have a tendency to blame others for her problems. While open to psychological treatment, she should be cautious about not terminating treatment too early once her current situational stress is reduced. Ms. S.'s score on the BAI indicated subjective anxiety in the severe range. Her score on the BDI-II indicated mild to moderate depressive mood; however, depressive symptoms did not meet criteria for a depressive disorder.

IMPRESSION/RECOMMENDATIONS:

The test results are consistent with impairments in sustained attention and working memory within the context of an overall at least average level of general intellectual functioning, which are most likely secondary to her diagnosis of generalized anxiety disorder (300.2). The test findings also showed a significant weakness in math skills that would be consistent with a diagnosis of mathematics disorder (315.1) that has likely been present since childhood.

The results of this evaluation indicate Ms. S. meets criteria for a disability relative to her peers and therefore is eligible for appropriate accommodations. The following testing accommodations are recommended:

1. That she be given 100 percent (double time) additional time to complete examinations because of reduced reading comprehension secondary to her anxiety disorder. She should also be allowed to take tests in a separate room to reduce the level of distraction.
2. That she be allowed to use a calculator on standardized testing (e.g., GRE) due to her diagnosis of mathematics disorder.
3. That she be given extra rest breaks during prolonged standardized testing (e.g., GRE) because of the above-noted disabilities.

In addition to the above academic specific recommendations, the following is also recommended to improve her functioning in day-to-day activities of daily living:

4. Although her ability to retain information once it is learned is intact, it is likely the functional memory problems she is experiencing are due to weaknesses in new learning/encoding as the result of attentional factors. Therefore, utilizing behavioral strategies such as the following to facilitate increased attention and recall of newly acquired information in her daily life may be beneficial:
 - Take frequent rest breaks from tedious work.

CONFIDENTIAL

Patient's Name: Ms. S.
Date of Evaluation: 07/01/2014

Page: 4

- Try to make your work environment as distraction-free as possible, such as working in a corner "cubicle" rather than in the middle of the room, or using noise-cancelling headphones when trying to focus on an attention-demanding task.
 - Use written reminders and notes to support verbal learning and recall. If attempting to learn something that you are reading, take written notes while reading, and then review these afterwards.
 - Work on one task at a time until completed. Try to minimize multitasking environments.
 - Write down all important information and upcoming events in one central location, such as a daily planner or appointment book. Do not use sticky notes or other individual pieces of paper to keep track of things, as these are easily lost or mixed up.
 - Keep your personal belongings in the same place in your home. Train yourself to put your wallet, keys, and so forth, in this spot immediately upon entering the house.
5. Ms. S. is welcome to schedule a follow-up one-hour appointment to review and discuss the results of this evaluation.

_____, Ph.D., ABPP-CN
Board Certified Neuropsychologist
Licensed Clinical Psychologist

cc: Dr. RN

Ms. S.

CONFIDENTIAL

Patient's Name: Ms. S.
Date of Evaluation: 07/01/2014

Page: 5

TESTING SUMMARY:	Raw test scores	Normative data	Current Level*
GENERAL FUNCTIONING			
WAIS-IV			
Full Scale IQ	--	SS = 89	Low Average-Average
Verbal Comprehension	--	SS = 88	Low Average-Average
Perceptual Reasoning	--	SS = 86	Low Average
Processing Speed	--	SS = 92	Average
Working Memory	--	SS = 86	Low Average
ATTENTION/PROCESSING SPEED			
WAIS-IV Coding	--	ss = 9	Average
WAIS-IV Symbol Search	--	ss = 8	Average
WAIS-IV Digit Span	5 F, 6 B	ss = 9	Average
WAIS-IV Arithmetic	--	ss = 6	Low Average
Semantic Fluency (total)	23 words	T = 66	Superior
FAS Test (average)	14 words	T = 57	High Average
Trail Making Test Part A	34"	T = 43	Average
Trail Making Test Part B	55"	T = 60	High Average
Ruff 2 & 7 Total Speed	--	T = 40	Low Average
Ruff 2 & 7 Total Accuracy	--	T = 52	Average
Stroop Color Word Score	--	T = 50	Average
VISUOSPATIAL			
WAIS-IV Block Design	--	ss = 7	Low Average
WAIS-IV Visual Puzzles	--	ss = 8	Average
Target Cancellation time (errors)		120" (0)	Within Normal Limits
LANGUAGE			
WAIS-IV Vocabulary	--	ss = 7	Low Average
WAIS-IV Information	--	ss = 8	Average
MEMORY			
CVLT-II			
Learning Trial 1	5/16	z = -1.5	Mild Impairment
Learning Trial 5	14/16	z = 0	Average
Interference Trial	5/16	z = -1.0	Low Average
Short Delay Recall	13/16	z = 0.5	Average
Long Delay Recall	14/16	z = 0.5	Average
Recognition	16/16	z = 0	Average
WMS-IV			
Logical Memory I	21/50	ss = 8	Average
Logical Memory II	19/50	ss = 9	Average
Visual Reproduction I	30/43	ss = 6	Low Average
Visual Reproduction II	19/43	ss = 8	Average
EXECUTIVE FUNCTIONS			
WAIS-IV Similarities	--	ss = 9	Average
WAIS-IV Matrix Reasoning	--	ss = 9	Average
WCST Categories (64 cards)	3/6	z = -1.4	Mild Impairment

*Based on age and/or education-matched normative data (as available)

SS = standard score; mean = 100, standard deviation = 15

ss = scaled score; mean = 10, standard deviation = 3

T = T-score; mean = 50, standard deviation = 10

z = z-score; mean = 0, standard deviation = 1

CONFIDENTIAL

Patient's Name: Ms. S.
Date of Evaluation: 07/01/2014

Page: 6

NELSON-DENNY READING TEST (Form G) - Standard Time Administration			
	Scaled Score	Grade Equivalent	Percentile (Grade 12, end of year norms)
Vocabulary	SS = 209	GE = 13.2	55%
Reading Comprehension	SS = 185	GE = 9.3	22%
TOTAL	SS = 197	GE = 11.1	37%
Reading Rate	SS = 193		37%

NELSON-DENNY READING TEST (Form G) - Extended Time Administration			
	Scaled Score	Grade Equivalent	Percentile (Grade 12 norms)
Vocabulary	SS = 214	GE = 13.8	62%
Reading Comprehension	SS = 190	GE = 9.7	28%
TOTAL	SS = 204	GE = 12.3	47%

WIDE RANGE ACHIEVEMENT TEST-4		
	Grade 12 Norms	
	SS	Level
Word Reading	SS = 97	Average
Sentence Comprehension	SS = 90	Average
Spelling	SS = 100	Average
Math Computation	SS = 56	Severe Impairment

CONFIDENTIAL

**TESTING AND ASSESSMENT:
A GUIDE TO GOOD PRACTICES FOR
WORKFORCE INVESTMENT
PROFESSIONALS**

U.S. Department of Labor
Employment and Training Administration

2006

Foreword

PURPOSE of the GUIDE

Human capital is our most valuable resource in today's globally competitive, highly technical, and increasingly diverse workplace. Investing in the development of our workforce is critical if we, as a nation, are to protect our economic well-being and maintain our standard of living. We develop our workforce in a number of ways by helping both individuals and organizations to identify and cultivate their abilities and competencies. Tests and assessments are often key tools in achieving this important goal.

This Guide serves to help career counselors and training and development professionals use assessment practices in appropriate ways to support the overarching goal of workforce development. It conveys the essential concepts of testing and assessment in easy-to-understand terms so that counselors, trainers, and other workforce development professionals can:

- ☐ evaluate and select assessment tools/procedures that provide career counseling and guidance, and aid in training and career development;
- ☐ administer and score assessment tools that are the most efficient and effective for their particular needs;
- ☐ interpret assessment results in an accurate manner; and
- ☐ understand the professional and legal standards to be followed when using tests and assessments in counseling, training, and other career development programs.

FORMAT of the GUIDE

This Guide is structured around a **set of assessment principles and their applications**. The information is organized so that readers from a variety of backgrounds will find the information presented in a clear and useful manner.

- ☐ Each chapter covers a critical aspect of the assessment process. The issues involved in each aspect are outlined at the beginning of each chapter.
- ☐ Thirteen principles of assessment are explained in the Guide. The last chapter (Chapter 9) summarizes the main points of the principles, serving as a review of the material discussed in the Guide.
- ☐ Appendix A offers a list of resource materials for those interested in more information on a particular topic, and Appendix B is a glossary for quick clarification of terms and concepts.

The Guide is designed to provide important information regarding testing as part of workforce development programs such as career counseling, training, and development. It gives general guidelines and must not be viewed as legal advice.

Acknowledgments

Testing and Assessment: A Guide to Good Practices for Workforce Investment Professionals (hereinafter referred to as the *Guide*) was produced and funded by the U.S. Department of Labor's Employment and Training Administration.

This Guide was prepared under Department of Labor grants with the North Carolina Employment Security Commission, Southern Assessment Research and Development Center, and National O*NET Consortium; the New York Department of Labor; and the Utah Department of Employment Security. The Guide was completed under the direction of David Rivkin and Phil Lewis. This Guide was adapted from *Testing and Assessment: An Employer's Guide to Good Practices*, written by Syed Saad, Gary W. Carter, Mark Rothenberg, and Enid Israelson. Dianne Brown Maranto and Elizabeth McKenzie revised the Guide and adapted it for the workforce investment system. Grateful acknowledgment is made to Patrice Gilliam-Johnson, Jonathan Levine, Cindy Bell, and Brenda Dunn for their contributions. Thanks are also given to Ann Kump, Helen Tannenbaum, Don Kreger, Kristin Fiske, and Marilyn Silver whose valuable suggestions were very much appreciated. Grateful acknowledgment is also made to Suzan Chastain, Department of Labor, Office of the Solicitor, Division of Civil Rights, and Hilary R. Weiner and Cynthia Misicka of the Equal Employment Opportunity Commission for consultant reviews and insights into the final preparation of this Guide.

Special Notice: Intended use of information presented in *Testing and Assessment: A Guide to Good Practices for Workforce Investment Professionals*.

The Guide is designed to provide important information regarding testing and assessment as part of workforce investment program activities such as career counseling, career training, and career development. It summarizes relevant laws and regulations. The information presented is not intended to be exhaustive. Additionally, the user should be aware that relevant new laws and regulations may have been instituted since development of the Guide. The Guide is not to be viewed as legal advice.

Copyright Information

Unless marked with a copyright notice, the materials contained in this Guide are part of the public domain and, therefore, may be used and reproduced without restriction. Any materials in this Guide that are copyrighted and protected under the federal Copyright Act of 1976 are marked with a copyright notice. Permission is granted to use such copyrighted materials for non-commercial, instructional, personal, or scholarly purposes. Any portion of this Guide reproduced for such a purpose must include a reference citation to the publication, including the title/date of the Guide, page number/s, and the authors. Use of the copyrighted material/s for any other purpose, particularly commercial use, without the prior, express written permission of the authors is prohibited.

This project was supported by the U.S. Department of Labor-ETA, under a grant to the North Carolina Employment Security Commission: Grant No. M-6403-7-00-96-60.

Table of Contents

Chapters	Page
Foreword.....	i
Acknowledgments	ii
Special Notice	ii
Copyright Information.....	ii
1 Assessment in Workforce Development: Career Counseling, Training, and Development	1-1
2 Understanding the Professional and Legal Context of Assessment in Workforce Development.....	2-1
3 Understanding Test Quality: Concepts of Reliability and Validity	3-1
4 Assessment Tools and Their Uses in Workforce Development	4-1
5 How to Select Tests: Standards for Evaluating Tests	5-1
6 Administering Assessment Instruments	6-1
7 Using, Scoring, and Interpreting Assessment Instruments	7-1
8 Issues and Concerns with Assessment	8-1
9 A Review: Principles of Assessment.....	9-1
 Appendices	
A Sources of Additional Information on Assessment in Workforce Development.....	A-1
B Glossary of Assessment Terms	B-1

CHAPTER 1 **Assessment in Workforce Development: Career Counseling, Training, and Development**

Assessment is a systematic approach to gathering information about individuals. This information can be used in a variety of ways to aid in workforce development.

Assessment is always conducted for a specific purpose. Career counselors may conduct assessment to provide career guidance to clients. Tests may provide information that helps individuals choose occupations in which they are likely to be successful and satisfied. In training and career development programs, tests are used to help identify employees or individuals who might benefit from either remedial or advanced training, or who are suitable for particular career development tracks. They also can be used to find out whether employees have mastered training materials.

Chapter Highlights

1. Assessment in workforce development
2. Assessment tools: tests and inventories
3. Relationship between workforce development and tests
4. What do tests measure?
5. Importance of using tests in a purposeful manner
6. Limitations of tests: fallibility of test scores

Principles of Assessment Discussed

Use assessment tools in a *purposeful manner*.
Use the *whole-person* approach to assessment.

1. Assessment in workforce development

Assessment can be used as an aid in workforce development in the following ways:

- **Career counseling.** Career counseling is an interactive process by which counselors and clients exchange and explore information concerning clients' backgrounds, experiences, interests, abilities, self-esteem, and other personal characteristics that help or inhibit their work readiness and career planning. Career counseling is a systematic approach to providing information and advice to clients in such areas as outreach programs, training, internships, apprenticeships, and job placement. Although the career counselor's primary concern is the client's career development, counselors also may provide screening and referral services to employers. Counselors use information gathered through assessment to understand and respond to clients' needs and concerns; clients use this information to

understand themselves better, clarify their goals and perspectives, and make plans for the future.

- **Training.** Training is often provided by employers to employees, usually for specific skills and abilities or for company procedures and rules. Training also can be provided by educational agencies, such as community colleges, or by private or governmental service providers as a means to help individuals gain skills to prepare them for employment. Training programs may be developed in-house or purchased off-the-shelf.

Assessment can help identify those applicants and employees who might benefit from either remedial or advanced training. It also is used to monitor individual progress through training. Although the primary focus of this Guide is on the use of assessments for individuals, it is worth noting that assessment also can be used to evaluate the effectiveness of the training program through pre- and post-testing of participants.

- **Development.** Career development programs may be provided by employers or workforce investment agencies to help individuals pursue specific career tracks. A career development program may include specific training, a variety of work assignments, and mentoring. Assessment can be used to select candidates for career development programs, to monitor individual progress through programs, and to aid in individual guidance.

2. Assessment tools: tests and inventories

There are many types of assessment tools that are used in the workforce investment system. These include traditional knowledge and ability tests, personality and interest inventories, and work samples or performance tests. In this Guide, the term *test* will be used as a generic term to refer to any instrument that measures traits, abilities, or other constructs.

Workforce development assessment tools differ in:

- **purpose**, e.g., career counseling, referral, selection, placement, or completion of training and development programs;
- **what they are designed to measure**, e.g., knowledge, skills, abilities, personality traits, work styles, work values, vocational interests, managerial potential, career success, and job satisfaction;
- **format**, e.g., paper-and-pencil, work sample, or computer application; and
- **level of standardization, objectivity, and quantifiability.** Assessment tools and procedures vary greatly on these factors. For example, there are subjective evaluations of informal interviews, highly structured achievement tests, and personality inventories with no specific right or wrong answers.

In training and development programs, there are several ways tests may be used as part of a decision-making process that affects an individual's employment status (e.g., access to training and, therefore, access to opportunities). **All** assessment tools used to make employment-related

decisions, regardless of their format, level of standardization, or objectivity are subject to professional and legal standards. Assessment tools used solely for career exploration or counseling are usually **not** held to these same legal standards. Counselors should be aware of these issues if they engage in referral of clients for employment and also to better understand the proper use of assessments in workforce development efforts.

3. Relationship between workforce development and tests

A test provides only *part of the picture* about a person. On the other hand, the assessment process combines and evaluates all the information gathered about an individual or a group of individuals to provide career guidance or to select individuals for training or development programs. Figure 1 below highlights the relationship between assessment tools and workforce development.

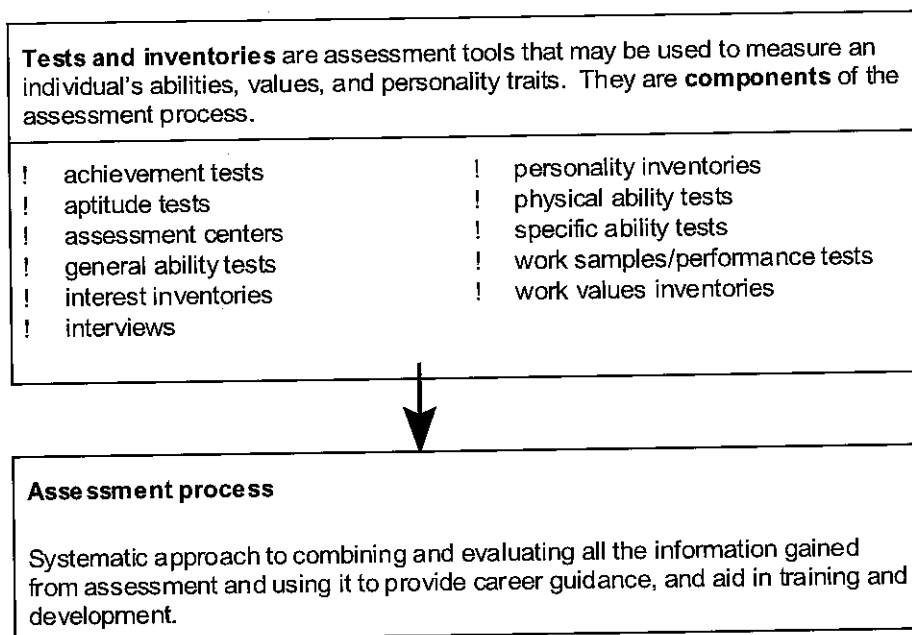


Figure 1. Relationship between assessment tools and the assessment process.

4. What do tests measure?

People differ in their relative knowledge, skills, abilities, competencies, personality, interests, and values. These characteristics are called *constructs*. For example, people skillful in verbal and mathematical reasoning are considered high on *mental ability*. Those who have little physical stamina and strength are assessed low on *endurance* and *physical strength*. The terms

mental ability, endurance, and physical strength are constructs. *Constructs* are used to identify personal characteristics and to distinguish between people in terms of how much they possess of such characteristics.

Constructs cannot be seen or heard, but we can observe their effects on other variables. For example, we don't observe physical strength, but we can observe people with great strength lifting heavy objects and people with limited strength attempting, but failing, to lift these objects. Tests give us information about characteristics we may not otherwise observe that, in turn, can be used to help individuals and organizations to develop their skill base or competencies.

Individuals vary in terms of constructs. These differences systematically affect their job and occupational suitability, career choices, job satisfaction, training needs, and overall career success.

These differences in characteristics are not necessarily apparent by simply observing an individual. In **career counseling**, assessment tools can be used to gather accurate information about career-relevant characteristics. For example, interest inventories are designed to measure people's likes and dislikes for various activities. Scores on an interest inventory summarize interest patterns of the client that can be used to gauge his or her "fit" with different jobs or occupations. Therefore, interest inventories can play an important role in career planning.

For **training**, tests can reveal skill and ability levels of individuals, indicating training needs. For example, a keyboard or typing test might be used to measure an individual's current abilities and can be used to determine whether or not training is needed. Assessment tools also can be used to predict an individual's success in training, aiding in selection of candidates for training programs. To give an example, an employee's score on a mechanical test reflects his or her mechanical aptitude as measured by the test. This score can be used to predict that person's likelihood of success in mechanical training. Skill or ability tests also can be used to assess an individual's progress in training or to evaluate the effectiveness of a training program. By pre- and post-testing trainees, skill or ability tests can tell us how effective the training is in general, not just for an individual.

In **career development** programs, tests can be used to identify and select suitable candidates for certain career tracks. For example, an ability measure might be used to determine managerial potential. Results of this assessment then can be used to place individuals in career tracks that lead to managerial positions. Part of this process also may include career guidance assessment, to help individuals determine their choices for career development programs. For example, a skills assessment might help an individual identify strengths that he or she can link to particular occupations. Tests also can be used to monitor an individual's progress through a specific career development program. They can help determine if the individual is acquiring the necessary skills or knowledge required to move to the next step in their career.

5. Importance of using tests in a purposeful manner

Assessment instruments, like other tools, can be extremely helpful when used properly, but counterproductive when used inappropriately. Often, inappropriate use stems from not having a clear understanding of what you want to measure and why you want to measure it. Having a clear understanding of the purpose of your assessment system is important in selecting the appropriate assessment tools to meet that purpose. This brings us to an important principle of assessment.

Principle of Assessment

Use assessment tools in a *purposeful manner*. It is critical to have a clear understanding of what needs to be measured and for what purpose.

Assessment strategies should be developed with a clear understanding of the knowledge, skills, abilities, characteristics, or personal traits you want to measure. It is also essential to have a clear idea of what each assessment tool you are considering using is designed to measure.

6. Limitations of tests: fallibility of test scores

Professionally developed tests and procedures that are used as part of a planned assessment program may help in career guidance, training, and development. However, it is essential to understand that *all assessment tools are subject to errors*, both in measuring a characteristic, such as verbal ability, and in predicting performance, such as success in training. This is true for all tests and procedures, regardless of how objective or standardized they might be.

- **Do not** expect any test or procedure to measure a personal trait or ability with perfect accuracy for every single person.
- **Do not** expect any test or procedure to be completely accurate in predicting performance or job satisfaction.

There will be cases when a test score or procedure will predict suitability for a particular occupation, when in fact the person would not prove to be satisfied in that occupation. There also will be cases in which an individual receiving a low score will be advised against a particular occupation, when in fact the person actually would be a capable and good worker.

Similarly, there will be cases where a test score or procedure will predict success in training for an individual who then does not succeed. Such errors in this context are called *selection errors*. Selection errors cannot be completely avoided in any assessment program.

Why use testing despite these errors? The answer is that appropriate use of professionally developed assessment tools on average enables individuals and organizations to make more effective decisions than use of simple observations or random decision making.

Using a single test or procedure will provide you with a limited view of a person's career interests or training needs. Moreover, you may reach a mistaken conclusion by giving too much weight to a single test result. On the other hand, using a variety of assessment tools enables you to get a more complete picture of the individual. The practice of using a variety of tests and procedures to more fully assess people is referred to as the *whole-person approach*. This will help reduce the number of selection errors made and will boost the effectiveness of your decision making. This leads to an important principle of assessment.

Principle of Assessment

Do not rely too much on any one test to make decisions. Use the *whole-person approach* to assessment.

CHAPTER 2 Understanding the Professional and Legal Context of Assessment in Workforce Development

This chapter introduces some of the major laws and professional guidelines that govern the use of tests in career counseling, job placement, employment, and/or training and development. It also describes some of the laws and identifies the testing standards adopted by major professional organizations involved with psychology, testing, and measurement.

Chapter Highlights

1. Overview of Pertinent Laws and Guidelines
2. Laws and Guidelines That Apply to Vocational Counseling Programs
 - 2.1 Title IX of the Education Amendments of 1972, as amended
 - 2.2 Section 504 of the Rehabilitation Act of 1973, as amended
 - 2.3 Career Education Incentive Implementation Act of 1977
 - 2.4 Vocational Educational Programs Guidelines of 1979
 - 2.5 Carl D. Perkins Vocational and Training Education Act of 1990, as amended
 - 2.6 Workforce Investment Act of 1998
 - 2.7 Responsibilities of Users of Standardized Tests (RUST) - 2003
 - 2.8 The Program Evaluation Standards - 1994
 - 2.9 Title VI of the Civil Rights Act of 1964, as amended
 - 2.10 Age Discrimination Act of 1975, as amended
 - 2.11 Title II of the Americans with Disabilities Act of 1990, as amended
3. Laws and Guidelines That Apply When Tests and Assessments Are Used in Employment Situations
 - 3.1 Title VII of the Civil Rights Act (CRA) of 1964, as amended
 - 3.2 Age Discrimination in Employment Act of 1967 (ADEA), as amended
 - 3.3 Uniform Guidelines on Employee Selection Procedures - 1978
 - 3.4 Title I of the Americans with Disabilities Act (ADA) of 1990
 - 3.5 Equal Employment Opportunity Commission (EEOC)
 - 3.6 Record-keeping of adverse impact and job-relatedness of tests
 - 3.7 Relationship between federal, state, and local employment laws
 - 3.8 Standards for Educational and Psychological Testing - 1999; Principles for the Validation and Use of Personnel Selection Procedures – 2003

Several laws protect individuals with disabilities. These laws include, but are not limited to, the Workforce Investment Act of 1998 (WIA) and Section 504 of the Rehabilitation Act of 1973, as amended (Section 504), and Americans with Disabilities Act of 1990 (ADA). Any reference in this Guide to ADA should be interpreted to include pertinent provisions of WIA and Section 504.

Principle of Assessment Discussed

Use only assessment instruments that are <i>unbiased</i> and <i>fair</i> to all groups.

1. Overview of Pertinent Laws and Guidelines

The number of laws and guidelines governing workforce development (employment, training, counseling) has increased over the past four decades. This increase resulted from developments in the fields of civil rights, industrial psychology, vocational education, and career counseling. This chapter is presented to provide workforce professionals important information that can help them successfully utilize assessments to assist clients in career development and career exploration.

The laws and guidelines discussed first in this chapter apply most directly to vocational counseling programs, for example, organizations which receive federal financial assistance, such as One-Stop Career Centers and vocational rehabilitation programs. Such programs conduct career counseling, including career exploration, career development, or training. The laws and guidelines discussed second in this chapter apply when employers use tests and other assessments in the context of the employment process. For example, an employer may provide career counseling or training or make an employment selection decision based on validated tests, inventories, and other assessment tools.

The general purpose of the employment laws discussed in this chapter is to prohibit discrimination in employment and provide equal employment opportunity for all. Discrimination occurs when employment decisions are based on race, sex, religion, ethnicity, age, or disability rather than on job-relevant knowledge, skills, abilities, and other characteristics. Employment practices also may be discriminatory if they disproportionately disadvantage a particular group of individuals. Employment practices that discriminate against people are called *unlawful* or *discriminatory employment practices*. Laws and guidelines directed towards organizations receiving federal financial assistance that provide vocational counseling services also are geared towards prohibiting discrimination (e.g., not making decisions based on race, sex, religion, ethnicity, age, or disability) in these types of programs and providing equal access to services for all. The summaries of the professional standards and guidelines in this chapter focus on their impact on workforce development programs.

Before you institute any policies based on these laws and regulations, read the specific laws carefully and consult with your legal advisors regarding the implications for your particular assessment program.

2. Laws and Guidelines That Apply to Vocational Counseling Programs

As discussed above, this first set of information applies most directly to career counseling, career development, and career exploration programs.

2.1 Title IX of the Education Amendments of 1972, as amended

This legislation prohibits recipients of federal financial assistance from discriminating on the basis of sex in educational programs or activities. Specifically, the statute states that “no person in the United States shall, on the basis of sex, be excluded from participation in, be denied the benefits of or be subjected to discrimination under any educational program or activity receiving federal financial assistance.” Title IX seeks to ensure that federal financial assistance is not used to support sex-based discrimination and that individuals have equal opportunities, without regard to gender, to pursue, engage or participate in and benefit from academic, extracurricular, research, occupational training, employment, and other educational programs or activities. To comply with this legislation, your program should use assessments in your workforce investment activities that are fair and unbiased for both men and women.

2.2 Section 504 of the Rehabilitation Act of 1973, as amended

The purpose of this Act is to empower individuals with disabilities to maximize their employment, independence, and integration into society. The Act aims to provide equal opportunity for persons with disabilities, especially in terms of gainful employment and participation in federally-funded programs and activities. Section 504 of the Rehabilitation Act prohibits discrimination on the basis of disability in programs and activities receiving federal financial assistance or conducted by an executive branch agency. No otherwise qualified individual with a disability may be excluded from participation in, or denied the benefits of, or otherwise be subjected to discrimination under any service, program or activity receiving federal financial assistance. Such services, programs, or activities must provide reasonable modifications to allow otherwise qualified individuals with disabilities to participate or benefit unless such modifications would fundamentally alter the nature of the service, program, or activity.

It is important that your program use assessments which are fair and unbiased for persons with disabilities. When selecting assessments for your counseling or training program, think about whether or not they are appropriate for clients with disabilities whom you serve. If an assessment tool or the administration of a test discriminates on the basis of disability, you may be required to provide a reasonable accommodation or modification (e.g., equipment, timing, administration procedures).

2.3 Career Education Incentive Implementation Act of 1977

This legislation further integrated the use of tests in career education, guidance, and counseling programs in public schools. One of its primary contributions has been to extend the use of career assessment instruments into middle and elementary schools. As a service provider, it may be useful for your agency to be knowledgeable about the application of career assessments in your local school district. Clients may come to your agency with previous assessment information that can help inform their career choices.

2.4 Vocational Education Programs Guidelines of 1979

In 1979, the Office of Civil Rights of the Department of Health, Education and Welfare (HEW) issued the *Vocational Educational Programs Guidelines for Eliminating Discrimination and Denial of Services on the Basis of Race, Color, National Origin, Sex and Handicap*. These guidelines were modeled after major civil rights legislation, including the Civil Rights Act of 1964 and the Rehabilitation Act of 1973. They prohibit the use of tests in ways that would deny or limit a person's opportunity to pursue training in vocational education programs based on his or her demographic background. For example, in counseling clients, if you are using a verbal and quantitative abilities test that has adverse impact, you may be limiting your recommendations, or a client's options, for career paths.

2.5 Carl D. Perkins Vocational and Training Education Act of 1990, as amended

The Vocational and Training Education Act promotes the use of federal funds to provide vocational-technical training programs and services to youth and adults and, in particular, special populations. The goals of programs funded by this Act are to provide participants with the academic and vocational skills needed to be successful contenders in a world market. In order for a state to receive funding under the Act, it must submit a state plan. The law states that programs must offer equal access to all participants.

The plans should cover such topics as the:

- type of career counseling and exploration services provided;
- resources and training that will be available to educators charged with carrying out the program;
- communication mechanisms that will be put in place to foster interaction between educators, students, and parents;
- academic and technical training programs that will be implemented, and the skills they will address; and
- programs that will be put in place to prepare participants for entry into post-secondary education or high skill and high wage occupations.

Assessments are a critical part of any plan submitted. The Act states that proper assessments should be used to evaluate the success of the program for each special population, and different assessments may need to be implemented for each special population. In developing and implementing your program, assessments can be used to:

- provide information for career exploration and career counseling;
- objectively identify the education and skill building participants require;
- identify and develop students' career goals and objectives;
- evaluate students' success; and
- evaluate the success of the program itself for particular populations.

The Act provides very specific details of requirements for programs seeking funding under this legislation.

2.6 Workforce Investment Act of 1998

The Workforce Investment Act of 1998 (WIA) replaced the Job Training Partnership Act (JTPA) as the fundamental federal initiative of workforce development. The Act was enacted on August 7, 1998, and became fully effective as of July 1, 2000. Title I builds upon the JTPA and approves the use of federal financial assistance to provide workforce investment activities through statewide and local workforce investment systems to increase employment outcomes and occupational skill attainment of participants. It provides funding for core, intensive, training and supportive services through three main areas of funding streams: youth, adults, and dislocated workers. It creates a One-Stop Delivery System that can be used by interested parties to explore work preparation and career development services and to access a variety of training, employment, and adult and occupational education programs. These One-Stop Delivery Systems offer core employment-related services and access to the other employment and training services funded under the Act as well as under other federal programs.

The federal government provides WIA financial assistance by formula to states. States allocate WIA financial assistance by formula to local areas for appropriate local programs. The level and mix of services provided to participants depends upon a determination of the services needed to obtain or retain employment. Assessments are critical tools in making such determinations. Assessments can assist in identifying their characteristics, including abilities, interests, interpersonal skills, and experience of participants. Results of the assessments should be used to determine the types of intervention, training, and experience individuals need to perform productively in an appropriate work environment. Assessments also can be used to determine individuals' progress in achieving their goals and the goals of the interventions.

2.7 Responsibilities of Users of Standardized Tests (RUST) - 2003

This is the joint policy statement of the Association for Assessment Counselors (AAC) and the American Counseling Association (ACA). RUST is an excellent resource guide on the standards by which tests should be used in counseling. It contains detailed guidelines on test selection, administration, scoring, and interpretation. Also discussed are the necessary qualifications of test users and the procedures for communicating test results to clients.

2.8 The Program Evaluation Standards - 1994

The Program Evaluation Standards address the development and use of assessments in conducting evaluations of educational and training programs and are very useful for professionals who need to monitor and report on the effectiveness of their workforce development programs. Their development was sponsored by the American Association of School Administrators, the American Educational Research Association, the American Evaluation Association, the American Federation of Teachers, the American Psychological Association, the Association for Assessment in Counseling, the Association for Supervision and Curriculum Development, the Canadian Society for the Study of Education, the Council of Chief State School Officers, the Council on Post-Secondary Accreditation, the National Association of Elementary School Principals, the National Association of Secondary School Principals, the National Council on Measurement in Education, the National Education Association, and the National School Boards Association.

2.9 Title VI of the Civil Rights Act of 1964, as amended

Title VI of the Civil Rights Act of 1964 prohibits recipients of federal financial assistance from discriminating on the basis of race, color, or national origin. Implementing regulations state that the registering, counseling, testing, recruitment, selection, referral to job openings, or referral for training opportunities, and all other activities performed by or through government service offices financed in whole or in part from federal funds, must be conducted without regard to race, color, or national origin. Use assessments that are fair and unbiased for persons of different races, color, or national origin.

2.10 Age Discrimination Act of 1975, as amended

This legislation prohibits discrimination on the basis of age in programs and activities receiving federal financial assistance. Unlike the Age Discrimination in Employment Act (ADEA), the Age Discrimination Act applies to all age levels.

To help your program comply with this legislation, you should implement assessments that are unbiased and fair to persons of different ages. In some cases, this might mean using different assessments for clients of different ages. This might be necessary because many assessments are designed for specific age groups.

2.11 Title II of the Americans with Disabilities Act of 1990, as amended

This legislation prohibits state and local governments from discriminating against qualified individuals with disabilities from participating in or deriving benefits from their services, programs, or activities. The state and local governments must make reasonable modifications in the policies, practices, or procedures when necessary to avoid discrimination on the basis of disability, unless making such changes would fundamentally alter the nature of the service, program, or activity.

To help your state or local workforce development program comply with this legislation, be sure to use assessments that are fair and unbiased for persons with disabilities that you serve. Choose assessments that are flexible, so reasonable modifications (equipment, set-up, timing, administration procedures) can be made for your clients with disabilities. For assistance in complying with Title II, you may contact the Department of Justice at (800) 514-0301 (voice) or (800) 514-0383 (TDD).

3. Laws and Guidelines That Apply When Tests and Assessments Are Used in Employment Situations

3.1 Title VII of the Civil Rights Act (CRA) of 1964, as amended

Title VII is legislation that prohibits discrimination in all terms and conditions of employment based on race, color, religion, sex, or national origin. The employment practices covered by this law include the following:

- | | | |
|-------------------------|----------------|----------------------|
| ▪ recruitment | ▪ hiring | ▪ job classification |
| ▪ transfer | ▪ training | ▪ promotion |
| ▪ performance appraisal | ▪ compensation | ▪ union or other |
| ▪ disciplinary action | ▪ termination | membership |
| | | ▪ fringe benefits |

Employers having 15 or more employees, public and private educational institutions, state and local governments, joint (labor-management) committees for apprenticeship and training, employment agencies, and labor unions are subject to this law.

Title VII authorizes the use of professionally-developed ability tests in making employment decisions. However, only instruments that are not designed, intended, or used to discriminate against any particular group can be used. Even professionally-developed tests, however, must meet the requirements of the Uniform Guidelines on Employee Selection Procedures (see Section 3.5 of this chapter).

Another important provision of Title VII relates to the use of group-based test score adjustments to maintain a representative workforce. The Act prohibits score adjustments, the use of different cut-off scores for different groups of test takers, or alteration of employment-related test results based on the demographics of the test takers. Such practices, which are

referred to as *race norming* or *within-group norming*, were used in the past by some employers and government agencies to avoid adverse impact.

3.2 Age Discrimination in Employment Act of 1967 (ADEA), as amended

This Act prohibits discrimination against employees or applicants age 40 or older in all aspects of the employment process, including hiring, firing, promotion, demotion, transfer, compensation, termination, and admission to training programs. Individuals in this group must be provided equal employment opportunity; discrimination in testing and assessment is prohibited. If an older worker charges discrimination under the ADEA, the employer may defend the practice if it can be shown that the job requirement is a matter of *business necessity*. Employers must have documented support for the argument they use as a defense.

The ADEA covers employers having 20 or more employees, labor unions, and employment agencies. Certain groups of employees are exempt from ADEA coverage, including public law enforcement personnel, such as police officers and firefighters. Uniformed military personnel also are exempt from ADEA coverage.

3.3 Uniform Guidelines on Employee Selection Procedures - 1978; adverse or disparate impact, approaches to determine existence of adverse impact, four-fifths rule, job-relatedness, business necessity, biased assessment procedures

In 1978, the EEOC and three other federal agencies (the Civil Service Commission (predecessor of the Office of Personnel Management) and the Labor and Justice Departments) jointly issued the *Uniform Guidelines on Employee Selection Procedures*. The *Guidelines* incorporate a set of principles governing the use of employee selection procedures according to applicable laws. They provide a framework for employers and other organizations for determining the proper use of tests and other selection procedures. The *Guidelines* are legally binding regulations under a number of civil rights laws, including Executive Order 11246, the Workforce Investment Act, and the Wagner-Peyser Act. In reviewing the testing practices of organizations under Title VII, the courts generally give great importance to the *Guidelines'* technical standards for establishing the job-relatedness of tests. Also, federal and state agencies, including the EEOC, apply the *Uniform Guidelines* in enforcing Title VII and related laws.

The *Guidelines* cover employers with 15 or more employees, labor organizations, and employment agencies. They also cover contractors and subcontractors to the federal government and some organizations receiving federal financial assistance. They apply to tests, inventories, and procedures used to make employment decisions. Employment decisions include hiring, promotion, referral, disciplinary action, termination, licensing, and certification. Training may be included as an employment decision if it leads to any of the actions listed above.

One of the basic principles of the *Uniform Guidelines* is that it is unlawful to use a test or selection procedure that creates adverse impact, unless justified. Adverse impact occurs when there is a substantially different rate of selection in hiring, promotion, or other employment decisions that works to the disadvantage of members of a race, sex, or ethnic group.

Different approaches exist that can be used to determine whether adverse impact has occurred. Statistical techniques may provide information regarding whether or not the use of a test results in adverse impact. Adverse impact is normally indicated when the selection rate for one group is less than 80% (4/5) that of another. This measure is commonly referred to as the four-fifths or 80% rule. However, variations in sample size may affect the interpretation of the calculation. For example, the 80% rule may not be accurate in detecting substantially different rates of selection in very large or small samples. When determining whether there is adverse impact in very large or small samples, more sensitive tests of statistical significance should be employed.

The *Guidelines* recommend the following actions when adverse impact occurs:

- Modify the assessment instrument or procedure causing adverse impact.
- Exclude the component procedure causing adverse impact from your assessment program.
- Use an alternative procedure that causes little or no adverse impact, assuming that the alternative procedure is substantially equally valid.
- Use the selection instrument that has adverse impact if the procedure is job-related and justified by business necessity, and there is no equally effective procedure available that has less adverse impact.

Demonstrating job-relatedness of a test is the same as establishing that the test may be validly used as desired. Chapter 3 discusses the concept of test validity and methods for establishing the validity or job-relatedness of a test.

Demonstrating the business necessity of using a particular assessment instrument involves showing that its use is essential to the safe and efficient operation of the business and that there are no alternative procedures available that are substantially equally valid to achieve the business objectives with a lesser adverse impact.

Another issue of importance discussed in the *Uniform Guidelines* relates to test fairness. The *Uniform Guidelines* define biased or unfair assessment procedures as those assessment procedures on which members of one race, sex, or ethnic group characteristically obtain lower scores than members of another group, and the differences in the scores are not reflected in differences in the job performance of members of the groups.

The meaning of scores on an unfair or biased assessment procedure will differ depending on the group membership of the person taking the test. Therefore, using biased tests can prevent employers from making equitable employment decisions. This leads to the next principle.

Principle of Assessment

Use only assessment instruments that are *unbiased* and *fair* to all groups.

Use of biased tools may result in discrimination against members of the lower scoring groups. However, use of fair and unbiased tests can still result in adverse impact in some cases. If you are developing your own test or procedure, expert help may be advisable to make sure your procedure is fair to all relevant groups. If you are planning to purchase professionally-

developed assessment tools, first evaluate the fairness of those you are considering by reading the test manuals and consulting independent reviews.

While the *Uniform Guidelines* focus on employment-related decisions, such as hiring, promotion, termination, or certification, they also can apply to training programs, particularly when tests are used to select individuals for training or career development programs. The *Uniform Guidelines* may apply to career counseling programs when referrals are based on assessments that may have adverse impact. Whether or not the *Uniform Guidelines* apply specifically to your program, you should be well-versed in the concept of adverse impact and the guidelines provided in this document.

3.4 Title I of the Americans with Disabilities Act (ADA) of 1990

Under the ADA, qualified individuals with disabilities must be given equal opportunity in all aspects of employment. The law prohibits employers with 15 or more employees, labor unions, and employment agencies from discriminating against qualified individuals with disabilities. Prohibited discrimination includes a failure to provide reasonable accommodation to persons with disabilities unless doing so would pose an undue hardship. Title I of the ADA also prohibits the denial of admission to an employment-related apprenticeship, counseling, or training program on the basis of a physical or mental disability.

A qualified individual with a disability is one who can perform the essential functions of a job, with or without reasonable accommodation.

- *Disability* is defined broadly to include any actual, perceived, or record of a physical or mental impairment that substantially limits one or more of an individual's major life activities, such as caring for oneself, walking, talking, hearing, or seeing. Some common examples include visual, speech, and hearing disabilities; epilepsy; specific learning disabilities; cancer; serious mental illness; AIDS and HIV infection; alcoholism; and past drug addiction. Noteworthy among conditions not covered are current illegal use of drugs, sexual behavior disorders, compulsive gambling, kleptomania, and pyromania.
- *Essential functions* are the primary job duties that are fundamental, and not marginal, to the job. Factors relevant to determining whether a function is essential include written job descriptions, the amount of time spent performing the function, the consequences of not requiring the function, and the work experiences of employees who hold the same or similar jobs.
- *Reasonable accommodation* is defined as a change in the job application and selection process, a change in the work environment, or a change in the manner in which the work is performed, that enables a qualified person with a disability to enjoy equal employment opportunities. Under this Act, qualified individuals with disabilities must be provided reasonable accommodation so they can perform the essential job functions, as long as this does not create undue hardship to the employer.
- *Undue hardship* is defined as significant difficulty or additional expense and is determined based on a number of factors. Some factors that are considered are the nature and net cost of the accommodation, the financial resources of the facility, the number employed at the facility, the effect on resources and operations, the overall

financial resources of the entire organization, and the fiscal relationship of the facility with the organization. An accommodation that is possible for a large organization may pose an undue hardship for a small organization.

The ADA has major implications for your assessment practices. In training and development programs, the ADA applies to employment opportunities. In employment counseling and guidance, reasonable accommodations in the assessment process also are required in providing services.

- In general, it is the responsibility of the individual with a disability to inform the service provider that an accommodation is needed. When an accommodation is requested, if necessary, the provider may ask for documentation of the applicant's disability and functional limitations for which he or she needs an accommodation.
- Reasonable accommodation may involve making the test site accessible or using an alternative assessment procedure. Administering tests that require individuals with disabilities to use their impaired abilities is prohibited unless the tests are intended to measure one of those abilities.
- One possible alternative procedure would be to use a form of the test that does not require use of the impaired ability. You may be required to provide an accommodation for the administration of the test. Another possibility is to use a procedure that compensates for the impaired ability, if appropriate. For example, allowing extra time to complete certain types of employment tests for someone with dyslexia or other learning disability, or providing a test with larger print, or supplying a reader to a visually impaired individual where appropriate would be considered reasonable accommodations.
- The ADA expressly prohibits making medical inquiries or administering medical examinations prior to making a job offer. This applies to selection of candidates for training and development programs as well, especially if such programs are linked to future employment or promotional opportunities. Before making medical inquiries or requiring medical exams, employers must make an offer of employment to the applicant. Employers may make medical inquiries or require medical exams of an employee only when doing so is work-related and justified by business necessity. All medical information employers obtain about applicants and employees is strictly confidential and must be treated as such. Access to and use of this information is also greatly restricted. Career counselors can help job seekers to understand their rights under the ADA.

Your organization should develop a written policy on conducting testing and assessment of individuals with disabilities. This will help ensure compliance with the provisions of the ADA.

If you need assistance in complying with the ADA, there are several resources you may contact:

- U.S. Equal Employment Opportunity Commission: (202) 663-4494, (800) 669-4000, TTY: (800) 669-6820
- The Job Accommodation Network: (800) 526-7234

- National Business and Disability Council: (516) 465-1515
- The American Foundation for the Blind: (202) 408-0200, (800) 232-5463
- Office of Disability Employment Policy: (866) 633-7365, TTY: (877) 889-5627
- Disability and Business Technical Assistance Centers: (800) 949-4232
- DisabilityInfo.gov (A Web site of information and resources):
<http://www.disabilityinfor.gov>.

3.5 Equal Employment Opportunity Commission (EEOC)

The EEOC is responsible for enforcing many of the federal laws prohibiting employment discrimination, including Title VII, the Age Discrimination in Employment Act, and the Americans with Disabilities Act. It receives, investigates, and processes charges of unlawful employment practices of employers filed by an individual, a group of individuals, or one of its commissioners. If the EEOC determines that there is "reasonable cause" that an unlawful employment practice has occurred, it is also authorized to sue on behalf of the charging individual(s) or itself.

3.6 Record-keeping of adverse impact and job-relatedness of tests

Employers are required to maintain a record of their employment-related activities, including statistics related to testing and adverse impact. Filing and record-keeping requirements for large employers (those with over 100 employees) are generally more extensive than those for employers with 100 or fewer employees. To learn more about the specific requirements, refer to EEOC regulations on record-keeping and reporting requirements under Title VII and the ADA (29 CFR Part 1602) and the *Uniform Guidelines*.

3.7 Relationship between federal, state, and local employment laws

Some states and localities have issued their own fair employment practices laws, and some have adopted the federal *Uniform Guidelines*. These state and local laws may be more stringent than corresponding federal laws. If there is a direct conflict, federal laws and regulations override any contradictory provisions of corresponding state or local laws. You should become thoroughly familiar with your own state and local laws on employment and testing, or consult with appropriate experts, before you initiate and operate an assessment program.

3.8 Standards for Educational and Psychological Testing - 1999; Principles for the Validation and Use of Personnel Selection Procedures - 2003

There are two resource guides published by major organizations in the testing field that will help you set up and maintain an assessment program. The principles and practices presented in these publications set the standards for professional conduct in all aspects of assessment.

- *The Standards for Educational and Psychological Testing*. This publication was developed jointly by the American Psychological Association (APA), the National Council on Measurement in Education (NCME), and the American Educational Research Association (AERA). The *Standards* are an authoritative and comprehensive source of information on how to develop, evaluate, and use tests and other assessment procedures in educational, employment, counseling, and clinical settings. Although developed as professional guidelines, they are consistent with applicable regulations and are frequently cited in litigation involving testing practices.
- *The Principles for the Validation and Use of Personnel Selection Procedures*. This publication was developed by the Society for Industrial and Organizational Psychology (SIOP). Like the *Standards*, the *Principles* also are an excellent guide to good practices in the choice, development, evaluation, and use of assessment tools. However, their main focus is on tools used in the personnel assessment context. The *Principles* explain their relationship to the *Standards* in the following way:

The *Standards* primarily address psychometric issues while the *Principles* primarily address the problems of making decisions in employee selection, placement, promotion, etc. The major concern of the *Standards* is general; the primary concern of the *Principles* is that performance on a test is related to performance on a job, in training, or other measures of job success.

Compatibility of the Standards and the Principles with the Uniform Guidelines

The *Uniform Guidelines* were intended to be consistent with generally accepted professional standards for validating and evaluating standardized tests and other selection procedures. In this regard, the *Guidelines* specifically refer to the *Standards*.

It is strongly encouraged that you become familiar with both the *Standards* and the *Principles* in addition to the *Uniform Guidelines*. Together, they can help you conduct assessment in workforce development programs in a manner consistent with legal and professional standards.

To summarize, this chapter has introduced several laws, regulations, standards, and guidelines that are relevant to assessment in workforce development and employment. Understanding them will help you to conduct assessments that are fair and useful within your workforce development program. Many of the laws cited in this chapter focus on employment-related decisions. However, they may relate to the use of assessments in training and development programs or in job referral. Your knowledge of such laws and regulations can help you assist clients or employees in their development process.

10

20

30

CHAPTER 3 Understanding Test Quality: Concepts of Reliability and Validity

Test *reliability* and *validity* are two technical properties of a test that indicate the quality and usefulness of the test. These are the two most important features of a test. You should examine these features when evaluating the suitability of the test for your use. This chapter provides a simplified explanation of these two complex ideas. These explanations will help you to understand reliability and validity information reported in test manuals and independent reviews and to use that information to evaluate the suitability of a test in your workforce development program.

Chapter Highlights

1. What makes a good test?
2. Test reliability
3. Interpretation of reliability information from test manuals and reviews
4. Types of reliability estimates
5. Standard error of measurement
6. Test validity
7. Methods for conducting validation studies
8. Using validity evidence from outside studies
9. How to interpret validity information from test manuals and independent reviews

Principles of Assessment Discussed

Use only reliable assessment instruments and procedures.

Use only assessment procedures and instruments that have been demonstrated to be valid for the specific purpose for which they are being used.

Use assessment tools that are appropriate for the target population.

1. What makes a good test?

A test used in career guidance or for training and development is considered good if the following can be said about it:

- The test measures what it claims to measure. For example, a test of mental ability does in fact measure mental ability and not some other characteristic.

- The test measures what it claims to measure consistently or reliably. This means that if a person were to take the test again, the person would get a *similar* test score.
- The test is purpose-relevant. In other words, the test measures one or more characteristics that are important to specific career decisions or for predicting or monitoring training and development outcomes.
- By using the test, more effective decisions can be made by and about individuals. For example, an interest inventory helps you to guide a client toward careers in which he or she is more likely to be satisfied. A mechanical aptitude test may help you predict who would benefit from mechanical training.

The degree to which a test has these qualities is indicated by two technical properties: reliability and validity.

2. Test reliability

Reliability refers to how dependably or consistently a test measures a characteristic. If a person takes the test again, will he or she get a similar test score, or a much different score? A test that yields similar scores for a person who repeats the test is said to measure a characteristic reliably.

How do we account for an individual who does not get exactly the same test score every time he or she takes the test? Some possible reasons are the following:

- **Test taker's temporary psychological or physical state.** Test performance can be influenced by a person's psychological or physical state at the time of testing. For example, differing levels of anxiety, fatigue, or motivation may affect an individual's test results.
- **Environmental factors.** Differences in the testing environment, such as room temperature, lighting, noise, or even the test administrator, can influence an individual's test performance.
- **Test form.** Many tests have more than one version or form. Items differ on each form, but each form is supposed to measure the same thing. Different forms of a test are known as *parallel forms* or *alternate forms*. These forms are designed to have similar measurement characteristics, but they contain different items. Because the forms are not exactly the same, a test taker might do better on one form than on another.
- **Multiple raters.** In certain tests, scoring is determined by a rater's judgments of the test taker's performance or responses. Differences in training, experience, and frame of reference among raters can produce different test scores for the test taker.

These factors are sources of chance or random measurement error in the assessment process. If there were no random errors of measurement, the individual would get the same test score, the individual's "true" score, each time. The degree to which test scores are unaffected by measurement errors is an indication of the **reliability of the test**.

Reliable assessment tools produce dependable, repeatable, and consistent information about people. In order to meaningfully interpret test scores and make useful training, development, or career-related decisions, you need reliable tools. This brings us to the next principle of assessment.

Principle of Assessment

Use only reliable assessment instruments and procedures. In other words, use only assessment tools that provide dependable and consistent information.

3. Interpretation of reliability information from test manuals and reviews

Test manuals and independent reviews of tests provide information on test reliability. The following discussion will help you interpret the reliability information about any test.

The reliability of a test is indicated by the **reliability coefficient**. It is denoted by the letter r and is expressed as a number ranging between 0 and 1.00, with $r = 0$ indicating no reliability, and $r = 1.00$ indicating perfect reliability. Do not expect to find a test with perfect reliability. Generally, you will see the reliability of a test as a decimal, for example, $r = .80$ or $r = .93$. The larger the reliability coefficient, the more repeatable or reliable the test scores. Table 1 serves as a general guideline for interpreting test reliability. However, do **not** select or reject a test solely based on the size of its reliability coefficient. To evaluate a test's reliability, you should consider the type of test, the type of reliability estimate reported, and the context in which the test will be used.

Table 1. General Guidelines for Interpreting Reliability Coefficients

Reliability coefficient value	Interpretation
.90 and up	excellent
.80 - .89	good
.70 - .79	adequate
below .70	may have limited applicability

4. Types of reliability estimates

There are several types of reliability estimates, each influenced by different sources of measurement error. Test developers have the responsibility of reporting the reliability estimates that are relevant for a particular test. Before deciding to use a test, read the test manual and any independent reviews to determine if its reliability is acceptable. The acceptable level of reliability will differ depending on the type of test and the reliability estimate used.

The discussion in Table 2 should help you develop some familiarity with the different kinds of reliability estimates reported in test manuals and reviews.

Table 2. Types of Reliability Estimates

-
- **Test-retest reliability** indicates the repeatability of test scores with the passage of time. This estimate also reflects the stability of the characteristic or construct being measured by the test.
Some constructs are more stable than others. For example, an individual's reading ability is more stable over a particular period of time than that individual's anxiety level. Therefore, you would expect a higher test-retest reliability coefficient on a reading test than you would on a test that measures anxiety. For constructs that are expected to vary over time, an acceptable test-retest reliability coefficient may be lower than is suggested in Table 1.

 - **Alternate or parallel form reliability** indicates how consistent test scores are likely to be if a person takes two or more forms of a test.
A high parallel form reliability coefficient indicates that the different forms of the test are very similar, which means that it makes virtually no difference which version of the test a person takes. On the other hand, a low parallel form reliability coefficient suggests that the different forms are probably **not** comparable; they may be measuring different things and, therefore, cannot be used interchangeably.

 - **Inter-rater reliability** indicates how consistent test scores are likely to be if the test is scored by two or more raters.
On some tests, raters evaluate responses to questions and determine the score. Differences in judgments among raters are likely to produce variations in test scores. A high inter-rater reliability coefficient indicates that the judgment process is stable and the resulting scores are reliable.
Inter-rater reliability coefficients are typically lower than other types of reliability estimates. However, it is possible to obtain higher levels of inter-rater reliabilities if raters are appropriately trained.

 - **Internal consistency reliability** indicates the extent to which items on a test measure the same thing.
A high internal consistency reliability coefficient for a test indicates that the items on the test are very similar to each other in content (homogeneous). It is important to note that the length of a test can affect internal consistency reliability. For example, a very lengthy test can spuriously inflate the reliability coefficient.
Tests that measure multiple characteristics are usually divided into distinct components. Manuals for such tests typically report a separate internal consistency reliability coefficient for each component in addition to one for the whole test.
Test manuals and reviews report several kinds of internal consistency reliability estimates. Each type of estimate is appropriate under certain circumstances. The test manual should explain why a particular estimate is reported.
-

5. Standard error of measurement

Test manuals report a statistic called the **standard error of measurement (SEM)**. It gives the margin of error that you should expect in an individual test score because of imperfect reliability of the test. The SEM represents the degree of confidence that a person's "true" score lies within a particular range of scores. For example, an SEM of "2" indicates that a test taker's "true" score probably lies within 2 points in either direction of the score he or she receives on the test. This means that if an individual receives a 91 on the test, there is a good chance that the person's "true" score lies somewhere between 89 and 93. The SEM is a useful measure of the accuracy of individual test scores. The smaller the SEM, the more accurate the measurement.

When evaluating the reliability coefficients of a test, it is important to review the explanations provided in the manual for the following:

- **Types of reliability used.** The manual should indicate why a certain type of reliability coefficient was reported. The manual should also discuss sources of random measurement error that are relevant for the test.
- **How reliability studies were conducted.** The manual should indicate the conditions under which the data were obtained, such as the length of time that passed between administrations of a test in a test-retest reliability study. In general, reliabilities tend to drop as the time between test administrations increases.
- **The characteristics of the sample group.** The manual should indicate the important characteristics of the group used in gathering reliability information, such as educational level, occupation, etc. This will allow you to compare the characteristics of the people you want to test with the sample group. If they are sufficiently similar, then the reported reliability estimates will probably hold true for your population as well.

For more information on reliability, consult the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association and the National Council on Measurement in Education), the Society for Industrial and Organizational Psychology's *Principles for the Validation and Use of Personnel Selection Procedures*, or any major textbook on psychometrics or employment testing. Appendix A lists some possible sources.

6. Test validity

Validity is the most important issue in selecting a test. Validity refers to the *characteristic* that the test measures and *how well* the test measures that characteristic.

- Validity tells you, for example, if the characteristic being measured by a test is related to job satisfaction or training achievement.
- Validity gives *meaning* to the test scores. Validity evidence indicates that there is linkage between test performance and the construct it is intended to measure. It can tell you what you may conclude or predict about someone from his or her score on the test.
- Validity also describes the *degree* to which you can make specific conclusions or predictions about people based on their test scores. In other words, it indicates the usefulness of the test.

It is important to understand the differences between *reliability* and *validity*. Validity will tell you how good a test is for a particular situation; reliability will tell you how trustworthy a score on that test will be. You cannot draw valid conclusions from a test score unless you are sure that the test is reliable. Even when a test is reliable, it may not be valid. You should be careful that any test you select is both reliable and valid for your situation.

A test's validity is established in reference to a specific purpose; the test may not be valid for different purposes. For example, the test you use to make valid predictions about someone's career satisfaction may not be valid for predicting his or her likelihood of success in leadership training. This leads to the next principle of assessment.

Principle of Assessment

Use only assessment procedures and instruments that have been demonstrated to be valid for the specific purpose for which they are being used.

Similarly, a test's validity is established in reference to specific groups. These groups are called the *reference groups*. The test may not be valid for different groups. For example, an assessment tool designed to determine the work values of adults may not be suitable for use with junior high school students. It is possible that many junior high school students are not vocationally mature enough for the assessment to be meaningful for them, or perhaps the reading level of the measure is not suitable for junior high school students. Similarly, a test used to predict success of managers in leadership skills training might not be suitable for computer technicians, even if the training program is similar. Computer technicians can be expected to have had very different backgrounds with regard to leadership skills than managerial staff, making the test valid for managers but not computer technicians.

Test developers have the responsibility of describing the reference groups used to develop the test. The manual should describe the groups for whom the test is valid and the interpretation of scores for individuals belonging to each of these groups. You must determine if the test can be used appropriately with the particular type of people you want to test. This group of people is called your *target population* or *target group*.

Principle of Assessment

Use assessment tools that are appropriate for the target population.

Your target group and the reference group do **not** have to match on all factors; they must be sufficiently similar so that the test will yield meaningful scores for your group. For example, a writing ability test developed for use with high school seniors may be appropriate for measuring the writing ability of administrative assistants, even though these groups do not have identical characteristics. In determining the appropriateness of a test for your target groups, consider factors such as occupation, reading level, cultural differences, and language barriers.

Recall that the *Uniform Guidelines* require assessment tools to have adequate supporting evidence for the conclusions you reach with them in the event adverse impact occurs. A valid assessment tool is one that measures an important characteristic of interest to you. Use of valid tools will, generally, enable you to make better decisions about career choices and training and development needs. Both from business-efficiency and legal viewpoints, it is essential to use only tests that are valid for your intended use.

7. Methods for conducting validation studies

The *Uniform Guidelines* discusses the following three methods of conducting validation studies and focuses mainly on the job-relatedness of a test, as pertinent to employment selection. This is also relevant to selection for training and career development programs. The *Guidelines* describes conditions under which each type of validation strategy is appropriate. It does not express a preference for any one strategy to demonstrate the job-relatedness of a test.

- **Criterion-related validation** requires demonstration of a correlation or other statistical relationship between test performance and job performance. In other words, individuals who score high on the test tend to perform better on the job than those who score low on the test. If the criterion is obtained at the same time the test is given, it is called concurrent validity; if the criterion is obtained at a later time, it is called predictive validity.
- **Content-related validation** requires a demonstration that the content of the test represents important job-related behaviors. In other words, test items should be relevant to and measure directly important requirements and qualifications for the job.
- **Construct-related validation** requires a demonstration that the test measures the construct or characteristic it claims to measure, and that this characteristic is important to successful performance on the job.

More recently, professional standards and scientific references (Guion, 1980; Messick, 1989; Cronbach, 1988; Camara & Brown, 1995) are discussing validity in slightly different terms. Construct validity encompasses all other forms of validity, and validation is the cumulative and on-going process of giving meaning to test scores. Rather than the traditional “3 Cs of Validity: Content, Criterion, & Construct,” validity is viewed as a unitary concept which may have many forms of evidence.

So, the overarching concept of validity is construct validity. A test must always measure some construct(s), and the validity evidence you collect for the test demonstrates its ability to measure the specified construct(s). This evidence may be obtained from a variety of sources, including test content, response processes, internal structure of the test, and external criteria. You may have more than one form of validity evidence for a single test.

As discussed in Chapter 1, a construct is any variable measured by a test: a trait, characteristic, knowledge, skill, ability, or interest, for example. Evidence that a test is measuring the construct(s) it is intended to measure can come from a variety of sources.

Content evidence relates the content of the test to the construct being assessed. This could be determined through subject matter experts and often through analyzing documentation of the development of a test, which should define the domain being assessed. For example, a test used to measure achievement in pilot training would be developed based on a thorough identification of the domain of knowledge, skills, and abilities necessary for being a pilot. This could be determined using pilots’ expert judgments to identify the domain and to determine whether the test’s items, structure, and format adequately measure the content.

Evidence of validity also can be obtained from an analysis of response processes. For example, a test of mathematical reasoning can be analyzed to determine whether examinees are performing the mental processes the test is intended to measure. If so, this is another form of validity evidence. The internal structure of a test can be studied to determine whether it supports validity for a test. Item difficulty levels can be examined to see if they reflect the intended level of difficulty of the construct being measured.

Tests also can be validated with evidence of their relationship to relevant criteria. Some examples of criteria are success in training, success on the job, or job satisfaction. Criterion-related evidence of validity is especially important when we are using a test to predict something. If the test can be shown to be related to the criteria it is intended to predict, this is evidence of the test's validity. For example, if a test is meant to predict successful completion of training in mechanics, it should be related to an assessment of training performance. If so, it has evidence of validity for this use and would be appropriate to use in selecting candidates for training.

Professionally developed tests should come with reports on validity evidence, including detailed explanations of how validation studies were conducted. If you develop your own tests or procedures, you will need to conduct your own validation studies. As the test user, you have the ultimate responsibility for making sure that validity evidence exists for the conclusions you reach using the tests. This applies to all tests and procedures you use, whether they have been bought off-the-shelf, developed externally, or developed in-house. If you purchase a professionally developed training program, you also must evaluate its assessment components for validity.

Validity evidence is especially critical for tests that have adverse impact. When a test has adverse impact, the *Uniform Guidelines* require that validity evidence for that specific employment decision be provided. This is especially relevant for tests used for selection into training and development programs. In career counseling, adverse impact rules may apply in the use of assessments for job referral.

When training is required for a particular job, it should be highly related to performing that job, which can be determined partly through a job analysis. **Job analysis** is a systematic process used to identify the tasks, duties, responsibilities, and working conditions associated with a job and the knowledge, skills, abilities, and other characteristics required to perform that job.

Job analysis information may be gathered by direct observation of people currently in the job, interviews with experienced supervisors and job incumbents, questionnaires, personnel and equipment records, and work manuals. In order to meet the requirements of the *Uniform Guidelines*, it is advisable that the job analysis be conducted by a qualified professional, for example, an industrial and organizational psychologist or other professional well trained in job analysis techniques. Job analysis information is central in determining training objectives, what to test for, and which tests to use.

8. Using validity evidence from outside studies

Conducting your own validation study is expensive, and, in many cases, you may not have enough individuals or other resources to make it feasible to conduct a study. Therefore, you may find it advantageous to use professionally developed assessment tools and procedures for which documentation on validity already exists. However, care must be taken to make sure that validity evidence obtained for an “outside” test study can be suitably “transported” to your particular situation, especially for training and development purposes.

The *Uniform Guidelines*, the *Standards*, and the *SIOP Principles* state that evidence of transportability is required. Consider the following when using outside tests:

- **Validity evidence.** The validation procedures used in the studies must be consistent with accepted standards.
- **Job similarity.** A job analysis should be performed to verify that your job or training program and the original job or training program are substantially similar in terms of ability requirements and work behavior.
- **Fairness evidence.** Reports of test fairness from outside studies must be considered for each protected group that is part of your labor market. When this information is not available for an otherwise qualified test, an internal study of test fairness should be conducted, if feasible.
- **Other significant variables.** These include the type of performance measures and standards used, the essential work activities performed, the similarity of your target group to the reference samples, as well as all other situational factors that might affect the applicability of the outside test for your use.

To ensure that the outside test you purchase or obtain meets professional and legal standards, you should consult with testing professionals. See Chapter 5 for information on testing consultants.

9. How to interpret validity information from test manuals and independent reviews

To determine if a particular test is valid for your intended use, consult the test manual and available independent reviews. (Chapter 5 offers sources for test reviews.) The information below can help you interpret the validity evidence reported in these publications.

- In evaluating validity information, it is important to determine whether the test can be used in the specific way you intended and whether your target group is similar to the test reference group.

Test manuals and reviews should describe:

- Available validation evidence supporting use of the test for specific purposes. The manual should include a thorough description of the procedures used in the validation studies and the results of those studies.
 - The possible valid uses of the test. The purposes for which the test can legitimately be used should be described, as well as any performance criteria that can be validly predicted.
 - The sample group(s) on which the test was developed. For example, was the test developed on a sample of high school graduates, managers, or clerical workers? What was the racial, ethnic, age, and gender mix of the sample?
 - The group(s) for which the test may be used.
- The *criterion-related validity* of a test is measured by the *validity coefficient*. It is reported as a number between 0 and 1.00 that indicates the magnitude of the relationship r between the test and a measure of training performance (criterion). The larger the validity coefficient, the more confidence you can have in predictions made from the test scores. However, a single test can never fully predict training performance because success in training depends on so many varied factors. Therefore, validity coefficients, unlike reliability coefficients, rarely exceed $r = .40$.

Table 3. General Guidelines for Interpreting Validity Coefficients

Validity coefficient value	Interpretation
above .35	very beneficial
.21 - .35	likely to be useful
.11 - .20	depends on circumstances
below .11	unlikely to be useful

As a general rule, the higher the validity coefficient, the more beneficial it is to use the test. Validity coefficients of $r = .21$ to $r = .35$ are typical for a single test and are considered useful. Validities for selection systems that use multiple tests will

probably be higher because different tools are being used to measure/predict different aspects of performance, whereas a single test is more likely to measure or predict fewer aspects of total performance. Table 3 serves as a general guideline for interpreting test validity for a single test.

For example, an abilities test with a reported validity coefficient of .30 is likely to be useful and would be appropriate to use in screening candidates for training. However, the same test with a validity coefficient of .10 is unlikely to be useful and should not be used in decision-making. An abilities test with a validity coefficient of .18 might be useful to monitor progress in training, but not to select candidates into development programs.

Evaluating test validity is a sophisticated task, and you might require the services of a testing expert/consultant. In addition to the magnitude of the validity coefficient, you should also consider at a minimum the following factors:

- Level of adverse impact associated with your assessment tool
- Selection ratio (number of applicants versus the number of openings)
- Cost of a referral error

- Cost of the assessment tool
- Probability of referring qualified individuals based on chance alone.

Here are three scenarios illustrating why you should consider these factors, individually and in combination with one another, when evaluating validity coefficients:

Scenario One

An employer has asked your agency to screen clients for a number of entry-level positions in their company. They inform you that they have not been satisfied with the performance of workers that they have hired in a number of positions. The company expects some level of employee error, but wants to reduce errors as part of a cost saving measure they want to put into operation. Therefore, in hiring for new openings in these positions, they would like your agency to screen potential applicants with an assessment tool before sending them for an interview for the job openings. Training time is relatively short for these positions since they are not highly technical and do not require a great deal of skill.

It is determined that many of the errors made by the current employees are due to poor clerical perception. You find two suitable clerical perception assessment tools. The first one has high validity, however, the cost is prohibitive and the administration and scoring time for your staff is extensive. The second assessment tool has a reasonable cost and has an administration and scoring time that is more in line with your agency's resources; however, the validity is considered only "likely to be useful." It is decided to use the second assessment tool for several reasons: it will be used only as one part of the selection process for this employer, it is likely to improve employee job performance, it is reasonably priced, it is fairly quick to administer and score, and no adverse impact has been associated with it to date.

Now, let's change the situation.

Scenario Two

Based on requests from local employers who are looking for highly capable applicants to fill a demanding and technical job, your agency is sponsoring a technical training course for customers in your area. The companies with job openings have informed you that for these particular positions, accuracy is very important; and employee errors could be both dangerous and expensive. Therefore, using an assessment tool that will help select the most competent trainees for the program is important. In addition, the significant expense of training is wasted whenever trainees drop out of the training program before completion or finish without acquiring the necessary competencies. After researching a number of assessment tools, an assessment is identified that will be used as one part of the screening process for potential trainees. While this assessment is a highly valid predictor of success for this occupation, it takes quite a bit of time to administer and is expensive. Nevertheless, your agency decides to use this assessment because its validity is considered to be "very beneficial" and no adverse impact has been associated with it to date. An assessment with lower validity would most likely not be acceptable considering the circumstances and the companies' job requirements. For persons who are not accepted into this training program, it

is important to provide career exploration activities and refer them to alternative opportunities to train for occupations in demand.

Here is another scenario that shows why you need to consider multiple factors when evaluating the validity of assessment tools.

Scenario Three

A company you are working with asks your agency to administer an assessment as part of your referral process to their company. When reviewing the documentation on this assessment you learn that even though it is considered to be a very effective selection system for the occupation, it also results in fairly high levels of adverse impact. In addition, you are concerned about the high cost of this particular assessment. Therefore, you conduct additional research to identify other possible assessment tools that might have less adverse impact and are less costly. In conducting your research, you carefully document your findings. Your research concludes that the other available assessment tools with lower adverse impact had substantially lower validity and were just as expensive. In addition, the selection tool will be used as only one part of the screening process. Other factors of consideration included the difficulty in hiring for this position, the “very beneficial” validity of the assessment, and the failed attempts to find equally valid alternative instruments with less adverse impact. As always, you consider any applicable state testing policies when making your assessment decisions. You also consult appropriate legal counsel to assure that your agency is in compliance with all applicable laws. Your agency decides that the selection tool is acceptable because making mistakes in hiring decisions would represent too great a risk for the company. Your agency plans to conduct periodic searches for new assessment tools that have “very beneficial validity” and with less adverse impact. In the meantime, your agency will use this assessment tool as part of the referral process.

Again, these examples demonstrate the complexity of evaluating the validity of assessments. Multiple factors often need to be considered. You might want to seek the assistance of a testing expert (for example, an industrial/organizational psychologist) to evaluate the appropriateness of particular assessments for various workforce investment customers and their individual situations.

When properly applied, the use of valid and reliable assessment instruments will help you and the clients you serve business and job seekers make better decisions. Additionally, by using a variety of assessment tools as part of an assessment program, you can more fully assess the skills and capabilities of individuals, while reducing the effects of errors associated with any one tool.

CHAPTER 4 Assessment Tools and Their Uses in Workforce Development

This chapter briefly describes different types of assessment tools and procedures that are used in workforce development programs. Included are techniques used in career guidance and counseling, training, and development programs. Results can help identify training needs, strengths, and possible career paths. Table 4, which appears at the end of this chapter, contains a brief description of the advantages and disadvantages of different types of assessment instruments.

Chapter Highlights

1. Preliminary assessment (career counseling)
2. Mental and physical ability tests
3. Achievement tests
4. Interest measures
5. Work and personal values measures
6. Personality inventories
7. Interviews
8. Assessment centers
9. Comprehensive career planning measures

It takes a good deal of knowledge and judgment to properly use assessment tools to make effective decisions regarding career counseling, training, and development. Many assessment tools and procedures require specialized training, education, or experience to administer and interpret correctly. These requirements vary widely, depending on the specific instruments being used. Check with the test publisher to determine whether you and your staff meet these requirements. To ensure that test users have the necessary qualifications, some test publishers and distributors require proof of qualifications before they will release certain tests.

1. Preliminary assessment

In counseling, the process of assessment begins before the first interview when the counselor reads the initial contact form completed by the client. This intake form often contains questions which can guide the initial counseling interview. Typical items include the client's name, address, age, educational and work history, reason for referral to counseling, and whether the client has been or is being counseled elsewhere. If the intake form is kept brief, it usually does not present a barrier to the client.

The intake form can help direct the initial interview, guiding the discussion toward the client's most significant problems. The assessment process will enable the counselor to evaluate the

client's needs and will help the counselor to select the most appropriate type of assessment tools for the individual. The counselor also can learn at this time whether the client was referred by another agency and whether any assessment was done at that agency.

2. Mental and physical ability tests

When properly applied, ability tests are among the most useful and valid tools available for predicting success in jobs and in training and for identifying training needs across a wide variety of occupations. Counselors and job placement staff frequently use ability tests when the client is considering an occupation or training program in which he or she has not had the opportunity to demonstrate competence. For example, if a client is considering a career in bookkeeping, a counselor might administer a quantitative abilities test to determine whether or not the individual has the necessary abilities for that occupation. Ability tests also can be used to identify employees or individuals who could benefit from training and to assess progress in training. A mechanical abilities test might be used to determine whether employees need training for new mechanical equipment or whether they have mastered training material. Mental ability tests are generally used to measure the ability to learn and perform particular job responsibilities. Like physical ability tests, they can be used in training and development programs. Examples of some mental abilities are verbal, quantitative, and spatial abilities. Physical ability tests usually encompass abilities such as strength, endurance, and flexibility.

- **General ability tests** typically measure one or more broad mental abilities, such as verbal, mathematical, and reasoning skills. These skills are fundamental to success in many different kinds of jobs and training, especially where cognitive activities such as reading, computing, analyzing, or communicating are involved.
- **Specific ability tests** include measures of distinct physical and mental abilities, such as reaction time, written comprehension, mathematical reasoning, and mechanical ability, which are important for many jobs and occupations. For example, good mechanical ability may be important for success in auto mechanic and engineering jobs; physical endurance may be critical for fire-fighting jobs.

Although mental ability tests are valid predictors of performance in many jobs, use of such tests in selection or referral contexts often results in adverse impact. For example, research suggests that mental abilities tests adversely impact some racial minority groups and, if speed is also a component of the test, older workers may be adversely impacted. Similarly, use of physical ability tests often results in adverse impact against women and older persons. See Chapter 7 for strategies to minimize adverse impact in your assessment program.

3. Achievement tests

Achievement tests, also known as proficiency tests, are frequently used to measure an individual's current knowledge or skills that are important to a particular job or a specific course of study. They may be used to assess an individual's progress in a specific training or career

development program, or to assess the need for training. These tests generally fall into one of the following formats:

- **Knowledge tests** typically involve specific questions to determine how much the individual knows about particular job tasks and responsibilities. Traditionally they have been administered in a paper-and-pencil format, but computer administration is becoming more common. Knowledge tests tend to have relatively high validity.
- **Work-sample or performance tests** require the individual to actually demonstrate or perform one or more job tasks. These tests, by their makeup, generally show a high degree of job-relatedness. Test takers generally view these tests as fairer than other types of tests. Use of these tests often results in less adverse impact than mental ability tests and job knowledge tests. They may be used in counseling to help a client get a realistic job preview, as well as to assess their achievement in a specific area. However, they can be expensive to develop and administer.

Achievement tests can also be important components of work readiness models. These models/systems measure an individual's readiness to enter into employment based on assessment results. Persons who reach certain levels of achievement or who have mastered particular competencies receive skill proficiency certificates which convey their level of mastery in terms of the work readiness model.

4. Interest measures

Interest measures usually assess a client's career-related interests by having the client respond positively or negatively to various activity statements or descriptions. Counselors use these measures to help with career and academic planning, and they are often the most non-threatening way to begin the career exploration process. They can be used to make clients aware of career possibilities they had not previously considered, confirm tentative choices, or help clients focus their attention on occupational areas that relate to their interests.

Two approaches are most often utilized in interest inventories - general scales and occupational scales. General scales measure a client's interests in broad categories (e.g., selling or physical performing). Occupational scales assess how similar a client's pattern of likes and dislikes is to those of workers in various occupations.

General scales usually require less knowledge about the world of work than do occupational scales. They often are used to assess high school students or inexperienced workers. Occupational scales provide more occupationally-specific information than do general scales. Therefore, occupational scales often are more appropriate instruments for the assessment of vocationally-sophisticated clients. Many interest measures offer both scales in one instrument.

Interest measures differ in their response formats. Some ask for "like" or "dislike" responses; others require forced choices between two or more items, and others ask for responses on a scale of "like a lot" to "dislike a lot."

Interest inventories are attractive as counseling tools because they can be administered relatively quickly, and their results tend to make intuitive sense to clients. Interest inventories typically do a good job of predicting job satisfaction. They typically do not predict job or training performance well. Therefore, interest measures are frequently used in counseling, but are rarely used for selection or referral purposes.

It is important to be aware that interest patterns can be unstable for young or inexperienced workers. Counselors may want to consider a new administration of an interest measure after six months or so when working with such clients.

5. Work and personal values measures

Interest measures ask people to consider whether they *like* or *dislike* activities performed on a job, whereas work values instruments evaluate the relative *importance* of job activities and conditions to them. Interest inventories address the question of what activities people like to perform; work values instruments ask which conditions people consider most important and rewarding to them.

Work values instruments typically ask clients to rate the importance of such characteristics as job security, salary, or the opportunity to display creativity on the job. Work values instruments are not used to make evaluative judgments. Rather, they are used to determine the relative importance of various job characteristics to the client. When people are employed in positions that meet their needs, they are more likely to be satisfied. Work values have been shown to be valid predictors of job satisfaction.

Personal values instruments address broader value structures that relate to personal lives as well as to careers. The importance of family, religion, and physical activity are examples of values that may be measured by a personal values instrument. Exploring personal values can help clarify work goals in the context of more general life goals.

Since they are used to explore clients' values rather than their behavior or abilities, work values instruments often are perceived by clients as less threatening than other assessments. They can act as springboards to discussions of vocational characteristics and choices. They also can pave the way for further testing. Values instruments can be used in conjunction with interest, ability, and other measures as part of a whole-person assessment program to help clients make vocational decisions.

6. Personality inventories

Some employers and counselors use personality inventories to learn about an individual's personal, emotional, and social traits and behaviors. Personality inventories designed for use in employment contexts are used to evaluate such characteristics as motivation, conscientiousness, self-confidence, or how well an employee might get along with fellow workers.

Research has shown that personality tests, in conjunction with other assessment instruments, can yield helpful predictions for career guidance, and for some training and development programs. For example, if a client is considering a career in customer service, a personality inventory that includes a measure of introversion/extroversion can provide useful information about that individual's suitability for working with people. A personality inventory that assesses leadership potential might be used to select individuals for a career development program in a managerial track.

Personality tests fall into two general categories: vocational and clinical instruments.

- Vocationally-oriented instruments ask clients to check off or rate items that best describe themselves. These measures can be used to evaluate such characteristics as motivation, conscientiousness, self-confidence, or how well an individual might get along with fellow workers.
- Clinical instruments are designed to screen for psychopathology. They can be either general purpose measures that produce scores on a number of scales or they can be specific purpose measures that screen for narrower types of pathology such as depression or eating disorders.

If you decide to use a personality instrument, be sure that it is appropriate for your purposes, and that you are trained to administer and interpret it. Clinical instruments usually are not appropriate for career counseling or development because they have been devised to detect psychopathology, not to assess job-relevant characteristics. Clinical tests are used in very limited employment contexts, primarily for jobs in which it is critical for the employer to have some idea about the applicant's state of mind, such as in the selection of law enforcement officers or nuclear power plant operators.

If a personality inventory or other assessment tool provides information that would lead to identifying a mental disorder or impairment, the tool is considered a medical exam under the ADA. The ADA permits medical examinations of applicants and employees only in limited circumstances.

There are a few additional concerns about personality tests. Since there are usually no right or wrong answers to the test items, test takers may provide socially desirable answers. However, sophisticated personality inventories often have "lie-scales" built in, which allow such response patterns to be detected. There is also a general perception that some of these tests ask personal questions that are only indirectly relevant to job or training performance. This may raise concern on the part of test takers that such tests are an invasion of privacy. Some of these concerns can be reduced by including personality tests only as one part of a broader assessment program.

7. Interviews

The interview is probably the most commonly used assessment tool. The interview can range from being totally unplanned, that is, *unstructured*, to carefully designed beforehand, that is, completely *structured*. The most structured interviews have characteristics such as standardized questions, trained interviewers, specific question order, controlled length of time, and a

standardized response evaluation format. At the other end of the spectrum, a completely unstructured interview would probably be done “off the cuff,” with untrained interviewers, random questions, and with no consideration of time. A structured interview will be more reliable and valid. When interviews are used in selection for training or development programs, reliability and validity become more important.

Counselors use interviews to help gather information about clients and clarify results of other assessments. In counseling job-seekers, it is important to appraise their interviewing skills and help them work on obstacles such as situational shyness, anxiety, and inappropriate body language. Every question in a job interview is geared towards selecting applicants; clients may need help to understand this. Counselors may need to explain how responses to seemingly innocuous questions might help or harm clients’ chances for a job offer. Even sophisticated clients may need a session focused on pinpointing their problem areas and anticipating the kinds of questions that might be asked.

Regardless of the extent to which the interview is structured or unstructured, the skill of the interviewer can make a difference in the quality of the information gathered. A skillful, trained interviewer will be able to ask relevant follow-up questions to clarify and explore issues brought up during the interview.

In selection, it is unlawful to ask questions about medical conditions and disabilities before a conditional job offer. Even if the job applicant volunteers such information, employers are not permitted to pursue inquiries about the nature of the medical condition or disability. Instead, they should refocus the interview so that emphasis is on the ability of the applicant to perform the job, not on the disability. Following a structured interview format can help interviewers avoid unlawful or inappropriate inquiries where medical conditions, disability, and age are concerned.

It is important to note that inquiries about race, ethnicity, or age generally are not expressly prohibited under the law, but usually serve no credible purpose in an interview. These types of questions also are closely scrutinized by organizations, including regulatory agencies interested in protecting the civil rights of applicants. Counselors can prepare their clients by discussing types of questions that are illegal for an employer to ask during a job interview and strategies for responding to illegal or inappropriate questions.

8. Assessment centers

Assessment centers are most widely used for managerial and other high level positions to assess managerial potential, promotability, problem-solving skills, and decision-making skills. They also may be an integral part of career development programs, for selection or for monitoring individual progress. In the assessment center approach, candidates are generally assessed with a wide variety of instruments and procedures. These could include interviews, ability and personality measures, and a range of standardized management activities and problem-solving exercises. Typical of these activities and exercises are in-basket tests, leaderless group discussions, and role-play exercises.

- In-basket tests ask the candidates to sort through a manager's "in-basket" of letters, memos, directives, and reports describing problems and scenarios. Candidates are asked to examine them, prioritize them, and respond appropriately with memos, action plans, and problem-solving strategies. Trained assessors then evaluate the candidates' responses.
- Leaderless group discussions are group exercises in which a group of candidates is asked to respond to various kinds of problems and scenarios without a designated group leader. Candidates are evaluated on their behavior in the group discussions. This might include their teamwork skills, their interaction with others, or their leadership skills.
- In role-play exercises, candidates are asked to pretend that they already have the job and must interact with another employee to solve a problem. The other employee is usually a trained assessor. The exercise may involve providing a solution to a problem that the employee presents or suggesting some course of action regarding a hypothetical situation. Candidates are evaluated on the behavior displayed, solutions provided, or advice given.

Assessors must be appropriately trained. Their skills and experience are essential to the quality of the evaluations they provide.

Assessment centers apply the whole-person approach to personnel assessment. They can be very good predictors of job performance and behavior when the tests and procedures making up the assessment center are constructed and used appropriately.

It can be costly to set up an assessment center. Large companies may have their own assessment centers; mid-size and smaller firms sometimes send candidates to private consulting firms for evaluation.

9. Comprehensive career planning measures

Comprehensive career planning measures combine several types of assessment instruments, such as measures of abilities, interests, and values, into a single tool. In addition to traditional career counseling, these measures may be used in career development programs. Such measures generate a number of reports or scores, often including a report that synthesizes the findings of the component scales into a set of vocational recommendations. A number of comprehensive career planning measures produce computer-generated narrative reports. Some of the comprehensive measures are based on self-estimates of levels of abilities and other characteristics, while others are based on standardized tests or a combination of self-estimates and standardized tests.

- Measures based on self-estimates can help clients to direct their thinking on vocational issues, but these measures are not designed to predict success in specific occupations. If clients are unable or unwilling to see a counselor, these instruments can serve as an alternative to counseling. Self-estimates are generally not used in selection and placement contexts.
- Measures incorporating standardized tests typically use several standardized measures, such as ability tests, interest inventories, and work values measures to provide

information that clients can use to explore career possibilities. Standardized tests are typically validated in terms of their success in predicting job performance, job or career satisfaction, or tenure.

Often, career development measures can be accessed through career information delivery systems. There are many web-based systems, and many states have developed their own local systems. These systems coordinate a variety of assessment tools used in career development with other client services that provide information about career options and opportunities. More information can be obtained about career information delivery systems through the Association of Computer-Based Systems for Career Information (ACSCI) at www.acsci.org.

Table 4. Main Advantages and Disadvantages of Different Types of Assessment Instruments

Type of assessment instrument	Advantages	Disadvantages
Ability tests	Mental ability tests ▪ Are among the most useful predictors of performance across a wide variety of jobs. ▪ Are usually easy and inexpensive to administer.	▪ Use of ability tests can result in high levels of adverse impact. ▪ Physical ability tests can be costly to develop and administer.
Achievement/proficiency tests	▪ In general, job knowledge and work-sample tests have relatively high validity. ▪ Job knowledge tests are generally easy and inexpensive to administer. ▪ Work-sample tests usually result in less adverse impact than ability tests and written knowledge tests.	▪ Written job knowledge tests can result in adverse impact. ▪ Work-sample tests can be expensive to develop and administer.
Interest measures	▪ Can be administered quickly. ▪ Predict job satisfaction. ▪ Are non-threatening to clients undergoing career counseling.	▪ May be unstable or unsuitable for young/inexperienced workers.
Work and personal values measures	▪ Work values instruments can be valid predictors of job or career satisfaction. ▪ Values instruments are generally not perceived as threatening by clients.	▪ Values instruments generally do not predict job performance. ▪ Personal values instruments may not relate directly to job choice.

Type of assessment instrument	Advantages	Disadvantages
Personality inventories	▪ Usually do not result in adverse impact. ▪ Predictive validity evidence exists for some personality inventories in specific situations. ▪ May help to reduce adverse impact when used in conjunction with other tests and procedures. ▪ Easy and inexpensive to administer.	▪ Need to distinguish between clinical and employment-oriented personality inventories in terms of their purpose and use. ▪ Possibility of faking or providing socially desirable answers. ▪ Concern about invasion of privacy (use only as part of a broader assessment battery).
Interviews	▪ Structured interviews, based on job analyses, tend to be valid. ▪ May reduce adverse impact if used in conjunction with other tests.	▪ Unstructured interviews typically have poor validity. ▪ Skill of the interviewer is critical to the quality of interview (interviewer training can help).
Assessment centers	▪ Good predictors of job and training performance, managerial potential, and leadership ability. ▪ Apply the whole-person approach to personnel assessment.	▪ Can be expensive to develop and administer. ▪ Specialized training required for assessors; their skill is essential to the quality of assessment centers
Comprehensive career planning measures	▪ Combine several types of assessment instruments into one comprehensive tool. ▪ Reports often synthesize findings into sets of vocational recommendations. ▪ Can help clients direct their thinking on vocational issues.	▪ They are only as good as the component instruments. ▪ They may be based on self-estimates, standardized tests, or a combination of the two. It is important to determine whether test results are based on self-report measures or standardized tests before making vocational decisions.

CHAPTER 5 How to Select Tests: Standards for Evaluating Tests

Previous chapters described a number of test types and their use as assessment tools in workforce development programs. Technical and legal issues that have to be considered in using tests also were discussed. In this chapter, information and procedures for evaluating tests will be presented.

Chapter Highlights

1. Sources of information about tests
2. Standards for evaluating a test: information to consider to determine suitability of a test for your use
3. Checklist for evaluating a test

Principle of Assessment Discussed

Use assessment instruments for which <i>understandable and comprehensive documentation</i> is available.
--

1. Sources of information about tests

Many assessment instruments are available for use in workforce development programs. Sources that can help you determine the tests that are appropriate for your situation are described below.

- **Test manual.** A test manual should provide clear and complete information about how the test was developed; its recommended uses and possible misuses; and evidence of reliability, validity, and fairness. The manual also should contain full instructions for test administration, scoring, and interpretation. In summary, a test manual should provide sufficient administrative and technical information to allow you to make an informed judgment as to whether the test is suitable for your use. You can order specimen test sets and test manuals from most test publishers. Test publishers and distributors vary in the amount and quality of information they provide in test manuals. The quality and comprehensiveness of the manual often reflect the adequacy of the research base behind the test. Do not mistake catalogs or pamphlets provided by test publishers and distributors for test manuals. Catalogs and pamphlets are marketing tools aimed at selling products. To get a balanced picture of the test, it is important to consult independently published critical test reviews in addition to test manuals.
- **Mental Measurements Yearbook (MMY).** The MMY is a major source of information about assessment tools. It consists of a continuing series of volumes. Each volume contains reviews of tests that are new or significantly revised since the publication of the

previous volume. New volumes do not replace old ones; rather, they supplement them. The MMY series covers nearly all commercially available psychological, educational, and vocational tests published for use with English-speaking people. There is a detailed review of each test by an expert in the field. A brief description of the test covering areas such as purpose, scoring, prices, and publisher also is provided. The MMY is published by the Buros Institute of Mental Measurements. Their web address is www.unl.edu/buros. The Buros Institute also makes test reviews available through a computer database. This database is updated monthly via an on-line computer service. This service is administered by the Bibliographic Retrieval Services (BRS).

- **Tests in Print (TIP).** TIP is another Buros Institute publication. It is published every few years and lists virtually every test published in English that is available for purchase at that time. It includes the same basic information about a test that is included in the MMY, but it does **not** contain reviews. This publication is a good starting place for determining the tests that are currently available.
- **Test Critiques.** Published by Pro-Ed, Inc., this publication provides practical and straightforward test reviews. It consists of several volumes, published over a period of years. Each volume reviews a different selection of tests. The subject index at the back of the most recent volume directs the reader to the correct volume for each test review.
- **Finding Information About Psychological Tests.** This pamphlet, available from the American Psychological Association (APA), is a guide for locating and using both published and unpublished tests. It includes the references above, as well as additional sources of information about tests. This and other information about testing can be found on the web at www.apa.org/science/testing.html.
- **Assessment Resources.** The Association for Assessment in Counseling and Education (AACE), a division of the American Counseling Association (ACA), also provides information on assessments and testing. Information and links to a variety of resources are available at <http://aac.ncat.edu/resources.html>.
- **Professional consultants.** There are many testing experts who can help you evaluate and select tests for your intended use. They can help you design assessment programs that are useful and comply with relevant laws.

If you are considering hiring a consultant, it is important to evaluate his or her qualifications and experience beforehand. Professionals working in this field generally have a Ph.D. in industrial/organizational psychology, counseling psychology, or a related field. Look for an individual with hands-on experience in the areas in which you need assistance. Consultants may be found in psychology or business departments at universities and colleges. Others serve as full-time consultants, either working independently, or as members of consulting organizations. Typically, professional consultants will hold memberships in organizations such as the APA (www.apa.org), the Society for Industrial and Organizational Psychology (www.siop.org), the American Society for Training and Development (www.astd.org), the Society for Human Resource Management (www.shrm.org), the International Personnel Management Association - Assessment Council (www.ipmaac.org), or other related professional organizations.

Reference libraries should contain the publications discussed above as well as others that will provide information about personnel tests and procedures. The *Standards for Educational*

and *Psychological Testing* and the *Principles for the Validation and Use of Personnel Selection Procedures* also can help you evaluate a test in terms of its development and use. In addition, these publications indicate the kinds of information a good test manual should contain. Carefully evaluate the quality and the suitability of a test before deciding to use it. Avoid using tests for which only unclear or incomplete documentation is available and tests that you are unable to thoroughly evaluate. This is the next principle of assessment.

Principle of Assessment

Use assessment instruments for which <i>understandable and comprehensive documentation</i> is available.
--

2. Standards for evaluating a test **Information to consider to determine suitability of a test for your use**

The following basic descriptive and technical information should be evaluated before you select a test for your use. In order to evaluate a test, you should obtain a copy of the test and test manual. Consult independent reviews of the test for professional opinions on the technical adequacy of the test and the suitability of the test for your purposes. This information applies to evaluating assessments that are part of professionally developed training programs, as well.

General information

- **Test description.** As a starting point, obtain a full description of the test. You will need specific identifying information to order your specimen set and to look up independent reviews. The description of the test is the starting point for evaluating whether the test is suitable for your needs.
 - **Name of test.** Make sure you have the accurate name of the test. There are tests with similar names, and you will want to look up reviews of the correct instrument.
 - **Publication date.** What is the date of publication? Is it the latest version? If the test is old, it is possible that the test content and norms for scoring and interpretation have become outdated.
 - **Publisher.** Who is the test publisher? Sometimes test copyrights are transferred from one publisher to another. You may need to call the publisher for information or for determining the suitability of the test for your needs. Is the publisher cooperative in this regard? Does the publisher have staff available to assist you?
 - **Authors.** Who developed the test? Try to determine the background of the authors. Typically, test developers hold a doctorate in industrial/organizational psychology, counseling psychology, psychometrics, or a related field and are associated with professional organizations such as the APA. Another desirable qualification is proven expertise in test research and construction.
 - **Forms.** Is there more than one version of the test? Are they interchangeable? Are forms available for use with special groups, such as non-English speakers or persons with limited reading skills?

- **Format.** Is the test available in paper-and-pencil and/or computer format? Is it meant to be administered to one person at a time, or can it be administered in a group setting?
- **Administration time.** How long does it take to administer?
- **Costs.** What are the costs to administer and score the test? This may vary depending on the version used and whether scoring is by hand, computer, or by the test publisher.
- **Staff requirements.** What training and background do staff need to administer, score, and interpret the test? Do you have suitable staff available now or do you need to train and/or hire staff?

Purpose, nature, and applicability of the test

- **Test purpose.** What aspects of training performance or career suitability do you need to measure? What constructs does the test measure? Does the manual contain a coherent description of these constructs? Is there a match between what the developer says the test measures and what you intend to measure? The test you select for your assessment should relate directly to career exploration or to your specific training and development program.
- **Similarity of reference group to target group.** The test manual will describe the characteristics of the reference group that was used to develop the test. How similar are your test takers, the target group, to the reference group? Consider such factors as age, gender, racial and ethnic composition, education, occupation, and cultural background. Do any factors suggest that the test may not be appropriate for your group? In general, the closer your group matches the characteristics of the reference group, the more confidence you will have that the test will yield meaningful scores for your group.
- **Similarity of norm group to target group.** In some cases, the test manual will refer to a norm group. A norm group is the sample of the relevant population on whom the scoring procedures and score interpretation guidelines are based. In some cases, the norm group may be the same as the reference group. If your target group differs from the norm group in important ways, then the test cannot be meaningfully used in your situation. For further discussion of norm groups, see Chapter 7.

Technical information

- **Test reliability.** Examine the test manual to determine whether the test has an acceptable level of reliability before deciding to use it. See Chapter 3 for a discussion of how to interpret reliability information. A good test manual should provide detailed information on the types of reliabilities reported, how reliability studies were conducted, and the size and nature of the sample used to develop the reliability coefficients. Independent reviews also should be consulted.
- **Test validity.** Determine whether the test may be validly used in the way you intend. Check the validity coefficients in the relevant validity studies. Usually the higher the validity coefficient, the more useful the test will be in predicting job success. See Chapter 3 for a discussion of how to interpret validity information. A good test manual will contain clear and complete information on the valid uses of the test, including how validation studies were conducted, and the size and characteristics of the validation samples. Independent test reviews will let you know whether the sample size was

sufficient, whether statistical procedures were appropriate, and whether the test meets professional standards.

- **Test fairness.** Select tests developed to be as fair as possible to test takers of different racial, ethnic, gender, and age groups. Read the manual and independent reviews of the test to evaluate its fairness to these groups. To secure acceptance by all test takers, the test also should appear to be fair. The test items should not reflect racial, cultural, or gender stereotypes or overemphasize one culture over another. The rules for test administration and scoring should be clear and uniform. Does the manual indicate any modifications that are possible and may be needed to test individuals with disabilities?
- **Potential for adverse impact.** The manual and independent reviews should help you to evaluate whether the test you are considering has the potential for causing adverse impact. As discussed earlier, mental and physical ability tests have the potential for causing substantial adverse impact if used in selection for training and development programs. However, they can be an important part of your assessment program. If these tests are used in combination with other tests and procedures, you will be able to obtain a better picture of an individual's training or career potential and reduce the effect of average score differences between groups on one test.

Practical evaluation

- **Test tryout.** It often is useful to try the test in your own organizational setting by asking employees of your organization to take the test and by taking the test yourself. Do not compute test scores for these co-workers or employees unless you take steps to ensure that results are anonymous. By trying the test, you will gain a better appreciation of the administration procedures, including the suitability of the administration manual, test booklet, answer sheets and scoring procedures, the actual time needed, and the adequacy of the planned staffing arrangements. The reactions of your co-workers or clients to the test may give you additional insight into the effect the test will have on examinees.
- **Cost-effectiveness.** Are there less costly tests or assessment procedures that can help you achieve your assessment goals? If possible, weigh the potential gain in diagnostic power or in training performance against the cost of using the test. Some test publishers and test reviews include an expectancy chart or table that you can consult to predict the expected level of performance of an individual based on his or her test score. However, make sure your target group is comparable to the reference group on which the expectancy chart was developed.
- **Independent reviews.** Is the information provided by the test manual consistent with independent reviews of the test? If there is more than one review, do they agree or disagree with each other? Information from independent reviews will prove most useful in evaluating a test.
- **Overall practical evaluation.** This involves evaluating the overall suitability of the test for your specific circumstances. Does the test appear easy to use or is it unsettling? Does it appear fair and appropriate for your target groups? How clear are instructions for administration, scoring, and interpretation? Are special equipment or facilities needed? Is the staff qualified to administer the test and interpret results, or would extensive training be required?

3. Checklist for evaluating a test

It is helpful to have an organized method for choosing the right assessment tools for your workforce development needs. A checklist can help you in this process. Your checklist should summarize the kinds of information discussed above. For example, is the test valid for your intended purpose? Is it reliable and fair? Is it cost-effective? Is the instrument likely to be viewed as fair and valid by the test takers? Also consider the ease or difficulty of administration, scoring, and interpretation given available resources. A sample checklist that you may find useful appears on the following page. Completing a checklist for each test you are considering will assist you in comparing them more easily.

CHECKLIST FOR EVALUATING A TEST

Characteristic to be measured by test (skill, ability, personality trait):	
Job/training characteristic to be assessed:	
Target population (education, or experience level, other background):	
TEST CHARACTERISTICS	
Test name:	Version:
Type: (paper-and-pencil, computer)	Alternate forms available:
Scoring method: (hand-scored, machine-scored)	
Technical considerations:	
Reliability: $r =$	Validity: $r =$ Reference/norm group:
Test fairness evidence:	
Adverse impact evidence:	
Applicability (indicate any special group)	
Administration considerations:	Administration time:
Materials needed (include start-up costs, operational and scoring cost):	Costs:
Facilities needed:	
Staffing requirements:	
Training requirements:	
Other considerations (consider clarity, comprehensiveness, utility):	
Test manual:	
Supporting documents from the publisher:	
Publisher assistance:	
Independent reviews:	
Overall evaluation:	

CHAPTER 6 Administering Assessment Instruments

Proper administration of assessment instruments is essential to obtaining *valid* or *meaningful* scores for your test takers. This chapter discusses how to administer assessment instruments so that you can be certain that the results will be valid and fair.

Chapter Highlights

1. Training and qualifications of administration staff
2. Following instructions and guidelines stated in the test manual
3. Ensuring suitable and uniform assessment conditions
4. Test administrator feedback
5. The pre-test counseling session
6. How much help to offer test takers
7. Test anxiety
8. Alternative assessment methods for special cases
9. Providing reasonable accommodation in the assessment process to people with disabilities
10. Administering computer-based tests
11. Obtaining informed consent of test takers and a waiver of liability claims
12. Self assessment
13. Maintaining assessment instrument security
14. Maintaining confidentiality of assessment results
15. Testing unionized employees

Principles of Assessment Discussed

Ensure that administration staff are *properly trained*.
Ensure that *testing conditions are suitable* for all test takers.
Provide *reasonable accommodation* in the assessment process for people with disabilities.
Maintain assessment instrument *security*.
Maintain *confidentiality* of assessment results.

1. Training and qualifications of administration staff

The qualifications and training required for a test administrator will depend on the nature and complexity of the test. The more complex the test administration procedures, the more training

an administrator will need. However, even simple-to-administer tests need trained staff to ensure valid results. Administrators should be given ample time to learn their responsibilities before they administer a test to clients. Your staff may need professional training in test administration that is offered by some test publishers.

Only those staff who can administer the test in a professional and satisfactory manner should be assigned test administration duties. Test administrators should be well organized and observant, speak well, and be able to deal comfortably with people. They also should be trained to handle special situations with sensitivity. For example, they should know how to respond to a test taker's request for an accommodation and be able to calm down those who may become overly anxious about taking a test. This leads to our next principle of assessment.

Principle of Assessment

Ensure that administration staff are <i>properly trained</i> .
--

2. Following instructions and guidelines stated in the test manual

Staff should be thoroughly familiar with the testing procedures before administering the test. They should carefully follow **all** standardized administration and scoring procedures as outlined in the test manual. Test manuals will indicate the test materials that are needed, the order of presentation, and the instructions that must be read verbatim. They also will indicate whether there are time limits, and, if so, what those time limits are. Any special instructions noted by the test manual should be observed. This includes meeting the requirements for specific equipment or facilities. Alterations can invalidate results.

3. Ensuring suitable and uniform assessment conditions

There are various extraneous influences that may affect the reliability and validity of an assessment procedure. To maintain the integrity of results, you and your staff should make sure that adverse conditions are minimized.

- **Choose a suitable testing location.** Obtain a room that is well-lit and well-ventilated, with acceptable room temperature. Make sure the room is free of noise, traffic, and other interruptions. Chairs should be comfortable, and tables should be at an appropriate height, with sufficient room for test booklets and answer sheets. Furthermore, testing facilities and conditions must be uniform for all test takers. This means that people taking the test in another room or at a different time should be in a substantially similar testing environment. As indicated in Chapter 3, these extraneous factors can affect the reliability and validity of test results.

- **Prepare the room and test materials ahead of time.** Chairs and tables should be set up in position. Staff should check that all needed test materials and equipment are available and in good condition.
- **Test taker readiness or suitability for testing.** Be alert to problems individuals may have in taking the test. Before the assessment begins, give them an overview of the test and ask whether anyone anticipates having a problem taking the test. Some test takers may have forgotten to bring their eyeglasses; others may have bad colds or other temporary illnesses. These individuals should be rescheduled. Others may have disabilities that require accommodations or an alternate assessment arrangement (see section on ADA in Chapter 2).
- **Uniform administration.** The practices and precautions discussed above should become standard procedures in preparing testing materials, equipment, and facilities. Also, make sure that all test takers understand the directions before the test begins and are ready to follow the standard set of instructions during the test. These steps will help ensure that the results reflect real differences among individuals, and not differences in test administration. This brings us to the next principle of assessment.

Principle of Assessment

Ensure that *testing conditions are suitable* for all test takers.

To maintain the integrity of test results, administrators need to be alert to test takers' activities throughout the session. For example, some individuals may lose their place in the test booklet or put answers in the wrong column on the answer sheet. Others may try to copy answers from someone else. An alert administrator will be able to correct these situations quickly before they invalidate the test takers' responses.

4. Test administrator feedback

Frequently, the test administrator will observe client behavior in the test session that would be of interest to the counselor. For example, one test taker might fill in the circles of an answer sheet for a speeded test with meticulous neatness and no regard for the clock, while another might display signs of unusual anxiety during the session. Moreover, the test administrator might observe behavior that a client has not demonstrated during counseling sessions. A client who appears to be reserved and cooperative during counseling may act disruptively during the test session.

In training and development programs using tests, it is a good practice to establish a procedure that allows the test administrator to report any extenuating circumstances in the testing situation. For example, if some disruption occurred during the test administration, this is relevant information in later interpreting or applying test results. The test administrator also might want to report questions asked by examinees. The feedback could help to clarify test instructions used in the future.

5. The pre-test counseling session

To get the best effort and cooperation from clients in counseling settings, it is important for them to understand the goals of assessment and how test results can benefit them. When clients take tests given by employers as part of the selection process, motivation is strong. To get hired, they must do as well as possible on the assessment procedures. When clients take tests suggested by a counselor, motivation may be questionable. Counselors can take steps during the pre-test counseling session to improve client interest and motivation.

The client should be consulted when planning a testing strategy. The counselor will have made some decisions as to what assessment tools are most appropriate for a particular client. The counselor then should explain to the client the benefits or usefulness of each of the assessment instruments in question. Most clients at this point will agree to all of the counselor's recommendations. The counselor might work on eliciting client feelings about the assessments to get some input and generate interest. And certainly, if the client is opposed to some or all of the assessment tools discussed, his or her feelings must be taken into account in the decision-making process.

Once the plan is agreed upon, the counselor should make sure that all clients are aware of what to expect during the assessment sessions. If an orientation booklet is available from the test developer or publisher that describes the test and gives sample questions, counselors should arrange for clients to review the booklet. Some clients who have limited recent test-taking experience or who have been out of school for a long time may need additional preparation. They may not only need an orientation to the types of questions that will be asked, but also may need to be familiarized with machine-scored answer sheets or with taking tests on a computer. For example, some clients may get very low scores on speeded tests by spending excessive amounts of time making sure that the boxes are filled out neatly on the answer sheet. Counselors may want to encourage clients to work quickly on such tests and be less concerned with neatness. The counselor's preparation session should familiarize clients with test structure and format. However, the counselor should not disclose any actual test items nor any other information that would compromise the validity or security of the test.

The pre-test counseling session also should include a discussion of what the client should expect from the results of the assessment process. Counselors should explain the type of information that will be furnished based on the assessments and how that information will be useful to the client. If ability or achievement tests are being scheduled, the possibility of the client obtaining low scores exists. It is helpful to broach this subject in the pre-test counseling session. Clients should understand that they may not do well in every area, that this is true for everyone, and that the results still will provide very useful information. Preparing the client for this kind of information will make it easier to talk about high and low scores in the post-test interview. It also may lower the client's anxiety about "failing." These types of discussions also may increase the likelihood that the client will show up for the test session. When they report for the post-test session, they probably will come with a more realistic idea of what to expect from the test results.

If a client is referred to an employer who uses assessment instruments as part of the selection procedure, the counselor should prepare the client for taking these instruments. Preparation should include a discussion of the particular test's purposes and formats, if known. If pre-test exercises or booklets are available, they should be discussed with the client. It is not appropriate or ethical to give clients information about a test that would invalidate the results or give them an unfair advantage over other applicants.

6. How much help to offer test takers

The test manual usually indicates the kind of assistance and information that can be provided to test takers during the test. Administration staff should be familiar with what is and is not permissible at each stage of the assessment process.

Some instruments allow the administrator to clarify the directions and practice exercises, but prohibit help with the actual test questions. This is generally true for ability and achievement tests. However, other assessment tools, such as interest inventories or personality tests, may allow for more assistance with the assessment.

In general, test takers should not be coached on how best to answer test questions. Administrators should not offer more information than is indicated in the instructions. If they do, some individuals will be given an unfair advantage.

7. Test anxiety

Most people feel some anxiety about taking a test. For some otherwise qualified individuals, test anxiety can have a paralyzing effect on their performance. There are a few things that can be done to alleviate anxiety.

- Written orientation materials are available for many tests. These materials describe the test and provide sample questions. If such materials exist, they should be made available to all test takers well in advance of the test date.
- Before the test begins, give test takers a brief orientation explaining the purpose of the test, the type of questions to expect, and how long the test will last.
- Start test sessions promptly. A long wait will raise the anxiety level among test takers. All testing materials, equipment, and facilities should be ready well in advance of the scheduled session. A well-run test session helps to reduce test anxiety.

8. Alternative assessment methods for special cases

Some individuals, because of cultural differences, poor skills in English, or limited formal education, cannot be evaluated properly using some of the available assessment instruments. Poor test performance may not be a reflection of their job-related knowledge, skills, or abilities, but rather may be due to the existence of a cultural or language barrier. Some of these tests may be available in appropriate foreign language versions or in a version suitable for individuals functioning at low literacy levels. Also, where appropriate, work samples and structured interviews should be considered as practical alternatives to written tests. At times, individual evaluations by outside agencies or consultants may be a suitable approach.

9. Providing reasonable accommodation in the assessment process to people with disabilities

The Americans with Disabilities Act has opened up employment opportunities for a great number of qualified persons with disabilities. These opportunities have enabled persons with disabilities to apply their skills and be successful in the world of work. Under the ADA, you are required to provide reasonable accommodation in the assessment process to qualified persons with disabilities. This leads to our next principle of assessment.

Principle of Assessment

Provide <i>reasonable accommodation</i> in the assessment process for people with disabilities.

Accommodation in the assessment process may involve ensuring physical accessibility to the test site, modifying test equipment or tests, or providing qualified assistance. Giving extra time on certain kinds of tests to test takers with dyslexia or other learning disabilities and administering a larger print version of a test to a person who is visually impaired are examples of reasonable accommodation. However, providing a reader for a reading comprehension test could invalidate the test results. You should become familiar with the accommodations that can be made for different conditions or circumstances without invalidating the test. Provide all test takers with descriptive information about the test in advance, so that they will have ample opportunity to request needed accommodations. When the need for accommodation is not obvious, you may ask for reasonable documentation of the disability's functional limitations for which accommodation is needed. The test taker, test manual, the test publisher, and several professional associations (listed in Chapter 2 and Appendix A) can help you determine the appropriate reasonable accommodations for particular situations. If an accommodation cannot be made without invalidating the test, alternative assessment strategies, such as a review of past job experience, a review of school records, or a brief job tryout, must be considered.

10. Administering computer-based tests

Many tests are now computer-based. Computers can be used to administer and score tests and print results. A number of computerized tests also provide extensive test interpretations.

Some computer-based tests are adaptive. Adaptive tests, as opposed to conventional tests, present test questions based on the responses of the test taker to previous questions, and so adjust for his or her level of ability. This allows for a more reliable measure of ability with fewer items administered.

Advantages to computer-based testing include:

- Administration procedures are the same for all test takers.
- The need for test administrators is reduced.
- Results can be available immediately.
- The test can be administered without delay to walk-in applicants.

Disadvantages of computer-based testing include:

- A computer is needed for each test taker.
- Some test takers may feel uncomfortable using a computer; this could raise anxiety levels and adversely affect scores of these individuals.

11. Obtaining informed consent of test takers

When a test taker gives informed consent, it implies that he or she understands the nature of the test, the reasons for it, and how the results will be used. When tests are being given for counseling purposes, counselors or staff should make sure that clients understand why they are being tested and what to expect of the assessment process. In some agency settings, informed consent is not required, but the client still should be apprised of the purpose and nature of assessments. In training and development contexts, the purpose of the assessment should be made explicit in all cases. Obtaining informed consent may not be legally required in this context, but it may help protect an organization from liability. Obtaining written consent does not relieve the organization of legal liability if applicable laws are violated.

12. Self assessment

In some cases, test instruments may be self-administered, scored, and interpreted. These may include abilities tests, interest and values inventories, and even personality inventories. Self assessment may be used in counseling, training and development, and program evaluation. When the focus of the assessment is to provide information to an individual about his or her knowledge, skills, abilities, or other characteristics, self-assessment tools can be appropriate methods.

Self assessment also may be useful when a program is short on staff or other resources. They can help clients learn self-assessment skills that may help them with their career development now and in the future.

If you are considering using self assessment as part of your workforce development program, it is important to ensure that the measures are appropriate for such use and have self-interpretable reports. Counselors may assist clients in a self-assessment process, but clients should be able to conduct the assessments independently.

13. Maintaining assessment instrument security

For tests used to make selection or referral decisions, test takers should not have an opportunity to view the test beforehand. This will help you obtain fair and valid results for your clients. To ensure this, keep test materials secure at all times. Store all materials relating to the test in locked rooms or cabinets when not in use and account for all materials that are used during the testing session. In general, test takers should not take any items from the testing room. Limit access to testing materials to staff involved in the assessment process. This brings us to the next principle of assessment.

Principle of Assessment

<i>Maintain assessment instrument security.</i>
--

In counseling, some instruments may not be held to the same strict standards of security that others are. For example, interest and values measures might be suitable for a client to take at home at their leisure. Tests that are intended to be used in self assessments also do not require the strictest security. To the extent that an instrument is used as information for individuals about their knowledge, skills, abilities, or other characteristics, test security may be less critical. If a test is used to help select candidates for training or development programs, security is paramount.

Security measures also are required when you use computer-based tests. Establish a password procedure for accessing computerized test materials and secure all related computer disks and manuals. Many computerized test developers encode test items and answer keys so that items cannot easily be read if electronic files are stolen.

When tests are used over a long period of time, it becomes increasingly likely that some test questions will leak out. To help maintain security, test developers periodically introduce new alternate forms. If alternate forms of the test are available, you can increase security by varying the form used.

14. Maintaining confidentiality of assessment results

Test results and answer sheets should be kept in a secure location. Results should be released only to those who have a legitimate need to know. Test results are confidential and should not be disclosed to another individual or outside organization without the informed consent of the test taker. This is the next principle of assessment.

Principle of Assessment

Maintain <i>confidentiality</i> of assessment results.
--

15. Testing unionized employees

Testing may be a mandatory subject of collective bargaining between management and labor unions. Therefore, unionized employers should not institute a testing program or revise a current program without first referring to the collective bargaining agreement.

CHAPTER 7 Using, Scoring, and Interpreting Assessment Instruments

This chapter describes some of the most common assessment instrument scoring procedures. It discusses how to properly interpret results and how to use them effectively. Other issues regarding the proper use of assessment tools also are discussed.

Chapter Highlights

1. Assessment instrument scoring procedures
2. Test interpretation methods: norm and criterion-referenced tests
3. Interpreting test results
4. Using test results in counseling
5. Using test results in training and development programs
6. Minimizing adverse impact

Principle of Assessment Discussed

Ensure that scores are interpreted properly.
--

1. Assessment instrument scoring procedures

Test publishers may offer one or more ways to score the tests you purchase. Available options may range from hand scoring by your staff to machine scanning and scoring done by the publisher. All options have their advantages and disadvantages. When you select the tests for use, investigate the available scoring options. Your staff's time, turnaround time for test results, and cost may all play a part in your purchasing decision.

- ☐ **Hand scoring.** The answer sheet is scored by counting the number of correct responses or by categorizing responses, often with the aid of a stencil. These scores may then have to be converted from the *raw score* count to a form that is more meaningful, such as a *percentile* or *standard score*. Staff must be trained on proper hand scoring procedures and raw score conversion. This method is more prone to error than machine scoring. To improve accuracy, scoring should be double checked. Hand scoring a test may take more time and effort, but it also may be the least expensive method when there are only a small number of tests to score.
 - ☐ **Computer-based scoring.** Tests can be scored using a computer and test scoring software purchased from the test publisher. When the test is administered in a paper-and-pencil format, raw scores and identification information must be key-entered by staff following the completion of the test session. Converted scores and interpretive reports can be printed
-

immediately. When the test is administered on the computer, scores are most often generated automatically upon completion of the test; there is no need to key-enter raw scores or identifying information. This is one of the major advantages of computer-based testing.

- **Optical scanning.** Machine scorable answer sheets are now readily available for many multiple choice tests. They are quickly scanned and scored by an optical mark reader. You may be able to score these answer sheets in-house or send them to the test publisher for scoring.
 - **On-site.** You will need a personal computer system (computer, monitor, and printer), an optical reader, and special test scoring software from the publisher. Some scanning programs not only generate test scores, but also provide test users with individual or group interpretive reports. Scanning systems can be costly, and the staff must learn to operate the scanner and the computer program that does the test scoring and reporting. However, using a scanner is much more efficient than hand scoring or key-entering raw scores when testing volume is heavy.
 - **Mail-in and fax scoring.** In many cases the completed machine-scannable answer sheets can be mailed or faxed to the test publisher. The publisher scores the answer sheets and returns the scores and test reports to the employer. Test publishers generally charge a fee for each test scored and for each report generated. For mail-in service, there is a delay of several days between mailing answer sheets and receipt of the test results from the service. Overnight mail by private or public carrier will shorten the wait, but will add to the cost. Some publishers offer a scoring service by fax machine. This will considerably shorten the turn-around time, but greater care must be taken to protect the confidentiality of the results.

2. Test interpretation methods: norm and criterion-referenced tests

Assessment instruments are used to make inferences about people's knowledge, skills, abilities, interests, traits, and values. What does the test score mean? To help answer this question, consider what the test is designed to accomplish. Does the test compare one person's score to those obtained by others in an occupation, or does it measure the absolute level of skill an individual has obtained? These two methods are described below.

- **Norm-referenced test interpretation.** In norm-referenced test interpretation, the scores that the applicant receives are compared with the test performance of a particular reference group. In this case, the reference group is the norm group. The norm group generally consists of large representative samples of individuals from specific populations, such as high school students, clerical workers, or electricians. It is their average test performance and the distribution of their scores that set the standard and become the test norms of the group. Interest, personality, and values inventories are norm-referenced instruments.

The test manual usually will provide detailed descriptions of the norm groups and the test norms. To ensure valid scores and meaningful interpretation of norm-referenced tests, make sure that your target group is similar to the norm group. Compare educational levels;

occupational, language, and cultural backgrounds; and other demographic characteristics of the individuals making up the two groups to determine their similarity.

- **Criterion-referenced test interpretation.** In criterion-referenced tests, the test score indicates the amount of skill or knowledge the test taker possesses in a particular subject or content area. The test score is not used to indicate how well the person does compared to others; it relates solely to the test taker's degree of competence in the specific area assessed. Criterion-referenced assessment is used in selecting for, and monitoring progress in, training and development programs.

A particular test score generally is chosen as the minimum acceptable level of competence.

How is a level of competence chosen? The test publisher may develop a mechanism that converts test scores into proficiency standards, or the company may use its own experience to relate test scores to competence standards.

It is important to ensure that all inferences you make on the basis of test results are well-founded. Only use tests for which sufficient information is available to guide and support score interpretation. Read the test manual for instructions in properly interpreting the test results. This leads to the next principle of assessment.

Principle of Assessment

Ensure that scores are interpreted properly.
--

3. Interpreting test results

Test results usually are presented in terms of numerical scores, such as raw scores, standard scores, and percentile scores. In order to interpret test scores properly, you need to understand the scoring system used.

□ **Types of scores**

- **Raw scores.** These refer to the unadjusted scores on the test. Usually the raw score represents the number of items answered correctly, as in mental ability or achievement tests. Some types of assessment tools, such as work value inventories and personality inventories, have no "right" or "wrong" answers. In such cases, the raw score may represent the number of positive responses for a particular trait. Raw scores do not provide much useful information. Consider a test taker who gets 25 out of 50 questions correct on a math test. It's hard to know whether 25 is a good score or a poor score. When you compare the results to all the other individuals who took the same test, you may discover that this was the highest score on the test. In general, for norm-referenced tests, it is important to see where a particular score lies within the context of the scores of other people. Adjusting or converting raw scores into standard scores or percentiles will provide you with this kind of information. For criterion-referenced tests, it is important to see what a particular score indicates about proficiency or competence.
- **Standard scores.** Standard scores are converted raw scores. They indicate where a person's score lies in comparison to a reference group. For example, if the test manual

indicates that the average or mean score for the group on a test is 50, then an individual who gets a higher score is above average, and an individual who gets a lower score is below average. Standard scores are discussed in more detail below in the section on standard score distributions.

- **Percentile score.** A percentile score is another type of converted score. An individual's raw score is converted to a number indicating the percent of people in the norm group who scored below the test taker. For example, a score at the 70th percentile means that the individual's score is the same as or higher than the scores of 70% of those who took the test. The 50th percentile is known as the median and represents the middle score of the distribution.

□ **Score distribution**

- **Normal curve.** A great many human characteristics, such as height, weight, math ability, and typing skill, are distributed in the population at large in a typical pattern. This pattern of distribution is known as the *normal curve* and has a symmetrical bell-shaped appearance. The curve is illustrated in Figure 2. As you can see, a large number of individual cases cluster in the middle of the curve. The farther from the middle or average you go, the fewer the cases. In general, distributions of test scores follow the same normal curve pattern. Most individuals get scores in the middle range. As the extremes are approached, fewer and fewer cases exist, indicating that progressively fewer individuals get low scores (left of center) and high scores (right of center).

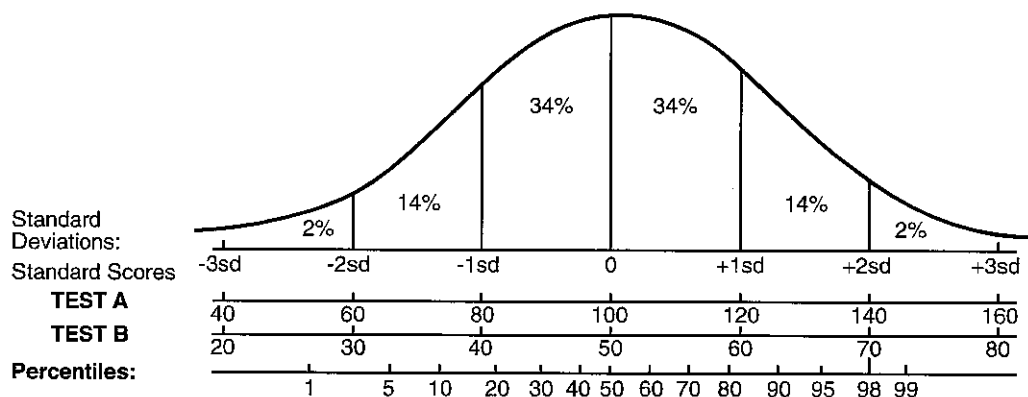


Figure 2. The normal curve illustrating standard score and percentile distribution.

- **Standard score distribution.** There are two characteristics of a standard score distribution that are reported in test manuals. One is the mean, a measure of central tendency; the other is the standard deviation, a measure of the variability of the distribution.
 - **Mean.** The most commonly used measure of central tendency is the mean or arithmetic average score. Test developers generally assign an arbitrary number to represent the mean standard score when they convert from raw scores to

standard scores. Look at Figure 2. Test A and Test B are two tests with different standard score means. Notice that Test A has a mean of 100 and Test B has a mean of 50. If an individual got a score of 50 on Test A, that person did very poorly. However, a score of 50 on Test B would be an average score.

- **Standard deviation.** The standard deviation is the most commonly used measure of variability. It is used to describe the distribution of scores around the mean. Figure 2 shows the percent of cases 1, 2, and 3 standard deviations (*sd*) above the mean and 1, 2, and 3 standard deviations below the mean. As you can see, 34% of the cases lie between the mean and +1 *sd*, and 34% of the cases lie between the mean and -1 *sd*. Thus, approximately 68% of the cases lie between -1 and +1 standard deviations. Notice that for Test A, the standard deviation is 20, and 68% of the test takers score between 80 and 120. For Test B the standard deviation is 10, and 68% of the test takers score between 40 and 60.
- **Percentile distribution.** The bottom horizontal line below the curve in Figure 2 is labeled "Percentiles." It represents the distribution of scores in percentile units. Notice that the median is in the same position as the mean on the normal curve. By knowing the percentile score of an individual, you already know how that individual compares with others in the group. An individual at the 98th percentile scored the same or better than 98% of the individuals in the group. This is equivalent to getting a standard score of 140 on Test A or 70 on Test B.

4. Using test results in counseling

The scoring and interpretation of test instruments have been addressed, but there are additional issues related to using results specifically in counseling settings. These relate to how results are given to examinees or clients and how they are reported within an agency.

- **Guidelines for the test interpretation session.** Before discussing results with clients, it is important for the counselor to review the results thoroughly and formulate a plan for the post-test discussion. Initially, the counselor may need to refer to the test manual to solidify his or her knowledge of the test itself. The test manual will discuss applications of the test, describe populations used to norm the instrument and suggest ways to interpret the results.

Understanding the strengths and limitations of the assessment instrument will enable the counselor to relate assessment results to the client's needs. The counselor should review the client's pattern of results in the context of other information about the client (e.g., results from other assessment tools, educational background, and vocational history).

The post-test session should aim to inform clients of the results, enlist clients' active participation in test interpretation, and help clients to formulate education, training, or career plans. It is advisable to get the client's reactions to the testing session before beginning a discussion of the client's results. This conversation might address how he or she felt about the assessment procedures and how the client thought he or she performed.

To open the discussion of test results, the counselor should review the purposes for which the tests were administered. It is usually better to discuss results in terms of specific goals, rather than as scores alone. Numerical scores often have very little meaning to clients. Clients should be provided with information about the implications of the scores. If specific scores are discussed, their meaning should be made clear in terms of career or occupational goals. If a battery of tests was given, it is helpful to remind the client what questions or parts of the test are associated with each score (e.g., "your clerical score comes from the part in which you had to decide whether the names were similar or different"). The counselor should explain what each of the scores is intended to indicate in terms the client can understand. For example, rather than telling a client he or she got a 70 on a work style scale, it would be more meaningful to explain that this indicates a preference for working with others rather than working alone.

The post-test session should leave clients with a picture of their skills, abilities, interests, or values as indicated by the assessment results. If the scores seem to be inconsistent with a client's vocational plans, the counselor should discuss the scores within the context of other available information about the person and reach some conclusions, if possible. A low score on a skills test may not indicate that the client should jettison his or her plan to pursue a job requiring those skills, but it might call for added steps to reach the goal. For example, the client may want to pursue more education or training to obtain the necessary skills. Your program can help the client obtain the necessary preparation to move forward toward the client's goals. In some cases, the client may want to consider changing career plans, but this should *never* be done on the basis of one test score alone. Results should be conveyed in terms of probabilities, rather than as definitive predictions. It is important to suggest options to clients rather than closing career doors. Test scores should be viewed together with other relevant factors such as education, work history, recreational activities, and other test scores, rather than in isolation.

If the test generates a profile, it is important to view the whole pattern as well as the individual scores. The profile can give you a more complete picture of the client. The individual scores may pinpoint particularly strong characteristics of the client, areas that need improvement, or areas that are not important to the client. Some instruments also provide scales that report the consistency of the client's profile; if such scales indicate that the client may be distorting the truth or failing to concentrate, it is advisable to probe the client's reaction to the test with particular care. It can be fruitful to look at clients' scores on individual scales in relation to their patterns of results as well as in comparison to the relevant population. For example, an interest scale score may reflect an average level of interest compared to the population, but it may represent a particular client's highest score. In the case of flat profiles, with no particular highs or lows, the counselor might find it useful to discuss the client's life experiences to elicit preferences. Sometimes, the flatness of the client's profile may have led the client to counseling in the first place.

- **Poor test results.** Counselors may shy away from suggesting ability or achievement tests for clients because the counselors fear poor results. It may be a wise decision not to test,

especially when the available assessment tools are not appropriate given the client's educational level.

Often, however, useful information about clients' abilities and skills can best be obtained from tests. In such cases, it can be a disservice to clients *not* to test them. One argument in favor of testing is that many clients, especially if they are reserved or inarticulate, may produce more positive test results than the counselor anticipates. In such cases, test scores may introduce many possible career directions to discuss with the client. Another argument for testing is that clients are best positioned for vocational choice if they show their strengths *and* weaknesses. They also can seek education and training to help improve their weaknesses and maybe even make them strengths.

When the client and counselor agree to an assessment plan, they are forming a contract of sorts. The client agrees to take the assessment instruments with sufficient concentration to produce valid results and to participate actively in the post-test session. In return, the counselor agrees to review the assessment results with the client and to work with the client in integrating the results with all other client information to help formulate a vocational plan. Both the counselor and client should recognize that low scores are possible. Low scores should not be given undue weight in the test interpretation session, but they also should not be overlooked or downplayed. If recognized as a possibility before the assessment takes place, low scores can be treated as useful information, not as barriers to discussion.

Conveying low scores can be uncomfortable for the counselor, but it is an important part of providing feedback to a client. The counselor's role is to discuss the client's scores with objectivity and empathy, without being judgmental, moralistic, or showing pity. It is the counselor's responsibility to help clients understand their strengths and weaknesses. If a score falls below the counselor's and/or the client's expectations, the counselor could explore reasons for the discrepancy (e.g., poor mathematics scores might indicate the client has "rusty" math skills), but should not necessarily downplay the validity of the score. If the area of weakness impinges on the client's vocational plans, those plans could either be rethought or the counselor might discuss ways to compensate for the weakness, such as remediation. Most importantly, the test score should not be regarded as the last word on a client's aptitude or motivation; the whole-person approach asserts that other test scores, educational and vocational history, interests, and life experiences also must be considered.

- **Conducting group test interpretation.** Group sessions allow counselors to use their time more efficiently, while giving clients the opportunity to interact with peers. A well-run group session can help the client to stay focused on the task. It also can generate questions that might be of interest to clients who might not have thought to raise those questions in individual sessions.

The major function of the group test interpretation session is to convey information about the test results to a number of clients at one time. Self assessments are ideal for group interpretations since results are written for the client. The counselor presents information about the test, explains what the scores mean, and shows how to use the results in career

planning. Clients are given only their own scores; they do not see others' test scores. In a well-run group session, the counselor does not need to speak to each member individually to interpret his or her scores; the general presentation permits each client to draw conclusions from his or her own test results. The counselor suggests to clients areas on which to focus, such as highest aptitudes and interests. Often counselors have resource books available at the session that clients can use to assist them in their planning.

When forming a group, it is reasonable to select clients who have some similarity in educational or vocational background. Because test interpretation groups are often formed before test results are known, test score similarity usually is not a major consideration in group membership. However, after tests are scored, counselors should review the results to see whether there are clients who have unusually low scores or scores that are highly incompatible with the clients' histories. These clients might be scheduled for individual test interpretation.

- **Counselor assessment instrument write-ups.** The form and content of a write-up are often governed by agency policy. As such, write-ups of test scores may vary as widely as the goals of the agency and the counseling program. Generally, the write-up is used to summarize the client's situation for the counselor and for other professionals who might inherit the case in the future or might be consulted. The write-up may be used to document the client's progress for the purpose of agency review. Frequently, the act of writing helps to clarify the counselor's thinking regarding the case. Generally, the counseling write-up contains a brief description of the client, an overview of the present situation that brought the individual to counseling, a description of the evaluation procedure, observations of the client's behavior in counseling, a review of the test scores, client's views, counselor conclusions, recommendations, and a summary. The write-up should describe how test results have been integrated with all other information about the client.

The principle of confidentiality and its converse, freedom of information, are important considerations. Write-ups of test scores should be written as if they might be seen by all who are entitled to access. In many agency settings, counseling write-ups can function as legal documents; accuracy is paramount in such cases and comprehensibility to laypersons can be important. As with any confidential information, only those staff members who need to see test write-ups should have access to them. Freedom of information laws assert clients' rights to see their records; keep this in mind when characterizing clients and their behavior.

5. Using test results in training and development programs

In training and development programs, tests are used in several ways: to help clients identify their training needs, for selection into training and development programs, for monitoring individual progress through programs, and for evaluating the effectiveness of programs.

- **Identify training needs.** Test results can be used to identify training needs for clients or for employees. Usually, some standard of performance on a test has been established to be associated with adequate job performance and scores falling below that level indicate the need for training. For example, we may have information that successful auto mechanics score above the 50th percentile on a mechanical abilities test. If a client scores below this, it may indicate the need for more training.

- **Selection into training and development programs.** The rank-ordering of test results, the use of cut-off scores, or some combination of these approaches is commonly used to assess the qualifications of people and to make selection decisions about them. Many programs may use a variety of tests and procedures in their assessment of candidates for training and development programs. In general, you can use a “multiple hurdles” approach or a “total assessment” approach, or a combination of the two. These are described below.
 - **Rank-ordering** is a process of arranging candidates on a list from highest score to lowest score based on their test results. In rank-order selection, candidates are chosen on a top-down basis.
 - **A cut-off score** is the minimum score that a candidate must have to qualify for entry into a training or development program. Generally the cut-off score is set at a level which is determined directly related to training success. Candidates who score below this cut-off generally are not considered for training.
 - **Multiple hurdles approach.** In this approach, candidates must pass each test or procedure (usually by scoring above a cut-off score) to continue within the assessment process. It may be used to reduce the total cost of assessment by administering less costly screening devices to everyone, but having only those who do well take the more expensive tests or other assessment tools.
 - **Total assessment approach.** In this approach, test takers are administered every test and procedure in the assessment program. The information gathered is used in a flexible or *counterbalanced* manner. This allows a high score on one test to be counterbalanced with a low score on another. For example, an applicant who performs poorly on a written test, but shows great enthusiasm for learning and is a very hard worker, may still be a candidate for training or development.

A key decision in using the total assessment approach is determining the relative weights to assign to each assessment instrument in the program.

Figure 3 is a simple example of combining assessment results from several tests and procedures to generate a weighted composite score.

Assessment instrument	Assessment score (0-100)	Assigned weight	Weighted total
Interview	80	8	640
Mechanical ability test	60	10	600
H.S. course work	90	5	450
			Total Score: 1,690

Figure 3. Score-sheet for selection into machinist training: Candidate A.

A company is selecting employee candidates to train for entry-level machinist. The assessment instruments consist of a structured interview, a mechanical ability test, and high school course work. After consultation with relevant staff and experts, a weight of 8 is assigned for the interview, 10 for the test, and 5 for course work. A sample score sheet for one candidate, Candidate A, is shown above. As you can see, although Candidate A scored lowest on the mechanical ability test, the weights of all of the assessment instruments as a composite allowed him/her to continue on as a candidate for the machinist training rather than being eliminated for consideration as a result of the one low score.

- **Monitoring individual progress through training/development.** Tests are an important part of monitoring an individual's progress in training and development programs. Individuals can benefit from specific feedback about their test results. It can help them to identify strengths and weaknesses and help point to the need for further training, or establish that they have achieved mastery of relevant material. Test results should be maintained confidentially in appropriate personnel files, and only staff who have a need to know the information, and a right to it, should be allowed access.
- **Evaluating program effectiveness.** Tests can be used to monitor the overall effectiveness of training programs, usually by pre-testing and post-testing trainees to establish their change in knowledge, skill, or ability. This may even be accomplished using some of the same instruments that are used to monitor individual progress.

6. Minimizing adverse impact

A well-designed assessment program will improve clients' ability to make effective career, training, and development related decisions. However, some predictors of job performance may exhibit adverse impact. As a test user, there are several good testing practices to follow to minimize the likelihood of adverse impact in conducting assessments for selection and referral purposes for training.

- Be clear about what needs to be measured and for what purpose. Use only assessment tools that are job-related and valid, and only use them in the way they were designed to be used.
- Use assessment tools that are appropriate for the target population.
- Do not use assessment tools that are biased or unfair to any group of people.
- Consider whether there are alternative assessment methods to use.
- Consider whether there is another way to use the test that either is free of or reduces adverse impact.
- To minimize the possibility of adverse impact, it is recommended that a test be used as only *one part* of a comprehensive assessment process. That is, apply the whole-person approach to your personnel assessment program. This approach will allow you to improve your assessment of the individual and reduce the effect of differences in average scores between groups on a single test.

CHAPTER 8 Issues and Concerns With Assessment

It is important to remember that an assessment instrument, like any tool, is most effective when used properly and can be very counterproductive when used inappropriately. In previous chapters, you have read about the advantages of using tests and procedures as part of your workforce development program. You also have read about the limitations of tests in providing a consistently accurate and complete picture of an individual's career-related qualifications and potential. This chapter highlights some important issues and concerns surrounding these limitations. Careful attention to these issues and concerns will help you produce a fair and effective assessment program.

Chapter Highlights

1. Deciding whether or not to use tests
2. Test anxiety
3. Fallibility of test scores
4. Appeals process and retesting
5. Qualifications of assessment staff
6. Misuse or overuse of tests
7. Ethnic, linguistic, and cultural differences and biases
8. Testing people with disabilities

1. Deciding whether or not to use tests

How successful is your current assessment program? Is it in need of improvement? The decision to use a test is an important one. You need to carefully consider several technical, administrative, and practical matters.

In counseling settings, consider how much additional time and effort will be involved in expanding your assessment program. As in every business decision, you will want to determine whether the potential benefits outweigh the expenditure of time and effort. Be sure to factor in all the costs, such as purchase of tests and staff time, and balance these against all the benefits, including potential increases in the effectiveness of the assessment process.

Before expanding your assessment program or deciding to use tests in training and development programs, it is important to have a clear picture of your organization's needs, the benefits you can expect, and the costs you will incur.

2. Test anxiety

Many people are intimidated at the mere thought of taking a test. Some may fear that testing will expose their weaknesses, and some may fear that tests will not measure what they really can do on the job. Also, some people may view certain tests as an invasion of privacy. This is especially true of personality tests.

Fear or mistrust of tests can lower the scores of some otherwise qualified candidates or clients. To reduce these feelings, it is important to take the time to explain a few things about the testing program before administering a test. Any explanation should, at a minimum, cover the following topics:

- why the test is being administered,
- confidentiality of test results, and
- how the test results will be used in the assessment process.

3. Fallibility of test scores

All assessment tools and procedures are subject to measurement errors. This means that a test neither measures a characteristic with perfect accuracy for all people nor fully accounts for their knowledge, skills, abilities, or other characteristics. Thus, there will always be some errors in selection or referral decisions that are made based on assessment results. This is true of all assessment procedures, regardless of how objective or standardized they might be.

It is, therefore, important not to rely entirely on any one assessment instrument in making counseling or training decisions. Using a variety of assessment tools will help you obtain a fuller and more accurate picture of an individual. Consider such information as an evaluation of a person's education, work experience, and other job-relevant factors in addition to standardized test results.

4. Appeals process and retesting

Every test taker should have a fair chance to demonstrate his or her best performance on an assessment procedure. However, at times this might not occur. If the results may not be valid for an individual, consider retesting or using alternative assessment procedures before screening the individual.

There are external circumstances or conditions that could invalidate the test results. These may include the test taker's state of mind or health at the time of the test, the conditions under which the test is given, and his or her familiarity with particular questions on the test. To give some specific examples, a person who has a child at home with the measles may not be able to concentrate on taking a vocabulary test. Someone sitting next to a noisy air conditioner may not

be able to concentrate on the test questions. On another day, under different circumstances, these individuals might obtain a different score.

If you believe that the test was not valid for an individual, you should consider a retest. If other versions of the test are not available, consider alternative means of assessment. Check the test manual for advice from the publisher regarding retesting. It is advisable to develop a policy on handling complaints regarding testing and appeals for retesting, so that these concerns can be resolved fairly and consistently.

5. Qualifications of assessment staff

Test results may not be accurate if the tests have not been administered and scored properly, or if the results are not interpreted appropriately. The usefulness of test results depends on proper administration, scoring, and interpretation. Qualified individuals must be chosen to administer and score tests and interpret test results. These individuals must be trained appropriately. Test manuals usually will specify the qualifications and training needed to administer and score the tests and interpret results.

6. Misuse or overuse of tests

A single test cannot be expected to be valid in all situations and for all groups of people. A test generally is developed to measure specific characteristics or to predict specific performance criteria for a particular group. For example, a test with items designed to assess the interests of adults may not be valid for identifying interests of junior high school students.

In addition, test results usually provide specific information that is valid for a specific amount of time. Therefore, it is unlikely to be appropriate to consider a client for entry into a career development program based on his or her test scores on a proficiency test taken 5 years earlier.

The test manual and independent reviews of the test remain your best guides for administering, scoring, and interpreting the test.

7. Ethnic, linguistic, and cultural differences and biases

The American workforce is made up of a diverse array of ethnic and cultural groups, including many persons for whom English is not the primary language. Some of these individuals may experience difficulty on standardized tests due to cultural differences or lack of mastery of the English language. Depending on the nature of the job for which they are applying, this could mean that their test scores will not accurately predict their true job potential.

Before selecting new tests, consider the composition of your potential client or candidate population. Are the tests appropriate for all of them? The test manuals may provide assistance in determining this. If you need further clarification, contact the test publisher.

There may be cases where appropriate standardized tests are not available for certain groups. You may have to rely on other assessment techniques, such as interviews and evaluations of education and work experience to make your assessment decisions.

8. Testing people with disabilities

The Americans with Disabilities Act protects qualified individuals with disabilities from discrimination in all aspects of employment and in the job-seeking process, including counseling and testing. Your staff should be trained to evaluate requests for reasonable accommodation and provide these accommodations if they are necessary and would not cause “undue hardship.” These situations must be handled with professionalism and sensitivity. Properly handled, this can be accomplished without compromising the integrity of the assessment process.

Accommodation may involve ensuring physical accessibility to the test site, modifying test equipment or tests, or providing other forms of assistance. Giving extra time for certain kinds of tests to test takers with dyslexia or other learning disabilities and administering a Braille version of a test for the blind may be examples of reasonable accommodation. See Chapters 2 and 6 for further discussions on testing people with disabilities.

CHAPTER 9 A Review: Principles of Assessment

Assessment tools, when used appropriately, can improve the quality of an agency's services by helping clients, counselors, and job placement specialists gather valuable information that can be used to help clients make appropriate career decisions. Employers can effectively use assessment instruments to measure job-relevant skills and capabilities of employees in training and development programs. To use assessment tools properly, administrators and staff must be aware of the inherent limitations of any assessment procedure, as well as the legal and professional issues involved when using assessment tools.

The Guide is organized around 13 important assessment principles and their applications. This final chapter brings all the principles together. They are listed below in the order of their appearance in the text, with the chapter number in parentheses. Together, the 13 principles provide a comprehensive framework for conducting an effective assessment program.

☐ **Use assessment tools in a *purposeful manner* (Chapter 1)**

Assessment instruments, like other tools, are helpful when used properly, but can be useless, harmful, or illegal when used inappropriately. Often, inappropriate use results from not having a clear understanding of what you want to measure and why you want to measure it. You should be clear about what you want to accomplish with your assessment program in order to select the proper tools to achieve those goals.

In selecting your assessment tools, always keep in mind the knowledge, skills, abilities, characteristics, and personal traits you want to measure, and the population to be assessed. Once you are clear about your purpose, you will be better able to select appropriate assessment tools and use those tools in an effective manner. Only use tests that are appropriate for your particular purpose.

☐ **Use the *whole-person approach* to assessment (Chapter 1)**

An assessment instrument may provide you with important career- or training-relevant information about an individual. However, no assessment tool is 100% reliable or valid; all are subject to errors, both in measuring job-relevant characteristics and in predicting job performance. Moreover, a single assessment instrument only provides you with a limited view of a person's qualifications. Using a variety of tools to measure skills, abilities, and other job-relevant characteristics provides you with a solid basis upon which to make important career and development decisions and minimizes adverse impact.

☐ **Use only assessment instruments that are *unbiased* and *fair* to all groups (Chapter 2)**

Using unbiased and fair tests will enable you to provide the most appropriate assessment services to clients. You should review the fairness evidence associated with assessment instruments before selecting tools by examining the test manual and independent test reviews.

☐ **Use only *reliable* assessment instruments and procedures (Chapter 3)**

If a person takes the same test again, will he or she get a similar score, or a very different score? A reliable instrument will provide accurate and consistent scores. To meaningfully interpret test scores and make useful career or employment-related decisions, use only reliable tools. Test manuals usually will provide a statistic known as the reliability coefficient, giving you an indication of a test's reliability. The higher the reliability coefficient, the more confidence you can have that the score is accurate.

☐ **Use only assessment procedures and instruments that have been demonstrated to be *valid for the specific purpose* for which they are being used (Chapter 3)**

Validity is the most important issue in selecting assessment tools. It refers to (1) the characteristic the assessment instrument measures, and (2) how well the instrument measures the characteristic. Validity is not a property of the assessment instrument itself; it relates to how the instrument is being used.

A test's validity is established in reference to a specific purpose; it may not be valid for different purposes. For example, a test that may be valid for assessing someone's job knowledge may not be valid for predicting his or her leadership skills. You must be sure that the instrument is valid for the purpose for which it is to be used. Selecting a commercially developed instrument does not relieve you of this responsibility.

The test manual usually provides a statistic, the validity coefficient, which will give an indication of the test's validity for a specific purpose under specific circumstances. It measures the degree of relationship between test performance and job performance (i.e., job-relatedness of the test).

☐ **Use assessment tools that are appropriate for the *target population* (Chapter 3)**

An assessment tool usually is developed for use with a specific group; it may not be valid for other groups. For example, a test designed to predict the work values of adults may not be valid for use with junior high school students. It is possible that many junior high school students are not vocationally mature enough for the assessment tool to be meaningful, or perhaps the reading level of the test is not suitable for junior high school students. Tests should be appropriate for the individuals you want to test, that is, your target population.

The manual should indicate the group or groups the test is designed to assess. Your target population should be similar to the group on which the test was developed or normed. In determining the appropriateness of an instrument for your target group, also consider such factors as reading levels, cultural backgrounds, and language barriers.

☐ **Use assessment instruments for which *understandable and comprehensive documentation* is available (Chapter 5)**

Are the instructions for administration and interpretation understandable? Is the information sufficiently comprehensive to evaluate the suitability of the instrument for your needs? Carefully evaluate the documentation provided by the test publisher to be sure that the tools you select do the job you want them to do and furnish you with the information you need. If the documentation is not understandable or complete, you run the risk of selecting inappropriate instruments.

Test manuals should provide information about both the development and psychometric characteristics of tests. They should cover topics such as procedures for administration, scoring and interpretation, the recommended uses of an instrument, the groups for whom the test is appropriate, and test norms. They also should include a description of the validation procedures used and evidence of validity, reliability, and test fairness.

☐ **Ensure that administration staff are *properly trained* (Chapter 6)**

Assessment instruments must be administered properly to obtain valid results. Consult the test publisher and administration manual for guidelines on the qualifications and training required for test administrators. These requirements will vary depending on the nature and complexity of the test. Only suitable staff should be selected. Administrators should be given ample time to learn their responsibilities and should practice by administering tests to other staff before administering tests to applicants. Some test publishers may run training sessions for test administration and interpretation.

Administration staff also should be trained to handle special situations with sensitivity. An example would be responding to a request for accommodation based on a disability.

☐ **Ensure that *testing conditions* are suitable for all test takers (Chapter 6)**

There are various extraneous influences that may affect the reliability and validity of an assessment procedure. For example, noise in the testing room, poor lighting, inaccurate timing, and damaged test equipment may adversely affect test takers. Staff should ensure that the testing environment is suitable and that administration procedures are uniform for all test takers.

☐ **Provide *reasonable accommodation* in the assessment process for people with disabilities (Chapter 6)**

To ensure that qualified individuals with disabilities have an equal chance to demonstrate their potential, accommodations in the assessment process may be necessary. Under the ADA, reasonable accommodation may involve ensuring physical accessibility to the test site, modifying test equipment or the testing process, or providing qualified assistance to the test taker. For example, administering a Braille version of a test, allowing extra time to complete the test, or supplying a reader may be appropriate. It is important to become familiar with the types of accommodations that can be made without invalidating test results. If reasonable accommodation involving test administration cannot be made, consider alternative assessment strategies.

☐ **Maintain assessment instrument *security* (Chapter 6)**

All materials used in the assessment process, whether paper-and-pencil or computer-based, must be kept secure. Lack of security may result in some test takers having access to test questions beforehand, thus invalidating their scores. To prevent this, test administrators should, for example, keep testing materials in locked rooms or cabinets and limit access to those materials to staff involved in the assessment process.

Some instruments used in counseling and in self-assessments are not held to the same strict standards of security. The purpose of testing often will dictate the importance of security for test materials. Security is also the responsibility of test developers. The security of a test may become compromised over time. To protect security, test developers periodically introduce new forms of tests.

☐ **Maintain *confidentiality* of assessment results (Chapter 6)**

Assessment results are highly personal. Workforce development professionals must respect the test taker's right to confidentiality. Assessment results should be shared only with those who have a legitimate need to know. This would include staff involved in interpreting assessment results or making work-related decisions. Personal information should not be released to other organizations or individuals without the informed consent of the test taker.

☐ **Ensure that scores are *interpreted properly* (Chapter 7)**

Tests are used to make inferences about people's knowledge, skills, abilities, traits, interests, and values. The inferences should be reasonable, well-founded, and not based upon stereotypes. If test scores are not interpreted properly, the conclusions drawn from them are likely to be invalid, thus leading to poor decision making. Ensure that there is solid evidence to justify your test score interpretations and the decisions you make based on those scores. The test manual should provide instructions for properly interpreting test results.

APPENDIX A: Sources of Additional Information on Assessment in Workforce Development

The following list of reference materials provides sources of information on specific topics and issues relating to testing and assessment. The main text has referred to many of the publications listed below. Others are included as general reference documents and as recommended readings.

American Educational Research Association, American Psychological Association and National Council on Measurement in Education. (1999). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association. (Or current version.)

American Society for Training and Development. (1996). *Training and Development Handbook: A Guide to Human Resource Development* (4th ed.). New York: McGraw-Hill.

Anastasi, A., & Urbina, S. (1997). *Psychological Testing* (7th edition). New York: Macmillan.

Arvey, R. D., & Faley, R. H. (1988). *Fairness in Selecting Employees*. Reading, MA: Addison-Wesley.

Association for Assessment in Counseling. (2003). *Responsibilities of Users of Standardized Tests (RUST)* (3rd edition). Alexandria, VA: Author.

Association for Assessment in Counseling. (2003). *Standards for Multicultural Assessment* (2nd ed.). Alexandria, VA: Author.

Boudreau, J. (1996). *Cumulative Supplement to Employment Testing Manual*. Boston: Warren, Gorham & Lamont.

Bruyère, S.M., & O'Keeffe, J. (Eds.). (1994). *Implications of the Americans with Disabilities Act for Psychology*. Washington, DC: American Psychological Association.

Bureau of National Affairs. (1990). *The Americans with Disabilities Act: A Practical and Legal Guide to Impact, Enforcement, and Compliance*. Washington, DC: Author.

Bureau of National Affairs Policy and Practice Series. (1992-1995). *Fair Employment Practices Manual #8*. Washington, DC: Author.

Buros Institute of Mental Measurements. Various. *Mental Measurements Yearbook*. Lincoln, NE: University of Nebraska Press.

Buros Institute of Mental Measurements. Various. *Tests in Print*. Lincoln, NE: University of Nebraska Press.

- Camara, W. J. & Brown, D. C. (1995). Educational and employment testing: Changing concepts in measurement and policy. *Educational Measurement: Issues and Practice*, 14 (1), 5-11.
- Cronbach, L. J. (1988). Five perspectives on the validity argument. In H. Wainer & H.I. Braun (Eds.), *Test Validity* (pp. 3-17). Hillsdale, NJ: Erlbaum.
- Douglas, J. A., Feld, D. E. & Asquith N. (1989). *Employment Testing Manual*. Boston, MA: Warren, Gorham & Lamont.
- Equal Employment Opportunity Commission. (1978). The Office of Personnel Management, U.S. Department of Justice and U.S. Department of Labor (1979). *Questions and Answers Clarifying and Interpreting the Uniform Guidelines on Employee Selection Procedures*. 29 CFR Part 1607 (1988).
- Equal Employment Opportunity Commission. (1978). The Office of Personnel Management, U.S. Department of Justice and U.S. Department of Labor (1979). *Uniform Guidelines on Employee Selection Procedures*. 41 CFR Part 603 (1978).
- Equal Employment Opportunity Commission. (1992). *A Technical Assistance Manual on the Employment Provisions (Title I) of the Americans with Disabilities Act*. Washington, DC: U.S. Government Printing Office.
- Equal Employment Opportunity Commission. (1992). EEOC Technical Assistance Manual on Employment Provisions of the Americans with Disabilities Act; *ADA Enforcement Guidance: Preemployment Disability Related Questions and Medical Examinations*.
- Equal Employment Opportunity Commission & U.S. Department of Justice. (1991). *Americans with Disabilities Act Handbook*. Washington, DC: Author.
- French, W. L. (2003). *Human resources management* (5th edition). Houghton Mifflin Co.: Boston, MA.
- Guion, R. M. (1997). *Assessment, Measurement, and Prediction for Personnel Decisions*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Guion, R. M. (1980). On trinitarian doctrines of validity. *Professional Psychology*, 11, 385-398.
- Hogan J., & Hogan, R. (Eds.) (1984-1990). *Business and Industry Testing: Current Practices and Test Reviews*. Austin, TX: PRO-ED.
- Joint Committee on Standards for Educational Evaluation. (1994). *The program evaluation standards*. Thousand Oaks, CA: Sage
- Joint Committee on Testing Practices. (2004). *Code of Fair Testing Practices in Education*. Washington, DC: Author.

- Keyser, D. J., & Sweetland, R. C. (Eds.). (1984-2005). *Test Critiques*. (Vols. I-XI). Austin, TX: PRO-ED.
- Messick, S. (1989). Validity. In R.L. Linn (Ed.) *Educational Measurement* (3rd ed., pp. 13-103). New York: American Council for Education, Macmillan.
- Murphy, K. R., & Davidshofer, C. O. (1988). *Psychological Testing: Principles and Applications*. Englewood Cliffs, NJ: Prentice Hall.
- Nunnally, J. C. (1978). *Psychometric Theory* (2nd ed.). New York: McGraw-Hill.
- Prediger, D. J. (1994). Multicultural assessment standards: A compilation for counselors. *Measurement and Evaluation in Counseling and Development*, 27, 68-73.
- Society for Industrial and Organizational Psychology, Inc. (2003). *Principles for the Validation and Use of Personnel Selection Procedures* (4th ed.). Bowling Green, OH: Author.
- U.S. Department of Justice. (1991). *The Americans with Disabilities Act: Questions and Answers*. Washington, DC: Civil Rights Division, U.S. Department of Justice.
- U.S. Department of Labor, Employment and Training Administration. (1993). *JTPA: Improving Assessment: A Technical Assistance Guide*. Washington, DC: Author.
- U.S. Office of Personnel Management. (1994). *Guide for Administering Written Employment Examinations to Persons with Disabilities*. Washington, DC: Author.

APPENDIX B: Glossary of Assessment Terms

ability test

A test that measures the current performance or estimates future performance of a person in some defined area of cognitive, psychomotor, or physical functioning.

achievement test

A test that measures acquired knowledge or skills, usually as the result of previous instruction.

adverse impact

A situation in which members of a particular race, sex, or ethnic group have a substantially lower rate of selection in hiring, promotion, or other employment decisions.

alternate forms

Two or more forms of a test that are similar in nature and intended to be used for the same purpose.

assessment

Any test or procedure used to measure an individual's employment or career-related qualifications or characteristics.

basic skills test

Assessments of basic or minimal competencies in such areas as reading, simple mathematics, basic writing abilities, and other skills that are widely required in training and employment settings.

coaching

Instructional activities designed to improve the test performance of prospective test takers.

compensatory approach

See *counterbalanced approach*.

concurrent validity

See *criterion-related validity*.

construct

A theoretical characteristic or concept (e.g., numerical ability, conscientiousness) that has been constructed to explain observable patterns of behavior.

construct-related validity

The extent to which a test measures a specific theoretical construct, characteristic, or trait. Examples of constructs are mechanical ability, physical endurance, and introversion.

content-related validity

The extent to which the content of a test samples or represents the subject area or behavior it is intended to measure.

converted score

A raw score that has been converted by numerical transformation (for example, to percentile ranks or standard scores) to facilitate comparison of individual scores with group norms.

correlation

A statistic that indicates the degree to which two variables relate to each other, such as a test score and job performance, or one test with another test.

counterbalanced approach

An approach to personnel assessment that allows high scores in one or more areas to be counterbalanced with low scores in another area.

criterion

A measure of performance, such as productivity rate, accident rate, or supervisory ratings. Test scores are used to predict criteria.

criterion-related validity

The degree to which scores on an assessment instrument correlate with some external criterion, such as job performance. When the assessment instrument and the criterion are measured at about the same time, it is called *concurrent validity*; when the criterion is measured at some future time, it is called *predictive validity*.

derived score

See *converted score*.

equivalent forms

See *alternate forms*.

expectancy table

A table that shows the probability of different criterion outcomes for each test score.

hurdles approach

See *multiple hurdles approach*.

inventory

A questionnaire or checklist that elicits information about an individual in such areas as work values, interests, attitudes, and motivation.

job analysis

A systematic process used to identify the tasks, duties, responsibilities, and working conditions associated with a job and the knowledge, skills, abilities and other characteristics required to perform that job.

mean

The average score in a group of scores, computed by adding all the scores and dividing the sum by the number of cases.

median

The middle score in a group of ranked scores. It is the point or score that divides the group into two equal parts. The median is also known as the 50th percentile.

multiple hurdles approach

An approach to selection decisions that requires a candidate to pass all tests in sequence in order to qualify.

normal curve

A mathematical curve that is the basis of many statistical analyses. The curve is bilaterally symmetrical, with a single bell-shaped peak in the center. Most distributions of human traits, such as height, mathematical ability, and manual dexterity, approximate the normal curve.

norm group

The population sample that is used to determine where a person's results fall in a specified distribution of scores.

norms

Descriptive statistics that are used to summarize the test performance of a specified group, such as a sample of workers in a specific occupation. Norms often are assumed to represent a larger population, such as all workers in an occupation.

parallel forms

See *alternate forms*.

percentile score

The score on a test below which a given percentage of scores fall. For example, a score at the 65th percentile is equal to or higher than the scores obtained by 65% of the people who took the test.

predictive validity

See *criterion-related validity*.

rank ordering

The process of ranking individuals based on their relative test scores, from the highest to the lowest score.

raw score

The obtained score on a test, usually determined by counting the number of correct answers.

reference group

The group of individuals used to develop a test.

reliability

The degree to which test scores are consistent, dependable, or repeatable.

reliability coefficient

A coefficient of correlation that indicates the degree to which test scores are dependable or repeatable.

standard deviation

A statistic used to describe the variability within a set of scores. It indicates the extent to which scores vary around the mean or average score.

standard error of measurement (SEM)

A statistic that gives an indication of the amount of error in a measurement system. It indicates a range within which a test taker's "true" score is likely to fall.

standard score

A score that describes the location of a person's score within a set of scores in terms of its distance from the mean in standard deviation units.

standardized test

A test developed using professionally prescribed methods that provides specific administration requirements, instructions for scoring, and instructions for interpreting scores.

target group

The population or group of individuals whom the test user wishes to assess.

test

Any instrument or procedure that samples behavior or performance. A personnel or employment test is the general term for any assessment tool used to measure an individual's employment qualifications, capabilities, or characteristics.

validity

The degree to which actions or inferences based on test results are meaningful or supported by theory and empirical evidence.

validity coefficient

A numerical index that shows the strength of the relationship between a test score and a criterion, such as job performance.

Assignment 4

Prior to beginning work on this assignment, review Chapters 8 and 9 in your textbook.

In this assignment, you will compare projective and objective methods of personality assessment. Research a minimum of three peer-reviewed articles in the Ashford University Library that were published within the last 15 years on these techniques. In your paper, you will provide an evaluation of these techniques organized according to the outline provided below. Use information from your researched peer-reviewed articles and required sources to support your work in each section.

Section 1: Objective Personality Assessment

- Define the term *objective* in objective methods of personality assessment.
- Summarize the features of objective methods of personality assessment, and provide at least three examples of these types of measures.
- Explain the assumptions on which objective methods are based, and provide an analysis of empirical research testing the validity of the assumptions you identified.
- Appraise the research exploring the technical adequacy (i.e., reliability and validity) of objective tests.
- Describe the impact of social and culture variability on the administration and interpretation of objective tests.

Section 2: Projective Personality Assessment

- Define the term *projective* in projective methods of personality assessment.
- Summarize the features of projective methods of personality assessment, and provide at least three examples of these types of measures.
- Explain the assumptions on which projective methods are based, and provide an analysis of empirical research testing the validity of the assumptions you identified.
- Appraise the research exploring the technical adequacy (i.e., reliability and validity) of projective tests.
- Describe the impact of social and culture variability on the administration and interpretation of projective tests.

Section 3: Synthesis, Conclusions, and Recommendations

- Write a brief one-paragraph scenario for a fictitious client. Include the following information: presenting concerns (reason for referral), age, gender, ethnicity, language(s), and any other significant information (e.g., military status, health issues, marital status, sexual orientation, etc.).
- Debate the arguments supporting and opposing the use of projective and objective personality assessments with your identified client.
- Select a minimum of one objective and one projective measure to use with your client. Compare the use of the selected projective and objective personality measures with your identified client.
- Analyze the advantages and limitations of each assessment measure you selected.
- Compose recommendations to improve the validity of personality assessment.

The Evaluation of Objective and Projective Measures of Personality

- Must be four to six double-spaced pages in length (not including title and references pages) and formatted according to APA style as outlined in the Ashford Writing Center. (Links to an external site.)
- Must include a separate title page with the following:
 - Title of paper
 - Student's name
 - Course name and number
 - Instructor's name
 - Date submitted
- Must use at least three peer-reviewed sources published within the last 15 years in addition to the course text.
- Must document all sources in APA style as outlined in the Ashford Writing Center.
- Must include a separate references page that is formatted according to APA style as outlined in the Ashford Writing Center.

GUIDELINES EDITORIAL

Advancing Personality Assessment Terminology: Time to Retire "Objective" and "Projective" As Personality Test Descriptors

Gregory J. Meyer

Department of Psychology
University of Toledo

John E. Kurtz

Department of Psychology
Villanova University

For decades psychologists have classified personality tests dichotomously as *objective* or *projective*. These terms appear in scientific articles and textbooks and have become so entrenched that it is common to see separate courses in graduate clinical programs using these labels in course titles (e.g., "Objective Assessment," "Projectives"). In the interest of advancing the science of personality assessment, we believe it is time to end this historical practice and retire these terms from our formal lexicon and general discourse describing the methods of personality assessment.

For personality tests, the term *objective* typically refers to instruments in which the stimulus is an adjective, proposition, or question that is presented to a person who is required to indicate how accurately it describes his or her personality using a limited set of externally provided response options (true vs. false, yes vs. no, Likert scale, etc.). What is *objective* about such a procedure is that the psychologist administering the test does not need to rely on judgment to classify or interpret the test-taker's response; the intended response is clearly indicated and scored according to a pre-existing key. As a result, however, the necessity for judgment is passed on to the test taker. She must interpret the question, consider her personal characteristics, evaluate herself relative to others as best she can, decide the extent to which the characteristic fits her personality, and then choose whether to honestly convey this information in her response.

On the other hand, the term *projective* typically refers to instruments in which the stimulus is a task or activity that is presented to a person who is required to generate a response with minimal external guidance or constraints imposed on the nature of that response. What is *projective* in a test like this is the requirement to generate a response in the face of

ambiguity; in so doing, the person projects or puts forward elements of her personal characteristics.

Unfortunately, the terms *objective* and *projective* carry multiple, often unclear, meanings, including some connotations that are very misleading when applied to personality assessment instruments and methods. For instance, the term *objective* implies accuracy and precision that is impervious to biasing influences. These are desirable and positive connotations. One problem is that these positive connotations are not fully warranted for the inventories to which they typically refer. Scoring errors are certainly one potential concern (e.g., Allard & Faust, 2000). More substantively, however, if the kind of self-report scales that are classified as *objective* actually were "objective" in a meaningful sense of that word, then there would not be such a huge literature examining the various response styles and biases that affect scores derived from these instruments. In fact, the literature addressing the topic of response styles, malingering, and test bias in these measures appears larger than the literature on any other focused issue concerning their validity or application. Beyond bias and frank distortion, Meehl (1945) pointed out more than half a century ago that the processes influencing a test-taker's response include ambiguity inherent in the test items, limitations in self-knowledge or self-perception, personal dynamics, and even projections. Another serious issue that results from applying the term *objective* to certain personality instruments is that those so labeled will tend to be viewed positively simply by virtue of the term's positive connotations. Tests that are not so categorized will tend to be viewed less positively, regardless of psychometric data, because they are, after all, not *objective*. Accordingly, an unintended consequence of this terminology is that it may encourage or per-

petuate prejudices regarding the many alternative methods of assessment that do not carry the objective label.

At the same time, the connotations of the term projective also do not always apply when considering the instruments typically classified as projective. For instance, responses to the Rorschach inkblots often have more to do with stimulus classification and problem solving styles than to projection in a classical Freudian sense of the term, where undesirable personal feelings or impulses are seen as residing outside the self (see Exner, 1989). Similar difficulties emerge when considering the expanded definition of the term projective as Frank (1939) first defined it in reference to types of personality tests. Frank considered a projective test one that would

induce the individual to reveal his way of organizing experience by giving him a field (objects, materials, experiences) with relatively little structure and cultural patterning so that the personality can project upon that plastic field his way of seeing life, his meanings, significances, patterns, and especially his feelings. Thus we elicit a projection of the individual personality's *private world* because he has to organize the field, interpret the material and react affectively to it. ... The important and determining process is the subject's personality which operates upon the stimulus-situation as if it had a wholly private significance for him alone or an entirely plastic character which made it yield to the subject's control. (*italics in the original*; pp. 402-403)

This conceptualization of a projective test implies that stimulus features or task requirements are essentially immaterial; personality characteristics will shine through with force and clarity regardless of the medium. Although desirable, this view is clearly incorrect. For instance, it is well documented that the largest source of variability in Rorschach scores is the number and complexity of responses given (e.g., Meyer, 1993, 1997). The personality characteristics associated with this style of responding are interpretively quite important in their own right. However, the presence of this response complexity confounds efforts to interpret the test scores that psychologists are most interested in interpreting (e.g., Exner, 2003).¹ The situation is similar with thematic storytelling techniques, in which the number of words given and the specific stimulus pictures selected for use exert a powerful influence on the final scores obtained (e.g., Blankenship et al., 2006; Hibbard et al., 1994; Pang & Schultheiss, 2005).

Thus, the old and familiar terminology of objective and projective personality tests has misleading connotations that will not serve the field well as we seek to have a more differentiated

understanding of assessment methods. A relevant question then becomes: What is better alternative terminology?

It is fairly easy to identify reasonable alternatives to supplant the term objective. Almost exclusively, this term has been applied to structured questionnaires that are completed by the target person him or herself. Consequently, a reasonable alternative is to refer to these tests as "self-report inventories" or "patient-rated questionnaires." Moreover, to advance the science of assessment, it is equally important to differentiate self-report inventories from inventories completed by knowledgeable informants. Given that sources of information in personality assessment are far from interchangeable (e.g., Achenbach, Krukowski, Dumenci, & Ivanova, 2005; Achenbach, McConaughy, & Howell, 1987; Costa & McCrae, 1992; De Los Reyes & Kazdin, 2005; Kraemer, Measelle, Ablow, Essex, Boyce, & Kupfer, 2003; Meyer, 2002; Meyer et al., 2001), it would be optimal to further differentiate all questionnaire methods by specifying the type of informant providing judgments. Thus, peer ratings would be labeled as such and differentiated from spouse-report scales, parent-rated questionnaires, and so forth.

It is not as easy to identify a single term or phrase that could supplant the term projective. In fact, when discussing this issue with colleagues, disagreements about a suitable substitute appear to be one of the greatest obstacles to change. No single term seems fully adequate. The instruments that are typically subsumed under the projective label include the Rorschach (1921/1942) and other inkblot tests (e.g., Holtzman, Thorpe, Swartz, & Herron, 1961), Murray's (1943) Thematic Apperception Test and the subsequently developed Picture Story Exercise stimuli (e.g., Smith, 1992), sentence completion measures, and various figure drawing tasks (e.g., Naglieri & Pfeiffer, 1992). The wide differences among these tasks makes it challenging to find a suitable alternative term that accommodates all of their diverse features. Some possibilities include "performance tasks," "behavioral tasks," "constructive methods," "free response measures," "expressive personality tests," "implicit methods," or even "attributive tests." It is unlikely that any one of these labels would satisfy all experts. However, it is the very difficulty of finding a suitable alternative that speaks to the inadvisability of using a global term to characterize the essence of all these measures. In turn, this highlights the need to drop the term projective from the assessment method lexicon.

One of the initial steps to advance the scientific understanding of any phenomenon is to name and classify its components in a meaningful way. The unsuitable and primitive nature of the term projective is revealed when trying to arrive at an umbrella label to characterize tasks as diverse as drawing one's family, telling stories in response to pictures, and stating what an inkblot looks like. Applying a global and undifferentiated term to such a diverse array of assessment tasks seems akin to physicians classifying medical tests as either "visual tests" or "nonvisual tests," with the visual category including tasks ranging from observing reflexes to endoscopy to MRI, and the nonvisual category including tasks ranging from palpation

¹The confounding influence of this so called "first factor" variance is pervasive with other instruments as well. An excellent discussion of the problem and of a sophisticated effort to mitigate its influence on the MMPI-2 can be found in the recently published Special Issue of the *Journal of Personality Assessment* (Meyer, 2006) dealing with the MMPI-2 Restructured Clinical Scales (Tellegen et al., 2003).

methods (e.g., abdominal tenderness) to olfactory methods (e.g., odors indicative of infection) to auditory methods (e.g., detecting wheezes with a stethoscope).

Just as it would be regressive to apply such a simplistic categorization to medical tests, the field of personality assessment will not advance by relying on crude terminology to globally characterize all the tasks that are not self-report questionnaires or informant rating scales. Thus, if one of the substitute terms noted above does not seem suitable to replace projective, it would be most optimal for clinicians, researchers, and teachers to simply refer to assessment tasks by their specific name, for example, the Rorschach Inkblot Method, Holtzman Inkblot Task, Murray's TAT, Loevinger's SCT. The *Journal of Personality Assessment* will facilitate the transition to more adequately differentiated assessment terminology by asking authors to avoid referring to categories of personality tests as objective or projective. We hope other assessment journals will join this effort and adopt a similar position.

This editorial guideline is not meant to imply that the words objective and projective cannot be used in the context of referring to specific data from personality instruments. It is certainly true that all personality tests can provide more or less objective data. It is also the case that instruments like the Rorschach or TAT can capture projected personality characteristics, whether defined narrowly as by Freud or more broadly as by Frank, and this can also occur when patients complete self-report inventories (Meehl, 1945). There is no problem if authors carefully and deliberately choose these terms to further scientific communication (e.g., when one is describing aspects of inkblot responses that are truly believed to indicate projected dynamics). Rather, our objection is with the reflexive use of historically ingrained terms that poorly describe the complex and distinctive methods used to assess personality.

ACKNOWLEDGMENTS

Our thoughts on this topic have benefited from helpful and insightful input from many people. Those who commented on this document included Robert Bornstein, Anita Boss, Virginia Brabender, Philip Caracena, Robert Erard, Barton Evans, Leonard Handler, Radhika Krishnamurthy, Robert McGrath, Joni Mihura, David Nichols, Bruce Smith, Donald Viglione, Irving Weiner, and Jed Yalof.

REFERENCES

- Achenbach, T. M., Krukowski, R. A., Dumenci, L., & Ivanova, M. Y. (2005). Assessment of adult psychopathology: Meta-analyses and implications of cross-informant correlations. *Psychological Bulletin*, 131, 361-382.
- Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin*, 101, 213-232.
- Allard, G., & Faust, D. (2000). Errors in scoring objective personality tests. *Assessment*, 7, 119-129.
- Blankenship, V., Vega, C. M., Ramos, E., Romero, K., Warren, K., Keenan, K., et al. (2006). Using the multifaceted Rasch model to improve the

- TAT/PSE measure of need for achievement. *Journal of Personality Assessment*, 86, 100-114.
- Costa, P. T., Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Odessa, FL: Psychological Assessment Resources.
- De Los Reyes, A., & Kazdin, A. E. (2005). Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin*, 131, 483-509.
- Exner, J. E. (1989). Searching for projection in the Rorschach. *Journal of Personality Assessment*, 53, 520-536.
- Exner, J. E. (2003). *The Rorschach: A Comprehensive System* (4th ed.). New York: Wiley.
- Frank, L. K. (1939). Projective methods for the study of personality. *Journal of Psychology*, 8, 389-413.
- Hibbard, S., Farmer, L., Wells, C., Difillipo, E., Barry, W., Korman, R., & Sloan, P. (1994). Validation of Cramer's Defense Mechanism Manual for the TAT. *Journal of Personality Assessment*, 63, 197-210.
- Holtzman, W. H., Thorpe, J. S., Swartz, J. D., & Herron, E. W. (1961). *Inkblot perception and personality*. Austin: University of Texas.
- Kraemer, H. C., Measelle, J. R., Ablow, J. C., Essex, M. J., Boyce, W. T., & Kupfer, D. J. (2003). A new approach to integrating data from multiple informants in psychiatric assessment and research: Mixing and matching contexts and perspectives. *American Journal of Psychiatry*, 160, 1566-1577.
- Meehl, P. E. (1945). The dynamics of "structured" personality tests. *Journal of Clinical Psychology*, 1, 296-303.
- Meyer, G. J. (1993). The impact of response frequency on the Rorschach constellation indices and on their validity with diagnostic and MMPI-2 criteria. *Journal of Personality Assessment*, 60, 153-180.
- Meyer, G. J. (1997). On the integration of personality assessment methods: The Rorschach and MMPI. *Journal of Personality Assessment*, 68, 297-330.
- Meyer, G. J. (2002). Implications of information-gathering methods for a refined taxonomy of psychopathology. In L. E. Beutler & M. Malik (Eds.), *Rethinking the DSM: Psychological perspectives* (pp. 69-105). Washington, DC: American Psychological Association.
- Meyer, G. J. (Ed.). (2006). *The MMPI-2 Restructured Clinical Scales* [Special Issue]. *Journal of Personality Assessment*, 87(2).
- Meyer, G. J., Finn, S. E., Eyde, L., Kay, G. G., Moreland, K. L., Dies, R. R., et al. (2001). Psychological testing and psychological assessment: A review of evidence and issues. *American Psychologist*, 56, 128-165.
- Murray, H. A. (1943). *Thematic Apperception Test manual*. Cambridge, MA: Harvard University Press.
- Naglieri, J. A., & Pfeiffer, S. I. (1992). Validity of the Draw A Person: Screening Procedure For Emotional Disturbance with a socially-emotionally disturbed sample. *Psychological Assessment*, 4, 156-159.
- Pang, J. S., & Schultheiss, O. C. (2005). Assessing implicit motives in U.S. college students: Effects of picture type and position, gender and ethnicity, and cross-cultural comparisons. *Journal of Personality Assessment*, 85, 280-294.
- Rorschach, H. (1942). *Psychodiagnostics* (5th ed.). Berne, Switzerland: Verlag Hans Huber. (Original work published 1921)
- Smith, C. P. (Ed.). (1992). *Motivation and personality: Handbook of thematic content analysis*. New York: Cambridge University Press.
- Tellegen, A., Ben-Porath, Y. S., McNulty, J. L., Arbisi, P. A., Graham, J. R., & Kaemmer, B. (2003). *The MMPI-2 Restructured Clinical Scales: Development, validation, and interpretation*. Minneapolis: University of Minnesota Press.

Gregory J. Meyer
Department of Psychology
Mail Stop 948
University of Toledo
2801 Bancroft Street
Toledo, OH 43606
Email: gmeyer@utnet.utoledo.edu

Copyright of Journal of Personality Assessment is the property of Lawrence Erlbaum Associates and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.

COMMENT

Beyond "Objective" and "Projective": A Logical System for Classifying Psychological Tests: Comment on Meyer and Kurtz (2006)

EDWIN E. WAGNER

Miami, Florida

I present a formal system that accounts for the misleading distinction between tests formerly termed *objective* and *projective*, duly noted by Meyer and Kurtz (2006). Three principles of Response Rightness, Response Latitude and Stimulus Ambiguity are shown to govern, in combination, the formal operating characteristics of tests, producing inevitable overlap between "objective" and "projective" tests and creating at least three "types" of tests historically regarded as being projective in nature. The system resolves many past issues regarding test classification and can be generalized to include all psychological tests.

Meyer and Kurtz (2006) acknowledged that "[o]ne of the initial steps to advance the scientific understanding of any phenomenon is to name and classify its components in a meaningful way" (p. 224). Yet, compelled by the evidence, Meyer and Kurtz recommended abandoning the hoary distinction between "objective" and "projective," advocating the use of specific titles such as Rorschach (1921/1942) Inkblot Method when other proposed alternatives seem unsuitable.

Following up on this recommendation, in this article, I present a system for classifying psychological tests that is formal, logical, and avoids difficulties that have arisen in the past when tests were categorized by similarities that often proved descriptive but superficial and tended to obfuscate fundamental incompatibilities. I show logically that (a) a sharp distinction between so-called objective and projective tests cannot be maintained, (b) instruments traditionally regarded as projective are of at least three different types, and (c) a practical scheme for classifying tests can be devised according to how they align on three dimensions or parameters termed *response rightness* (or rightness for short), *response latitude* (latitude), and *stimulus ambiguity* (ambiguity).

I state from the outset that these parameters have nothing to do with projection per se. They do embody properties, however, that militate against, or are conducive to, the emergence of material traditionally recognized in the literature as indicative of projection and including but not limited to responses that (a) represent repressions that, in the Freudian sense, the subject would consciously deny or disavow; (b) reflect, often symbolically, important beliefs, feelings, or action tendencies; or (c) involve an interpretation not directly attributable to the stimulus itself as when a subject imputes movement to a static inkblot.

It is recognized, of course, that even tests that used to be called projective do not necessarily yield the kinds of responses listed previously—at least not all of the time.

RESPONSE RIGHTNESS

Rightness pertains to whether a test requires predetermined answers to specific questions. Rightness is dichotomous. One group of tests can be described as *restrictive* and possessing right or wrong answers in which *right* is defined as a response that is correct, partially correct, or incorrect. The other group can be termed *permissive*, featuring tests that allow subjects to give responses of their own choosing, all of which are acceptable provided they fall within the prescribed task guidelines such as making up a story. Intelligence, aptitude, achievement, and neuropsychological tests usually call for an accurate (right) responses and are restrictive. Personality and interest tests are generally characterized by permissiveness and, in fact, test takers are often explicitly informed that "there are no right or wrong answers."

Elimination of the rightness requirement, however, does not grant complete freedom, and permissiveness is always constrained by the nature of the task. The Minnesota Multiphasic Personality Inventory-2 (MMPI-2; Butcher et al., 2001), the Rorschach (1921/1942), and the Incomplete Sentences Blank (ISB; Rotter, Lah, & Rafferty, 1992) are all permissive (no right or wrong answers) but require different response modes, i.e., true-false selections, visual perceptions, and stem-completing statements. That is, permissiveness does not imply that a test taker can respond in any fashion whatsoever.

RESPONSE LATITUDE

Latitude refers to the number of responses or choices that can be elicited by the task and significantly affects the way a test behaves. Response latitude lies along a continuum but for convenience can be viewed as being essentially trichotomous,

TABLE 1.—Classification of illustrative tests in terms of response rightness, response latitude, and stimulus ambiguity.

Latitude	Rightness	Nonambiguous			
		Low Ambiguity		Medium Ambiguity	High Ambiguity
Low	Restrictive	Wechsler Adult Intelligence Scale—III: Matrix Reasoning; Bennett Mechanical Comprehension Test; Halstead Category Test (Reitan, 1993)		Structured-Objective Rorschach Test (Stone, 1958)	
	Permissive				
High	Restrictive	Minnesota Multiphasic Personality Inventory—2; Strong-Campbell Interest Inventory (Strong, Campbell, Harmon, & Hansen, 1994)			
	Permissive				
Infinite	Restrictive	Wechsler Adult Intelligence Scale—III: Picture Arrangement			
	Permissive				
	Restrictive	Rokeach Value Survey; The Color Pyramid Test			
	Permissive				
	Restrictive	Wechsler Adult Intelligence Scale—III: Comprehension			
	Permissive				
	Restrictive	House-Tree-Person Test; Incomplete Sentence Blank			
	Permissive				
		Thematic Apperception Test; Hand Test		Rorschach; Holtzman Inkblot Technique	Symbol Elaboration Test

consisting of two clusters and an endpoint. The first cluster involves a limited number of response alternatives and is characteristic of many tests formerly referred to as objective that utilize true-false, multiple choice, and scaling formats. Tests with limited latitude are usually easier to score because they are less dependent on expert judgment.

The second cluster, positioned an appreciable distance from the first, includes a large but finite number of response possibilities such as might be obtained from tasks involving ordering, arranging, and assembling; and at the extreme end of the continuum are situated theoretically infinite response possibilities. That is, although middle cluster options may be very large, they can be calculated in advance and are finite. The first cluster is rather narrow, varying from yes-no (true-false) through multiple choice to ratings or selections along a preference scale. The second cluster is widely dispersed and directly related to the number of orderings, arrangements, and constructions that the task entails. Infinity, of course, has no limit.

Tests that are characterized by both permissiveness and potentially infinite response latitudes such as the Rorschach and Thematic Apperception Test (TAT; Murray, 1943) can be scored, but because the latitude is unlimited, it is difficult to devise a system that will accommodate all the response possibilities—which is one of the reasons for the proliferation of scoring systems for such tests.

STIMULUS AMBIGUITY

Ambiguity requires an unresolved, semistructured perceptual stimulus. In most cases, such stimuli are visual, although they need not be. Stimulus ambiguity lies along a continuum in which, perceptually speaking, both extremes are unambiguous. At some theoretical midpoint, a stimulus is optimally ambiguous as far as evoking meaningful material for personality evaluation is concerned.

One end of the ambiguity continuum is so familiar that one tends to take it for granted. It consists of statements, questions, and instructions with clear intent and high structure that are nonambiguous. At the opposite extreme, however, are stimuli that are so unstructured that they too are no longer ambiguous such as blank sheets of paper.

An ISB stem is straightforward given suitable reading comprehension, and the task, although permissive, is unambiguous. At the other extreme, a blank sheet of paper, the stimu-

lus for the Draw-A-Person (D-A-P) test (Machover, 1949), is also unambiguous because it is completely lacking in structure, and there is nothing to interpret. The D-A-P task, like the ISB, is permissive but not ambiguous. In contrast, a Rorschach plate is semistructured and obligates the subject to convert an inkblot into a meaningful image before articulating a response. Responding to the ISB and D-A-P involve decisions about what to write and what to draw, whereas interpreting a Rorschach plate calls for a resolution of a semistructured perceptual ambiguity before a response can be formulated.

The blank TAT card, like the blank sheet of paper for the D-A-P, is highly permissive but unambiguous. Once some structure is imposed on such mediums, however, they can become ambiguous. A drawing test, such as the Symbol Elaboration Test (Krout, 1950), is ambiguous because it requires the subject to complete simple, semistructured line segments. In this context, it is important not to confuse permissiveness and ambiguity. The two combine and along with latitude determine the overall nature of the task, but permissiveness is cognitive, and ambiguity is perceptual. Although a test such as the ISB may seem far less ambiguous than the D-A-P, neither test makes use of an ambiguous stimulus. The ISB seems more like tests psychologists are used to dealing with, however, inasmuch as it is printed on a standard form. Further, an ISB stem significantly orients the nature of the fill in and is, therefore, less permissive than the D-A-P in which only the gender of the to-be-drawn figure is typically specified in advance.

The nature and quality of a semiambiguous stimulus will, of course, also influence a subject's reactions and guide interpretations, even for tests that seem closely related—especially if there are nuanced differences in permissiveness as well as ambiguity. The Rorschach and the Holtzman Inkblot Technique (HIT; Holtzman, Thorpe, Swartz, & Herron, 1961) are both inkblot tests, but the HIT contains more plates of minimal ambiguity over all 45 cards and is less permissive, requiring just one response per card. Therefore, the Rorschach can be regarded as being more permissive and more ambiguous.

COMMENTARY AND ILLUSTRATIONS

Table 1 illustrates how various representative tests can be classified in terms of rightness, latitude, and ambiguity.

Based on the overlaps among tests created by these parameters, it is obvious that Meyer and Kurtz (2006) were correct in that there are no iron clad rules that can unequivocally classify all tests as being either objective or projective. There are, however, some generalizations that can be derived from these paradigms that provide insights into how certain classes of psychological tests must perform function.

The necessary but not sufficient condition for most tests formerly regarded as being objective would be limited latitude, and the necessary but not sufficient requirement for most techniques previously treated as projective would be permissiveness. Clarification of the properties of instruments formerly regarded as projective is especially difficult because tests that seem most capable of revealing projection (as previously defined) appear to be permissive with some degree of perceptual ambiguity and infinite response latitude. In the past, however, many so-called projective techniques have lacked one or more of these attributes.

Most intelligence, achievement, and aptitude tests are nonpermissive, unambiguous, and response limited, but again, there are exceptions. The Wechsler Adult Intelligence Scale-III (WAIS-III; Wechsler, 1997) is a hybrid including subtests with response latitudes ranging from limited to large to infinite. In terms of rightness, however, all WAIS-III subtests are restrictive.

Most tests that used to be referred to as objective personality tests, such as the MMPI-2, are permissive, response limited, and unambiguous. This combination of parameters would also characterize most interest tests.

Noting the effects of large but finite response latitudes can now explicate some putative projective techniques that have been difficult to classify in the past. The Color Pyramid Test (Pfister, 1950) is an interesting example of a so-called projective technique that is permissive, latitude large (but not infinite), and nonambiguous. It consists of colored objects used to build a pyramid, and there are only so many ways to put the pieces together. Similarly, the Rokeach (1967) Value Survey, which requires individuals to place sets of values in rank order, would be classified as permissive, latitude large, and unambiguous. It seems doubtful that such tests would yield projective material in the psychoanalytic sense, whereas tests that are latitude infinite do, on occasion, reveal such insights. For example, on the ISB, a patient produced the following completion, evoking the classic psychoanalytic notion of the mother figure as the evil, wicked spider: "A mother eats her children's hearts like a big fat spider." Tests with infinite latitude can also, albeit infrequently, reflect deviancy via responses that contradict reality, for example, "I suffer because vampires bite me at night."

Tests that are ambiguous as well as permissive and response infinite, such as the Rorschach, seem even more capable of eliciting projective responses, at least on occasion. For example, the subject who sees "Two dogs trying to get a bone but they'll never be satisfied" (Card II) does seem to be revealing a possible oral fixation and/or sibling rivalry. Like the ISB, the Rorschach can reflect deviancy when reality is contradicted, for example, "Pink tigers" (Card VIII, side Ds); but because an ambiguous stimulus is involved, deviancy can also be suspected when obvious proprieties of the blot are violated, for example, "A red, white, and blue banner" (top of achromatic Card VI).

Perceptual ambiguity does, therefore, appear to be of real clinical value in personality assessment, but because lack of

structure lies along a continuum, it is difficult to judge how much ambiguity is optimum. It does appear, however, that highly unstructured stimuli do not lend themselves well to projections. The TAT and the Hand Test (Wagner, 1962/1983) are examples of tests with stimuli that are not markedly ambiguous yet can elicit meaningful projective responses. Consistent with this finding, it is hypothesized that amorphous stimuli, such as cloud pictures or line squiggles (Krout, 1950), that are positioned toward the unstructured end of the spectrum are less effective as projective mediums.

At any rate, based on the foregoing analysis, it is now possible to state that much of the confusion regarding the untenable objective-projective separation is attributable to the fact that there are actually three types of instruments with different operating characteristics, all of which have, at one time or another, received the appellation projective technique:

1. Type A projection: Permissive, large response latitude, unambiguous (example—the Color Pyramid Test).
2. Type B projection: Permissive, infinite response latitude, unambiguous stimuli (example—the ISB).
3. Type C projection: Permissive, infinite response latitude, ambiguous stimuli (example—the Rorschach).

TOWARD A MORE MEANINGFUL CLASSIFICATION OF PSYCHOLOGICAL TESTS

It has been suggested that tests can be meaningfully classified and therefore better understood by abrogating the distinction between objective and projective tests, as recommended by Meyer and Kurtz (2006), and substituting categories based on rightness, latitude, and ambiguity. For pedagogical reasons, in writing textbooks or instructing students, a few other appropriate descriptors could easily be added to flesh out the taxonomic skeleton. For example, a mechanical comprehension test could be characterized as a series of pictures designed to illustrate mechanical principles and possessing the formal properties of rightness (the selections are right or wrong), limited response latitude (multiple choice format), and nonambiguity (the pictures and questions are distinct and of clear intent). Additionally, the Rorschach could be described as a personality test consisting of 10 inkblots and possessing the formal properties of permissiveness, infinite response latitude, and stimulus ambiguity (semistructured inkblots).

REFERENCES

- Bennett, G. K. (1994). *Bennett Mechanical Comprehension Test* (2nd ed.). San Antonio, TX: The Psychological Corporation.
- Butcher, J., Graham, J., Ben-Porath, Y. S., Tellegen, A., Dahlstrom, W., & Kaemmer, B. (2001). *Minnesota Multiphasic Personality Inventory-2: MMPI: Manual for administration, score, and interpretation* (Rev. ed.). Minneapolis: University of Minnesota Press.
- Holtzman, W. H., Thorpe, J. S., Swartz, J. D., & Herron, E. W. (1961). *Inkblot perception and personality—Holtzman Inkblot Technique*. Austin: University of Texas Press.
- Krout, J. (1950). *Symbol Elaboration Test*. Chicago, IL: Psychological Institute.
- Machover, K. (1949). *Personality projections in the drawing of the human figure*. Springfield, IL: Thomas.
- Meyer, G. J., & Kurtz, J. E. (2006). Advancing personality assessment terminology: Time to retire "objective" and "projective" as personality test descriptors. *Journal of Personality Assessment*, 87, 223–225.
- Murray, H. A. (1943). *Thematic Apperception Test manual*. Cambridge, MA: Harvard University Press.

Editorial

25 Years in Promoting Projective Assessment: A Silver Jubilee Tribute

Historically, projective techniques have left an impressive, yet contentious, footprint worldwide (Groth-Marnat, 2009; Handler & Thomas, 2014; Musewicz et al., 2009; Piotrowski, 2015; Piotrowski, Keller, & Ogawa, 1993; Teglassi, 2010; Wood et al., 2011). Interestingly, survey data from recent 'test use' studies point to a rather bleak view regarding the status of projective techniques. Ready and Veague (2014) reported that no projective tests ranked among the top 10 tests being taught in clinical psychology programs. Wright et al. (2016), in a national sample of professional psychologists, found that the Rorschach was the only projective method used frequently (ranked in top 13 tests) among a myriad of psychological assessment domains. But do these reports provide an accurate portrait regarding the extent of clinical emphasis devoted to projective techniques in both training and practice? Undoubtedly, there is a perennial need to empirically address the clinical breadth and status of specific projective methods and their place in contemporary assessment practices.

As 2018 denotes the 25th year of continuous publication of the *SIS Journal of Projective Psychology & Mental Health*, this editorial attempts to elucidate the issue regarding the extent of **current** usage and emphasis on projective techniques in the mental health field. To that end, I provide data-based evidence in support of the sustainability of projective testing in contemporary professional psychology, based on data reported in very recent studies, from 2013-2017, on a) training in psychological testing, and b) assessment practices by professionals. Moreover, to provide a complete composite on the use of projective methods, survey-based findings in the periodicals literature as well as data reported in recent dissertation research are presented.

First, an objective appraisal regarding research interest in projective techniques seems in order. An online search of the database PsycINFO (conducted September 28, 2017), on the term 'projective techniques', yielded 1,857 articles, 43 dissertations, and 73 book chapters since the year 2000. Based on this search, Table 1 presents the top journals, in rank order, where major research findings on projective techniques have been published. Surely this illustration does not reflect a moribund state of affairs regarding projective assessment.

Table 1. Major Publication Outlets

Journal of Personality Assessment
SIS Journal of Projective Psychology & Mental Health
Psychologie Clinique et Projective
Qualitative Market Research
Rorschachiana
Perceptual & Motor Skills
Attachment & Human Development
Bulletin de Psychologie

Pertinent to the debate regarding projective tests, an interesting Delphi poll study on "discredited" tests, found that Human-Figure-Drawings, H-T-P, TAT, Rorschach, and Sentence completion methods were not considered 'probably discredited' (Norcross et al., 2006). Although some lesser-known projective instruments have fallen out-of-favor, a recent study reported that the Somatic Inkblot Series has attracted robust research attention worldwide (Piotrowski, 2017a; see Dubey et al. in this issue, for a review). Thus, while projective techniques have lost some of their luster over the decades, based on recent survey-based findings, there is still avid interest on projective approaches in mental health assessment by devoted clinicians.

Table 2 presents 10 survey-based studies, reported over the past 5 years, on the topic of assessment practices in both professional training and practice settings. Given the

proliferation of broad-band personality tests and brief, symptom-focused scales that have been introduced in the professional literature over the past 2 decades, it is rather commendable that several projective techniques continue to rank among the top 50 most popular psychological tests. Moreover, while projective methods have indeed been de-emphasized in graduate clinical training (Piotrowski, 2015b), several projective tests continue to be embraced by clinical faculty at internship sites (see Bates, 2016; Ready et al., 2016; Stedman et al., 2017). In fact, internship directors encourage pre-internship instruction, obtained during doctoral clinical training; in projective assessment methods (see Faith, 2016). The key factor, in terms of sustainability of projective techniques, will be centered on the status of personality assessment in the professional curriculum (Evans & Finn, 2017; Piotrowski, 2017b).

Thus, based on the findings of these recent reports, Projective methods indeed have a worthwhile place in the clinical

armamentarium- Undoubtedly, not as lofty a place as in the decades past, but still cherished by a select minority of proponents who appreciate the challenge in trying to understand the complexities of how individuals perceive and deal with both internal and external reality.

So as SIS celebrates its 25th Jubilee, projective assessment is represented well in this commemorative issue on topics such as the Somatic Inkblot Series, the Rorschach, the TAT, Sentence Completion methods, and assessment training. Indeed, verbal expressive techniques (Panek et al., 2013) and idiographic assessment (Beltz et al., 2016) continue to attract both research and clinical attention. It is apparent that the projective assessment enterprise is not moribund—well exemplified by Teglassi (2013, p. 113), “Insofar as variation in projective test responses are caused by constructs’ capturing phenomena that matter in the lives of individuals, projective techniques are legitimate tools for science and practice.”

Table 2. Recent (2013-2017) Survey-Based Studies on Testing Emphasis with Projective Techniques

Study	Country	Sample	Major Findings
Neukrug et al. (2013)	USA	Based on survey data from 210 counselor educators across the U.S., this study examined graduate-level coverage of assessment instruments by instructors	TAT ranked #12; H-T-P #13; Rorschach #16; Kinetic Family Drawings #30; Sentence completion #36; Children's Apperception Test #40; Human Figure Drawings #48.
Peterson et al. (2014)	USA	926 counselors (clinical mental health, school, occupational) rated tests of all types regarding usage	Amongst a copious set of testing instruments, House-Tree-Person ranked 17 th ; Human-Figure-Drawings 21 st ; DAP 35 th ; TAT 40 th ; Kinetic Family Drawings 47 th ; Sentence completion 52 nd ; Rorschach 57 th .
Ready & Veague (2014)	USA	Compared training in psychological assessment across 3 training models (Clinical-Science, Scientist-Practitioner, Practitioner-Scholar) in APA-Accredited programs	No projective tests ranked in the top 10; only practitioner-scholar programs offer limited coverage on projective techniques.
Wechsler et al. (2014)	Iberian/Latin-	Test development & usage in Portugal, Spain, Argentina,	Projective tests very popular in Venezuela; Rorschach somewhat

	American countries	Venezuela, and Brazil	popular in Brazil and Argentina; Spain and Portugal indicated low usage of projective tests. Sentence completion methods are not used by practitioners in any of these countries.
Ready et al. (2016)	USA & Canada	236 internship directors' views on pre-doctoral academic training in testing and assessment	36% of directors endorsed the Rorschach; 25% favored Sentence completion tests; 19% approved projective drawings; PT is somewhat more emphasized in child clinics.
Wright et al. (2016)	USA	Views of 279 professional APA psychologists on assessment/testing practices	In the projective testing category, only the Rorschach ranked in 'Top 10'.
Kohns (2016)	USA	Experiences of 148 advanced clinical students from both APA & non-APA professional psychology programs regarding assessment training	Academic exposure to projective assessment was rather high (30 coursework hours); but mostly focused on Rorschach CS & R-PAS.
Bates (2016)	USA	182 APPIC internship directors' views on extent of emphasis in specific assessment/testing training	Sites emphasizing PT usage by interns (TAT-45%; Sentence completion-44%; Rorschach-40%; Drawings-32%); although there has been a noted decrease in emphasis on most PT in recent years, directors prefer incoming interns to have academic exposure to projective assessment.
Mihura et al. (2017)	USA	Assessment/testing training reported by 83 APA clinical psychology programs	63% of programs 'covered' the TAT & Rorschach; Sentence completion (47%); Figure drawings (36%); Roberts Apperception Test (27%); PT was more emphasized in practitioner-focused clinical programs.
Stedman et al. (2017)	USA	335 APPIC psychology internship sites	Across child, adult, and 'mixed' program types, the Rorschach was used by interns (in 26-35% of sites); Thematic tests (in 19-41%); Sentence Completion (in 18-41%); Drawing tests (in 9-36%).

References:

- Bates, S. (2016). Internship directors' perspectives on psychological assessment training: Current status and emerging trends. Unpublished doctoral dissertation, Pepperdine University.
- Beltz, A.M., Wright, A., Sprague, B.N., & Molenaar, P. (2016). Bridging the nomothetic and idiographic approaches to the analysis of clinical data. *Assessment*, 23(4), 447-458.
- Evans, F.B., & Finn, S.E. (2017). Training and consultation in psychological assessment with professional psychologists: Suggestions for enhancing the profession and individual practices. *Journal of Personality Assessment*, 99(2), 175-185.
- Faith, A. (2016). Internship directors' perspectives on emerging trends in psychological assessment training and practice. Unpublished doctoral dissertation, Pepperdine University.
- Groth-Marnat, G. (2009). *Handbook of Psychological Assessment* (5th ed.). Hoboken, NJ: Wiley.
- Handler, L., & Thomas, A.D. (Eds.) (2014). *Drawings in assessment and psychotherapy: Research and applications*. New York: Routledge.
- Kohns, B.C. (2016). Doctoral training in psychological assessment and competency in assessment ethics. Unpublished doctoral dissertation, Alliant International University, San Diego.

- McCloskey, L.C. (2014). Construct and incremental validity of the Rotter Incomplete Sentences Blank in adult psychiatric outpatients. *Psychological Reports: Measures & Statistics*, 114(2), 363-375.
- Mihura, J.L., Roy, M., & Graceffo, R.A. (2017). Psychological assessment training in clinical psychology doctoral programs. *Journal of Personality Assessment*, 99(2), 153-164.
- Musewicz, J., Marczyk, G., Knauss, L., & York, D. (2009). Current assessment practice, personality measurement, and Rorschach usage by psychologists. *Journal of Personality Assessment*, 91(5), 453-461.
- Neukrug, E., Peterson, C.H., Bonner, M., & Lomas, G.I. (2013). A national survey of assessment instruments taught by counselor educators. *Counselor Education & Supervision*, 52, 207-219.
- Norcross, J.C., Koocher, G.P., & Garofalo, A. (2006). Discredited psychological treatments & tests: A Delphi poll. *Professional Psychology: Research and Practice*, 37(5), 515-522.
- Panek, P.E., Jenkins, S.R., Hayslip, B., & Moske, A.K. (2013). Verbal expressive personality testing with older adults: 25+ years later. *Journal of Personality Assessment*, 95(4), 366-376.
- Peterson, C.H., Lomas, G.I., Neukrug, E.S., & Bonner, M.W. (2014). Assessment use by counselors in the United States. *Journal of Counseling & Development*, 92, 90-98.
- Piotrowski, C. (2017a). Rorschach research through the lens of bibliometric analysis: Mapping investigatory domain. *Journal of Proj. Psychol. & Mental Health*, 24(1), 34-38.
- Piotrowski, C. (2017b). The linchpin on the future of projective techniques. *Journal of Proj. Psychol. & Mental Health*, 24(2), 71-73.
- Piotrowski, C. (2015a). Projective techniques usage worldwide: A review of applied settings 1995-2015. *Journal of the Indian Academy of Applied Psychology*, 41(3), 9-19.
- Piotrowski, C. (2015b). Clinical instruction on projective techniques in the USA: A review of academic training settings 1995-2014. *Journal of Proj. Psycho. & Mental Health*, 22(2), 83-92.
- Piotrowski, C., Keller, J.W., & Ogawa, T. (1993). Projective techniques: An international perspective. *Psychological Reports*, 72, 179-182.
- Ready, R.E., & Veague, H.B. (2014). Training in psychological assessment: Current practices of clinical psychology programs. *Professional Psychology: Research and Practice*, 45(4), 278-282.
- Ready, R.E., Santorelli, G.D., Lundquist, T.S., & Romano, F.M. (2016). Psychology internship directors' perceptions of pre-internship training preparation in assessment. *North American Journal of Psychology*, 18(2), 317-334.
- Stedman, J.M., McGeary, C.A., & Essery, J. (2017). Current patterns of training in personality assessment during internship. *Journal of Clinical Psychology*, 73, in press.
- Teglasi, H. (2013). The scientific status of projective techniques as performance measures of personality. In D.H. Saklofske, C.R. Reynolds, & V.L. Schwane (Eds.), *Oxford handbook of child psychological assessment* (pp. 113-128). New York: Oxford University Press.
- Teglasi, H. (2010). *Essentials of TAT and other storytelling assessments* (2nd ed.). New York: Wiley.
- Wechsler, S.M., Oakland, T., Leon, C., et al. (2014). Test development and use in five Iberian Latin American countries. *International Journal of Psychology*, 49(4), 233-239.
- Wood, J.M., Nezworski, M.T., Lilienfeld, S.O., & Garb, H.N. (2011). *What's wrong with the Rorschach: Science confronts the controversial inkblot test*. New York: Wiley.
- Wright, C.V., et al. (2017). Assessment practices of professional psychologists: Results of a national survey. *Professional Psychology: Research and Practice*, 48(2), 73-78.

Chris Piotrowski, PhD
 Senior Editor
 University of West Florida USA
 Email: cpiotrowski@uwf.edu