

Prior to beginning work on this discussion, read Chapter 11 in the text, the articles by Baez (2013), Hogan, Barrett, and Hogan (2007), Morgeson, Campion, and Dipboye (2007), Peterson, Griffith, Isaacson, O'Connell, and Mangos (2011), and the Maximizing Human Potential Within Organizations (Links to an external site.) and Building Better Organizations (Links to an external site.) brochures on the Society for Industrial and Organizational Psychology (SIOP) website.

Evaluate the MMP1-2-RF Police Candidate Interpretive Reports for Mr. C (Links to an external site.) and Ms. D. (Links to an external site.) For this discussion, you will take on the role of an industrial-organizational psychologist recently awarded a contract to evaluate potential police candidates. The purpose of the evaluations is to determine the psychological capability of the applicants to be certified as police officers in your state. The applicants you are examining are applying for certification and will be vested with a position of public trust. If certified as police officers, the individuals will likely be required at some future time to exercise significant physical strength and undergo high emotional stress. As the examining psychologist, you are required to comment on the applicants' social comprehension, judgment, impulse control, potential for violence, and/or any psychological traits that might render her or him psychologically at risk to be certified. The state requires that each applicant's examination include the following elements:

Interview and History: The psychologist must personally interview the applicant and provide a summary of the applicant's personal, educational, employment, and criminal history.

Required Personality Test: The applicant shall be administered any current standard form of the Minnesota Multiphasic Personality Inventory-2 (MMPI-2) by the licensed psychologist who interviewed the individual, or by a paraprofessional employed by and under the direct control and supervision of that licensed psychologist.

Other Testing Methods: If (after conducting the required test) the licensed psychologist is unable to certify the applicant's psychological capability or risk to exercise appropriate judgment and restraint to be certified as a police officer, the psychologist is directed to personally employ whatever other psychological measuring instrument(s) and/or technique(s) deemed necessary to form her or his professional opinion. The use of any such instrument(s) and/or technique(s) requires a full and complete written explanation to the commission.

For the purposes of this discussion, assume the interview and history information reported to you by Mr. C. and Ms. D. is unremarkable and that neither candidate communicated anything to you during the interview that raised concerns about her or his capabilities to exercise appropriate judgment and restraint to be certified as a police officer. Review the MMP1-2-RF Police Candidate Interpretive Reports for Mr. C. (Links to an external site.) and Ms. D. (Links to an external site.) and evaluate the professional interpretation of this testing and assessment data from an ethical perspective. In your initial post, communicate your conclusions about Mr. C. and Ms. D., either recommending certification or communicating reservations. After you have made your decision, begin the section on each candidate with one of the following statements, identifying each candidate by name.

To recommend certification: I have examined [insert applicant's name], and it is my professional opinion that this person is psychologically capable of exercising appropriate judgment and restraint to be certified as a police officer.

Follow the above statement with a one-paragraph rationale for your conclusion based on the available test results.

Follow the rationale with a brief comparison of any additional assessment instruments you might consider administering beyond the MMPI-2-RF. Debate the pros and cons of the potential use of other assessments. Explain any ethical implications that may arise from the interpretation of this data.

To communicate reservations: I have examined [insert applicant's name], and it is my professional opinion that this person is psychologically at risk for exercising appropriate judgment and restraint to be certified as a police officer.

Follow the statement with a one-paragraph rationale for your conclusion based on the available test results.

Follow the rationale with a brief comparison of any additional assessment instruments you recommend administering beyond the MMPI-2-RF. Debate the pros and cons of using other assessments. Explain any ethical implications that may arise from the interpretation of this data.

What Is SIOP?

The Society for Industrial and Organizational Psychology (SIOP) is an international professional organization with an annual membership of more than 8,000 industrial-organizational (I-O) psychologists. SIOP's mission is to enhance human well-being and performance in organizational and work settings by promoting the science, practice, and teaching of I-O psychology. The Society provides a platform for scientists, academics, consultants, and practitioners to collaborate, implement, and evaluate cutting-edge solutions to workplace challenges.

SIOP Member Snapshot

- With an annual membership of more than 8,000 across the United States and the world, SIOP is a diverse group.
- SIOP members are experts, most of them possessing doctoral degrees.
- SIOP is Division 14 of the American Psychological Association and an organizational affiliate of the Association for Psychological Science.
- SIOP members have testified before Congress, written federal legislation, and have served as consultants and experts for numerous companies from small private firms to large multinational corporations.



Online Resources

More information about I-O psychology can be found on the SIOP website.

www.siop.org

On SIOP's website you will find information about the Society's annual conference and Leading Edge Consortium as well as an array of free resources, including webinars, white papers, continuing education credit listings, graduate training program information, job listings, a book store, and pages devoted specifically to student, professional, educator, and media needs.



SOCIETY for
INDUSTRIAL and
ORGANIZATIONAL
PSYCHOLOGY

SCIENCE FOR A SMARTER WORKPLACE

Society for Industrial
and Organizational Psychology
440 E. Poe Rd. Suite 101
Bowling Green, OH 43402
www.siop.org

Industrial and Organizational
Psychology:

**Science for a
Smarter Workplace**



SOCIETY for INDUSTRIAL and
ORGANIZATIONAL PSYCHOLOGY
SCIENCE FOR A SMARTER WORKPLACE

A majority of our lives are spent working.

What happens at work can have a huge impact on people and the organizations for which they are employed. Industrial-organizational (I-O) psychologists use psychology to improve the workplace for both employees and employers.

What Is I-O Psychology?

I-O psychology is the scientific study of working and the application of that science to workplace issues facing individuals, organizations, and teams. It lies at the crossroads of business and psychology. Just as finance professionals advise organizations on how to maximize their financial capital, I-O psychologists teach organizations how to maximize their human capital.

I-O psychologists apply the scientific method to investigate issues of critical relevance to individuals, businesses, and society. Through these efforts, they advance the science of work, leading to smarter and more innovative evidence-based HR and organizational tools and solutions. I-O psychologists work toward—

Maximizing Employee Potential Through:

- **Testing:** test development, interpretation, and analysis, including tests of job knowledge, skills, reasoning, personality, and physical abilities; assessment centers; multimedia testing (Web-based, video, etc.).

- **Selection and Promotion:** recruiting, hiring, structured interviews, succession planning, performance appraisal/management.

- **Performance Evaluations/Assessments:** design of job performance measurement systems for feedback and performance improvement.

- **Training and Development:** computer-based learning; executive coaching, management development, mentoring, and leadership; competency modeling; team design and training; measuring effectiveness.

- **Employee Attitudes and Satisfaction:** involvement and empowerment; retention; job satisfaction; burnout, conflict, and stress; aging and retirement; gender issues; resignation and voluntary turnover.

- **Employee Motivation:** factors that motivate employees to perform effectively.

- **Compensation and Benefits:** pay, perks, rewards, and recognition.

- **Employee Behavior:** harassment, absenteeism, and discipline.

Maximizing Organizational Potential Through:

- **Change Management:** mergers and acquisitions, group processes, process reengineering, productivity and quality improvement, strategic planning.

- **Job Design and Evaluation**

- **Organizational Structure:** workforce planning; downsizing and rightsizing.

- **Cross-Cultural and Diversity Issues:** global, cross-cultural and diversity issues.

- **Impact of Technology in the Workplace**
- **Customer Service Issues**

- **Strategic Planning:** understanding and advising on how to use human capital to achieve organizational goals.

- **Workplace Health:** ergonomics, human factors, safety, stress, workplace violence.

- **Legal:** analysis of issues and expert testimony on EEO/AA, OSHA, and other issues; discrimination; jury decision processes.

- **Work-Life Programs:** flexible work arrangements, work-life balance, working parents, and telecommuting.

Job Titles of I-O Psychologists:

President, Vice President, Director, Staff Member, Manager, Principal, Consultant of:

Organizational Development; Organization Effectiveness; Talent Management; Management Development; Workforce Insights; Human Resources; Human Resources Research; Employee Relations; Training and Development; Leadership Development; Selection Systems; Evaluation and Assessment; Testing Programs; Leadership Research; Applied Behavioral Research; Optimization; Leadership, Learning, and Performance; Career Planning

Full, Associate, Assistant Professor of:

Industrial-Organizational Psychology, Industrial and Organizational Psychology, Organizational Psychology, Industrial Psychology, Work Psychology, (General) Psychology, Management, Organizational Behavior

Other titles include:

HR Practice Leader, Behavioral Analyst, Behavioral Scientist, Talent Management Specialist, Executive Coach, Leadership Coach, Employment Testing Professional, Testing Specialist, Assessment and Selection Specialist, Employment Law Expert, Research Analyst (I-O Psychology)



Resources for Students and Educators

The SIOP website, www.siop.org, contains information specifically for students and educators under the "Students" and "Educators" tabs, such as:

- Graduate Training Program (GTP) Information with a listing and search of I-O graduate programs, internship and program ranking information, and a helpful essay on how to choose a graduate program
- Guidelines for education and training at the doctoral and master's levels
- I-O job and internship openings
- Free educational webinars
- *What's in a Name?* A document describing I-O career and job titles
- *Psychologists at Work: What do I-O Psychologists Really Do?* A series of profiles of SIOP members' careers
- SIOP member and salary surveys
- PowerPoint "lectures" in introductory I-O psychology

http://www.onetonline.org/link_summary/19-3032.00#WagesEmployment



About SIOP

The Society for Industrial and Organizational Psychology (SIOP) is an international professional organization with an annual membership of more than 9,000 industrial-organizational (I-O) psychologists.

SIOP's mission is to enhance human well-being and performance in organizing the science, and work settings by promoting the science, practice, and teaching of I-O psychology. More interesting facts about SIOP include:

- With members in various employment settings (academia, government, corporations, etc.) across the United States and the world, SIOP is diverse
- SIOP members are experts, most of them possessing doctoral degrees
- SIOP members have testified before Congress, written federal legislation, and have served as consultants and experts for numerous companies from small private firms to large multinational corporations
- SIOP is Division 14 of the American Psychological Association and an organizational affiliate of the Association for Psychological Science



SOCIETY for
INDUSTRIAL and
ORGANIZATIONAL
PSYCHOLOGY

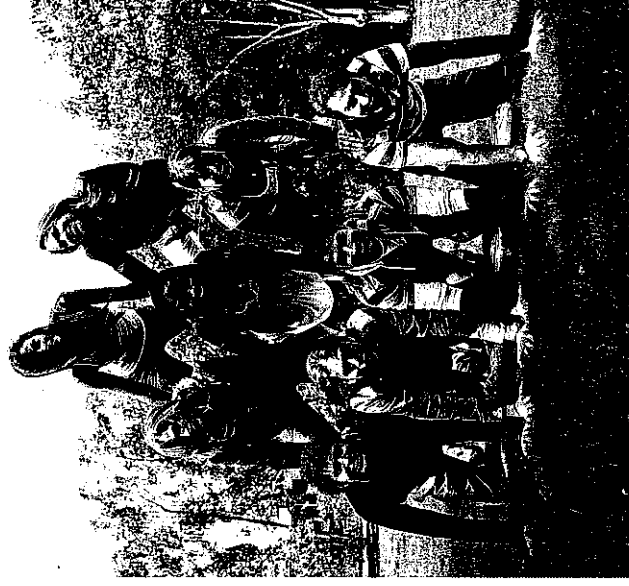
SCIENCE FOR A SMARTER WORKPLACE

440 East Poe Rd. Suite 101
Bowling Green, OH 43402
419-353-0032
siop@siop.org

www.SIOP.org

Join one of the highest paid fields in psychology!*

Build Better Organizations: Become an Industrial-Organizational Psychologist



SOCIETY for
INDUSTRIAL and
ORGANIZATIONAL
PSYCHOLOGY

SCIENCE FOR A SMARTER WORKPLACE

What Is Industrial-Organizational Psychology?

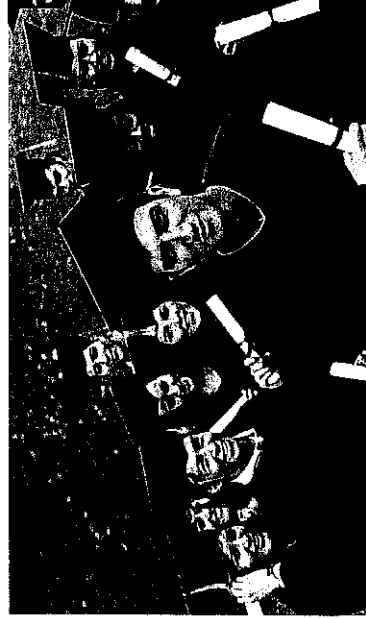
Industrial-organizational (I-O) psychology is the scientific study of working and the application of that science to workplace issues facing individuals, teams, and organizations.

The scientific method is applied to investigate issues of critical relevance to employees, businesses, and society, such as testing/assessment, staffing, performance management, employee stress/well-being, teamwork, organizational development and change management, work-life balance, and leadership. Through these efforts, I-O psychologists advance the science of work, leading to smarter and more innovative evidence-based tools and solutions.

What Do I-O Psychologists Do?

I-O psychologists work in many areas that contribute to organizations' success, effectiveness, and bottom line. They:

- Recruit employees that best fit jobs and organizations
- Develop and retain the best employees
- Create fair, legal, and efficient hiring practices



- Develop performance management systems
- Minimize absenteeism and counterproductive behaviors
- Eliminate harassment/discrimination
- Maximize productivity and efficiency of employees
- Foster teamwork and develop teams
- Increase motivation and dedication
- Identify and train leaders
- Develop compensation and benefits practices
- Consult during downsizing, rightsizing, mergers/ acquisitions
- Understand and manage organizational culture and climate

Job Titles for I-O Psychologists

President, Vice President, Director, Manager, Principal, Consultant of:

Organizational Development, Organizational Effectiveness, Organizational Capability, Talent Management, Management Development, Workforce Insights, Human Resources, Human Resources Research, Employment Relations, Training and Development, Leadership Development, Selection Systems, Evaluation and Assessment, Testing, Leadership Research, Assessment, Applied Behavioral Research

Full, Associate, Assistant Professor of:

Industrial-Organizational Psychology, Industrial and Organizational Psychology, Organizational Psychology, Industrial Psychology, Work Psychology, (General) Psychology, Management, Organizational Behavior, Industrial Relations

Other titles include:

HR Practice Leader, Behavioral Analyst, Behavioral Scientist, Talent Management Specialist, HR Organizational Development Specialist, Executive Coach, Career Coach, Testing Specialist, Leadership Coach

Top 3 Reasons to Become an I-O Psychologist:

- Benefit from the variety, independence, and flexibility of being a highly paid professional in a growing field.
- Develop a deep understanding of people and how they behave.
- Learn how to help both businesses and employees achieve their full potential.



CHAPTER 11

Industrial, Occupational, and Career Assessment

TOPIC 11A Industrial and Organizational Assessment

11.1 The Role of Testing in Personnel Selection

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch11lev1sec1#ch11lev1sec1>)

11.2 Autobiographical Data

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch11lev1sec2#ch11lev1sec2>)

11.3 The Employment Interview

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch11lev1sec3#ch11lev1sec3>)

11.4 Cognitive Ability Tests

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch11lev1sec4#ch11lev1sec4>)

11.5 Personality Tests

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch11lev1sec5#ch11lev1sec5>)

11.6 Paper-and-Pencil Integrity Tests

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch11lev1sec6#ch11lev1sec6>)

11.7 Work Sample and Situational Exercises

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch11lev1sec7#ch11lev1sec7>)

11.8 Appraisal of Work Performance

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch11lev1sec8#ch11lev1sec8>)

11.9 Approaches to Performance Appraisal

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch11lev1sec9#ch11lev1sec9>)

11.10 Sources of Error in Performance Appraisal

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch11lev1sec10#ch11lev1sec10>)

In this chapter we explore the specialized applications of testing within two distinctive environments—occupational settings and vocational settings. Although disparate in many respects, these two fields of assessment share essential features. For example, legal guidelines exert a powerful and constraining influence upon the practice of testing in both arenas. Moreover, issues of empirical validation of methods are especially pertinent in occupational and areas of practice. In **Topic 11A** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch11#ch11box1>), Industrial and Organizational Assessment, we review the role of psychological tests in making decisions about personnel such as hiring, placement, promotion, and evaluation. In **Topic 11B**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch11lev1sec10#ch11box2>) , Assessment for Career Development in a Global Economy, we analyze the unique challenges encountered by vocational psychologists who provide career guidance and assessment. Of course, relevant tests are surveyed and catalogued throughout. But more important, we focus upon the special issues and challenges encountered within these distinctive milieus.

Industrial and organizational psychology (I/O psychology) is the subspecialty of psychology that deals with behavior in work situations (**Borman, Ilgen, Klimoski, & Weiner, 2003** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib182>)). In its broadest sense, I/O psychology includes diverse applications in business, advertising, and the military. For example, corporations typically consult I/O psychologists to help design and evaluate hiring procedures; businesses may ask I/O psychologists to appraise the effectiveness of advertising; and military leaders rely heavily upon I/O psychologists in the testing and placement of recruits. Psychological testing in the service of decision making about personnel is, thus, a prominent focus of this profession. Of course, specialists in I/O psychology possess broad skills and often handle many corporate responsibilities not previously mentioned. Nonetheless, there is no denying the centrality of assessment to their profession.

We begin our review of assessment in the occupational arena by surveying the role of testing in personnel selection. This is followed by a discussion of ways that psychological measurement is used in the appraisal of work performance.

11.1 THE ROLE OF TESTING IN PERSONNEL SELECTION

Complexities of Personnel Selection

Based upon the assumption that psychological tests and assessments can provide valuable information about potential job performance, many businesses, corporations, and military settings have used test scores and assessment results for personnel selection. As Guion (1998

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib671>)) has noted, I/O research on personnel selection has emphasized criterion-related validity as opposed to content or construct validity. These other approaches to validity are certainly relevant but usually take a back seat to criterion-related validity, which preaches that current assessment results must predict the future criterion of job performance.

From the standpoint of criterion-related validity, the logic of personnel selection is seductively simple. Whether in a large corporation or a small business, those who select employees should use tests or assessments that have documented, strong correlations with the criterion of job performance, and then hire the individuals who obtain the highest test scores or show the strongest assessment results. What could be simpler than that?

Unfortunately, the real-world application of employment selection procedures is fraught with psychometric complexities and legal pitfalls. The psychometric intricacies arise, in large measure, from the fact that job behavior is rarely simple, unidimensional behavior. There are some exceptions (such as assembly-line production) but the general rule in our postindustrial society is that job behavior is complex, multidimensional behavior. Even jobs that seem simple may be highly complex. For example, consider what is required for effective performance in the delivery of the U.S. mail. The individual who delivers your mail six days a week must do more than merely place it in your mailbox. He or she must accurately sort mail on the run, interpret and enforce government regulations about package size, manage pesky and even dangerous animals, recognize and avoid physical dangers, and exercise effective interpersonal skills in dealing with the public, to cite just a few of the complexities of this position.

Personnel selection is, therefore, a fuzzy, conditional, and uncertain task. Guion (1991) has highlighted the difficulty in predicting complex behavior from simple tests. For one thing, complex behavior is, in part, a function of the situation. This means that even an optimal selection approach may not be valid for all candidates. Quite clearly, personnel selection is not a simple matter of administering tests and consulting cutoff scores.

We must also acknowledge the profound impact of legal and regulatory edicts upon I/O testing practices. Given that such practices may have weighty consequences—determining who is hired or promoted, for example—it is not surprising to learn that I/O testing practices are rigorously constrained by legal precedents and regulatory mandates. These topics are reviewed in **Topic 12A**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch12#ch12box1>) , Psychological Testing and the Law.

Approaches to Personnel Selection

Acknowledging that the interview is a widely used form of personnel assessment, it is safe to conclude that psychological assessment is almost a universal practice in hiring decisions. Even by a narrow definition that includes only paper-and-pencil measures, at least two-thirds of the companies in the United States engage in personnel testing (Schmitt & Robertson, 1990

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1471>)). For purposes of personnel selection, the I/O psychologist may recommend one or more of the following:

- Autobiographical data
- Employment interview
- Cognitive ability tests
- Personality, temperament, and motivation tests
- Paper-and-pencil integrity tests
- Sensory, physical, and dexterity tests
- Work sample and situational tests

We turn now to a brief survey of typical tests and assessment approaches within each of these categories. We close this topic with a discussion of legal issues in personnel testing.

11.2 AUTOBIOGRAPHICAL DATA

According to Owens (1976

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1262>), application forms that request personal and work history as well as demographic data such as age and marital status have been used in industry since at least 1894. Objective or scorable autobiographical data—sometimes called **biodata** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm01#bm01gloss36>)—are typically secured by means of a structured form variously referred to as a biographical information blank, biographical data form, application blank, interview guide, individual background survey, or similar device. Although the lay public may not recognize these devices as true tests with predictive power, I/O psychologists have known for some time that biodata furnish an exceptionally powerful basis for the prediction of employee performance (Cascio, 1976

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib278>) ; Ghiselli, 1966 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib584>) ; Hunter & Hunter, 1984). An important milestone in the biodata approach is the publication of the *Biodata Handbook*, a thorough survey of the use of biographical information in selection and the prediction of performance (Stokes, Mumford, & Owens, 1994

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1581>)).

The rationale for the biodata approach is that future work-related behavior can be predicted from past choices and accomplishments. Biodata have predictive power because certain character traits that are essential for success also are stable and enduring. The consistently ambitious youth with accolades and accomplishments in high school is likely to continue this pattern into adulthood. Thus, the job applicant who served as editor of the high school newspaper—and who answers a biodata item to this effect—is probably a better candidate for corporate management than the applicant who reports no extracurricular activities on a biodata form.

The Nature of Biodata

Biodata items usually call for “factual” data; however, items that tap attitudes, feelings, and value judgments are sometimes included. Except for demographic data such as age and marital status, biodata items always refer to past accomplishments and events. Some examples of biodata items are listed in **Table 11.1** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch11lev1sec2#ch11tab1>).

Once biodata are collected, the I/O psychologist must devise a means for predicting job performance from this information. The most common strategy is a form of empirical keying not unlike that used in personality testing. From a large sample of workers who are already hired, the I/O psychologist designates a successful group and an unsuccessful group, based on performance, tenure, salary, or supervisor ratings. Individual biodata items are then contrasted for these two groups to determine which items most accurately discriminate between successful and unsuccessful workers. Items that are strongly discriminative are assigned large weights in the scoring scheme. New applicants who respond to items in the keyed direction, therefore, receive high scores on the biodata instrument and are predicted to succeed. Cross validation of the scoring scheme on a second sample of successful and unsuccessful workers is a crucial step in guaranteeing the validity of the biodata selection method. Readers who wish to pursue the details of empirical scoring methods for biodata instruments should consult Murphy and Davidshofer (2004 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1190>)), Mount, Witt, and Barrick (2000

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1184>)), and Stokes and Cooper (2001 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1579>)).

TABLE 11.1 Examples of Biodata Questions

How long have you lived at your present address?

What is your highest educational degree?

How old were you when you obtained your first paying job?

How many books (not work related) did you read last month?

At what age did you get your driver's license?

In high school, did you hold a class office?

How punctual are you in arriving at work?

What job do you think you will hold in 10 years?

How many hours do you watch television in a typical week?

Have you ever been fired from a job?

How many hours a week do you spend on hobbies?

How many job projects did you manage in the last year?

In college, did you participate in a sports team?

How many hours per month do you volunteer?

What is your attitude toward others who use marijuana?

The Validity of Biodata

The validity of biodata has been surveyed by several reviewers, with generally positive findings (Breauigh, 2009 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib213>) ; Stokes et al., 1994 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1581>) ; Stokes & Cooper, 2004 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1580>)). An early study by Cascio (1976 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib278>)) is typical of the findings. He used a very simple biodata instrument—a weighted combination of 10 application blank items—to predict turnover for female clerical personnel in a medium-sized insurance company. The cross-validated correlations between biodata score and length of tenure were .58 for minorities and .56 for nonminorities.¹ (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch11lev1sec2#ch11fn01>) Drakeley et al. (1988 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib434>)) compared biodata and cognitive ability tests as predictors of training success. Biodata scores possessed the same predictive validity as the cognitive tests. Furthermore, when added to the regression equation, the biodata information improved the predictive accuracy of the cognitive tests.

In an extensive research survey, Reilly and Chao (1982 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1347>)) compared eight selection procedures as to validity and adverse impact on minorities. The procedures were biodata, peer evaluation, interviews, self-assessments, reference checks, academic achievement, expert judgment, and projective techniques. Noting that properly standardized ability tests provide the fairest and most valid selection procedure, Reilly and Chao (1982

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1347>)) concluded that only biodata and peer evaluations had validities substantially equal to those of standardized tests. For example, in the prediction of sales productivity, the average validity coefficient of biodata was a very healthy .62.

Certain cautions need to be mentioned with respect to biodata approaches in personnel selection. Employers may be prohibited by law from asking questions about age, race, sex, religion, and other personal issues—even when such biodata can be shown empirically to predict job performance. Also, even though the incidence of faking is very low, there is no doubt that shrewd respondents can falsify results in a favorable direction. For example, Schmitt and Kuncce (2002 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1470>)) addressed the concern that some examinees might distort their answers to biodata items in a socially desirable direction. These researchers compared the scores obtained when examinees were asked to elaborate their biodata responses versus when they were not. Requiring elaborated answers reduced the scores on biodata items; that is, it appears that respondents were more truthful when asked to provide corroborating details to their written responses.

Recently, Levashina, Morgeson, and Campion (2012 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib971>)) proved the same point in a large scale, high-stakes selection project with 16,304 applicants for employment. Biodata constituted a significant portion of the selection procedure. The researchers used the response elaboration technique (RET), which obliges job applicants to provide written elaborations of their responses. Perhaps an example will help. A naked, unadorned biodata question might ask:

- How many times in the last 12 months did you develop novel solutions to a work problem in your area of responsibility?

Most likely, a higher number would indicate greater creativity and empirically predict superior work productivity. The score on this item would be combined with others to produce an overall biodata score used in personnel selection. But notice that nothing prevents the respondent from exaggeration or outright lying. Now, consider the original question with the addition of response elaboration:

- How many times in the last 12 months did you develop novel solutions to a work problem in your area of responsibility?
- For each circumstance, please provide specific details as to the problem and your solution.

Levashina et al. (2012

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib971>)) found that using the RET technique produced more honest and realistic biodata scores. Further, for those items possessing the potential for external verification, responses were even more realistic. The researchers conclude that RET decreases faking because it increases accountability.

As with any measurement instrument, biodata items will need periodic restandardization. Finally, a potential drawback to the biodata approach is that, by its nature, this method captures the organizational status quo and might, therefore, squelch innovation. Becker and Colquitt (1992) discuss precautions in the development of biodata forms.

The use of biodata in personnel selection appears to be on the rise. Some corporations rely on biodata almost to the exclusion of other approaches in screening applicants. The software giant Google is a case in point. In years past, the company used traditional methods such as hiring candidates from top schools

who earned the best grades. But that tactic now is used rarely in industry. Instead, many corporations like Google are moving toward automated systems that collect biodata from the many thousands of applicants processed each year. Using online surveys, these companies ask applicants to provide personal details about accomplishments, attitudes, and behaviors as far back as high school. Questions can be quite detailed, such as whether the applicant has ever published a book, received a patent, or started a club. Formulas are then used to compute a score from 0 to 100, designed to predict the degree to fit with corporate culture (**Ottinger & Kurzon, 2007**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1261>)). The system works well for Google, which claims to have only a 4 percent turnover rate.

There is little doubt, then, that purely objective biodata information can predict aspects of job performance with fair accuracy. However, employers are perhaps more likely to rely upon subjective information such as interview impressions when making decisions about hiring. We turn now to research on the validity of the employment interview in the selection process.

¹The curious reader may wish to know which 10 biodata items could possess such predictive power. The items were age, marital status, children's age, education, tenure on previous job, previous salary, friend or relative in company, location of residence, home ownership, and length of time at present address. Unfortunately, Cascio (1976

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib278>)) does not reveal the relative weights or direction of scoring for the items.

11.3 THE EMPLOYMENT INTERVIEW

The employment interview is usually only one part of the evaluation process, but many administrators regard it as the vital make-or-break component of hiring. It is not unusual for companies to interview from 5 to 20 individuals for each person hired! Considering the importance of the interview and its huge costs to industry and the professions, it is not surprising to learn that thousands of studies address the reliability and validity of the interview. We can only highlight a few trends here; more detailed reviews can be found in Conway, Jako, and Goodman (1995

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib348>), Huffcutt (2007 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib792>), Guion (1998 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib671>), and Schmidt and Zimmerman (2004 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1466>)).

Early studies of interview reliability were quite sobering. In various studies and reviews, reliability was typically assessed by correlating evaluations of different interviewers who had access to the same job candidates (Wagner, 1949

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1703>); Ulrich & Trumbo, 1965 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1678>)).

The interrater reliability from dozens of these early studies was typically in the mid-.50s, much too low to provide accurate assessments of job candidates. This research also revealed that interviewers were prone to halo bias and other distorting influences upon their perceptions of candidates. *Halo bias*—discussed in the next topic—is the tendency to rate a candidate high or low on all dimensions because of a global impression.

Later, researchers discovered that interview reliability could be increased substantially if the interview was jointly conducted by a panel instead of a single interviewer (Landy, 1996

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib947>). In addition, structured interviews in which each candidate was asked the same questions by each interviewer also proved to be much more reliable than unstructured interviews (Borman, Hanson, & Hedge, 1997; Campion, Pursell, & Brown, 1988

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib264>). In these studies, reliabilities in the .70s and higher were found.

Research on validity of the interview has followed the same evolutionary course noted for reliability: Early research that examined unstructured interviews was quite pessimistic, while later research using structured approaches produced more promising findings. In these studies, interview validity was typically assessed by correlating interview judgments with some measure of on-the-job performance. Early studies of interview validity yielded almost uniformly dismal results, with typical validity coefficients hovering in the mid-.20s (Arvey & Campion, 1982

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib62>)).

Mindful that interviews are seldom used in isolation, early researchers also investigated incremental validity, which is the potential increase in validity when the interview is used in conjunction with other information. These studies were predicated on the optimistic assumption that the interview would contribute positively to candidate evaluation when used alongside objective test scores and background data. Unfortunately, the initial findings were almost entirely unsupportive (Landy, 1996

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib947>)).

In some instances, attempts to prove incremental validity of the interview demonstrated just the opposite, what might be called decremental validity. For example, Kelly and Fiske (1951 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib875>)) established that interview information actually decreased the validity of graduate student evaluations. In this early and classic study, the task was to predict the academic performance of more than 500 graduate students in psychology. Various combinations of credentials (a form of biodata), objective test scores, and interview were used as the basis for clinical predictions of academic performance. The validity coefficients are reported in **Table 11.2**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch11lev1sec3#ch11tab2>). The reader will notice that credentials alone provided a much better basis for prediction than credentials plus a one-hour interview. The best predictions were based upon credentials and objective test scores; adding a two-hour interview to this information actually decreased the accuracy of predictions. These findings highlighted the superiority of actuarial prediction (based on empirically derived formulas) over clinical prediction (based on subjective impressions). We pursue the actuarial versus clinical debate in the last chapter of this text.

Studies using carefully structured interviews, including situational interviews, provide a more positive picture of interview validity (Borman, Hanson, & Hedge, 1997; Maurer & Fay, 1988; Schmitt & Robertson, 1990 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1471>)). When the findings are corrected for restriction of range and unreliability of job performance ratings, the mean validity coefficient for structured interviews turns out to be an impressive .63 (Wiesner & Cronshaw, 1988 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1755>)). A meta-analysis by Conway, Jako, and Goodman (1995

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib348>)) concluded that the upper limit for the validity coefficient of structured interviews was .67, whereas for unstructured interviews the validity coefficient was only .34. Additional reasons for preferring structured interviews include their legal defensibility in the event of litigation (Williamson, Campion, Malo, and others, 1997 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1761>)) and, surprisingly, their minimal bias across different racial groups of applicants (Huffcutt & Roth, 1998 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib793>)).

TABLE 11.2 Validity Coefficients for Ratings Based on Various Combinations of Information

<i>Basis for Rating</i>	<i>Correlation with Academic Performance</i>
Credentials alone	.26
Credentials and one-hour interview	.13
Credentials and objective test scores	.36
Credentials, test scores, and two-hour interview	.32

Source: Based on data in Kelly, E. L., & Fiske, D. W. (1951

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib875>)). *The prediction of performance in clinical psychology*. Ann Arbor: University of Michigan Press.

In order to reach acceptable levels of reliability and validity, structured interviews must be designed with painstaking care. Consider the protocol used by Motowidlo et al. (1992

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1182>)) in their research on structured interviews for management and marketing positions in eight telecommunications companies. Their interview format was based upon a careful analysis of critical incidents in marketing and management. Prospective employees were asked a set of standard questions about how they had

handled past situations similar to these critical incidents. Interviewers were trained to ask discretionary probing questions for details about how the applicants handled these situations. Throughout, the interviewers took copious notes. Applicants were then rated on scales anchored with behavioral illustrations. Finally, these ratings were combined to yield a total interview score used in selection decisions.

In summary, under carefully designed conditions, the interview can provide a reliable and valid basis for personnel selection. However, as noted by Schmitt and Robertson (1990 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1471>)), the prerequisite conditions for interview validity are not always available. Guion (1998 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib671>)) has expressed the same point:

A large body of research on interviewing has, in my opinion, given too little practical information about how to structure an interview, how to conduct it, and how to use it as an assessment device. I think I know from the research that (a) interviews can be valid, (b) for validity they require structuring and standardization, (c) that structure, like many other things, can be carried too far, (d) that without carefully planned structure (and maybe even with it) interviewers talk too much, and (e) that the interviews made routinely in nearly every organization could be vastly improved if interviewers were aware of and used these conclusions. There is more to be learned and applied. (p. 624)

The essential problem is that each interviewer may evaluate only a small number of applicants, so that standardization of interviewer ratings is not always realistic. While the interview is potentially valid as a selection technique, in its common, unstructured application there is probably substantial reason for concern.

Why are interviews used? If the typical, unstructured interview is so unreliable and ineffectual a basis for job candidate evaluation, why do administrators continue to value interviews so highly? In their review of the employment interview, Arvey and Campion (1982 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib62>)) outline several reasons for the persistence of the interview, including practical considerations such as the need to sell the candidate on the job, and social reasons such as the susceptibility of interviewers to the illusion of personal validity. Others have emphasized the importance of the interview for assessing a good fit between applicant and organization (Adams, Elacqua, & Colarelli, 1994 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib08>); Latham & Skarlicki, 1995 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib955>)).

It is difficult to imagine that most employers would ever eliminate entirely the interview from the screening and selection process. After all, the interview does serve the simple human need of meeting the persons who might be hired. However, based on 50 years worth of research, it is evident that biodata and objective tests often provide a more powerful basis for candidate evaluation and selection than unstructured interviews.

One interview component that has received recent attention is the impact of the handshake on subsequent ratings of job candidates. Stewart, Dustin, Barrick, and Darnold (2008 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1576>)) used simulated hiring interviews to investigate the commonly held conviction that a firm handshake bears a critical nonverbal influence on impressions formed during the employment interview. Briefly, 98 undergraduates underwent realistic job interviews during which their handshakes were surreptitiously

rated on 5-point scales for grip strength, completeness, duration, and vigor; degree of eye contact during the handshake also was rated. Independent ratings were completed at different times by five individuals involved in the process. Real human-resources professionals conducted the interviews and then offered simulated hiring recommendations. The professionals shook hands with the candidates but were not asked to provide handshake ratings because this would have cued them to the purposes of the study. This is the barest outline of this complex investigation. The big picture that emerged was that the quality of the handshake was positively related to hiring recommendations. Further, women benefited more than men from a strong handshake. The researchers conclude their study with these thoughts:

The handshake is thought to have originated in medieval Europe as a way for kings and knights to show that they did not intend to harm each other and possessed no concealed weapons (**Hall & Hall, 1983** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib686>)). The results presented in this study show that this age-old social custom has an important place in modern business interactions. Although the handshake may appear to be a business formality, it can indeed communicate critical information and influence interviewer assessments. (p. 1145)

Perhaps this study will provide an impetus for additional investigation of this important component of the job interview.

Barrick, Swider, and Stewart (2010

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib98>)) make the general case that initial impressions formed in the first few seconds or minutes of the employment interview significantly influence the final outcomes. They cite the social psychology literature to argue that initial impressions are nearly instinctual and based on evolutionary mechanisms that aid survival. Handshake, smile, grooming, manner of dress—the interviewer gauges these as favorable (or not) almost instantaneously. The purpose of their study was to examine whether these “fast and frugal” judgments formed in the first few seconds or minutes even before the “real” interview begins affect interview outcomes. Participants for their research were 189 undergraduate students in a program for professional accountants. The students were pre-interviewed for just 2-3 minutes by trained graduate students for purposes of rapport building, before a more thorough structured mock interview was conducted. After the brief pre-interview, the graduate interviewers filled out a short rating scale on liking for the candidate, the candidate’s competence, and perceived “personal” similarity. The interviewers then conducted a full structured interview and filled out ratings. Weeks after these mock interviews, participants engaged in real interviews with four major accounting firms (Deloitte Touche Tohmatsu, Ernst & Young, KPMG, and PricewaterhouseCoopers) to determine whether they would receive an offer of an internship. Just over half of the students received an offer. Candidates who made better first impressions during the initial pre-interview (that lasted just 2-3 minutes) received more internship offers ($r = .22$) and higher interviewer ratings ($r = .42$). In sum, initial impressions in the employment interview do matter.

11.4 COGNITIVE ABILITY TESTS

Cognitive ability can refer either to a general construct akin to intelligence or to a variety of specific constructs such as verbal skills, numerical ability, spatial perception, or perceptual speed (**Kline, 1999** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib902>)). Tests of general cognitive ability and measures of specific cognitive skills have many applications in personnel selection, evaluation, and screening. Such tests are quick, inexpensive, and easy to interpret.

A vast body of empirical research offers strong support for the validity of standardized cognitive ability tests in personnel selection. For example, Bertua, Anderson, and Salgado (2005) conducted a meta-analysis of 283 independent employee samples in the United Kingdom. They found that general mental ability as well as specific ability tests (verbal, numerical, perceptual, and spatial) are valid predictors of job performance and training success, with validity coefficients in the magnitude of .5 to .6. Surveying a large number of studies and employment settings, Kuncel and Hezlett (2010) summarized correlations between cognitive ability and seven measures of work performance as follows:

Job performance, high complexity:	.58
Job performance, medium complexity:	.52
Job performance, low complexity:	.40
Training success, civilian:	.55
Training success, military:	.62
Objective leader effectiveness:	.33
Creativity:	.37

Beyond a doubt, there is merit in the use of cognitive ability tests for personnel selection.

Even so, a significant concern with the use of cognitive ability tests for personnel selection is that these instruments may result in an adverse impact on minority groups. *Adverse impact* is a legal term (discussed later in this chapter) referring to the disproportionate selection of white candidates over minority candidates. Most authorities in personnel psychology recognize that cognitive tests play an essential role in applicant selection; nonetheless, these experts also affirm that cognitive tests provide maximum benefit (and minimum adverse impact) when combined with other approaches such as biodata. Selection decisions never should be made exclusively on the basis of cognitive test results (**Robertson & Smith, 2001**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1376>)).

An ongoing debate within I/O psychology is whether employment testing is best accomplished with highly specific ability tests or with measures of general cognitive ability. The weight of the evidence seems to support the conclusion that a general factor of intelligence (the so-called *g* factor) is usually a better predictor of training and job success than are scores on specific cognitive measures—even when several specific cognitive measures are used in combination. Of course, this conclusion runs counter to common sense and anecdotal evidence. For example, Kline (1993 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib901>)) offers the following vignette:

The point is that the *g* factors are important but so also are these other factors. For example, high *g* is necessary to be a good engineer and to be a good journalist. However for the former high spatial

ability is also required, a factor which confers little advantage on a journalist. For her or him, however, high verbal ability is obviously useful.

Curiously, empirical research provides only mixed support for this position (**Gottfredson, 1986; Larson & Wolfe, 1995** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib953>) ; **Ree, Earles, & Teachout, 1994**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1342>)). Although the topic continues to be debated, most studies support the primacy of *g* in personnel selection (**Borman et al., 1997; Schmidt, 2002**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1464>)). Perhaps the reason that *g* usually works better than specific cognitive factors in predicting job performance is that most jobs are factorially complex in their requirements, stereotypes notwithstanding (**Guion, 1998** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib671>)). For example, the successful engineer must explain his or her ideas to others and so needs verbal ability as well as spatial skills. Since measures of general cognitive ability tap many specific cognitive skills, a general test often predicts performance in complex jobs as well as, or better than, measures of specific skills.

Literally hundreds of cognitive ability tests are available for personnel selection, so it is not feasible to survey the entire range of instruments here. Instead, we will highlight three representative tests: one that measures general cognitive ability, a second that is germane to assessment of mechanical abilities, and a third that taps a highly specific facet of clerical work. The three instruments chosen for review—the Wonderlic Personnel Test-Revised, the Bennett Mechanical Comprehension Test, and the Minnesota Clerical Test—are merely exemplars of the hundreds of cognitive ability tests available for personnel selection. All three tests are often used in business settings and, therefore, worthy of specific mention. Representative cognitive ability tests encountered in personnel selection are listed in **Table 11.3**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch11lev1sec4#ch11tab3>) . Some classic viewpoints on cognitive ability testing for personnel selection are found in Ghiselli (**1966**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib584>)), Hunter and Hunter (1984), and Reilly and Chao (**1982**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1347>)). More recent discussion of this issue is provided by Borman et al. (1997), Guion (**1998**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib671>)), and Schmidt (**2002** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1464>)).

TABLE 11.3 Representative Cognitive Ability Tests Used in Personnel Selection

General Ability Tests

Shipley Institute of Living Scale
Wonderlic Personnel Test-Revised
Wesman Personnel Classification Test
Personnel Tests for Industry

Multiple Aptitude Test Batteries

General Aptitude Test Battery
Armed Services Vocational Aptitude Battery
Differential Aptitude Test
Employee Aptitude Survey

Mechanical Aptitude Tests

Bennett Mechanical Comprehension Test
Minnesota Spatial Relations Test
Revised Minnesota Paper Form Board Test
SRA Mechanical Aptitudes

Motor Ability Tests

Crawford Small Parts Dexterity Test
Purdue Pegboard
Hand-Tool Dexterity Test
Stromberg Dexterity Test

Clerical Tests

Minnesota Clerical Test
Clerical Abilities Battery
General Clerical Test
SRA Clerical Aptitudes

*Note: SRA denotes Science Research Associates. These tests are reviewed in the *Mental Measurements Yearbook* series.*

Wonderlic Personnel Test-Revised

Even though it is described as a personnel test, the Wonderlic Personnel Test-Revised (WPT-R) is really a group test of general mental ability (Hunter, 1989

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib800>) ; Wonderlic, 1983 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1783>)). The revised version was released in 2007 and is now named the Wonderlic Contemporary Cognitive Ability Test. We refer to it as the WPT-R here. What makes this instrument somewhat of an institution in personnel testing is its format (50 multiple-choice items), its brevity (a 12-minute time limit), and its numerous parallel forms (16 at last count). Item types on the Wonderlic are quite varied and include vocabulary, sentence rearrangement, arithmetic problem solving, logical induction, and interpretation of proverbs. The following items capture the flavor of the Wonderlic:

1. REGRESS is the opposite of
 - a. ingest
 - b. advance
 - c. close

- d. open
- 2. Two men buy a car which costs \$550; X pays \$50 more than Y. How much did X pay?
 - a. \$500
 - b. \$300
 - c. \$400
 - d. \$275
- 3. HEFT CLEFT—Do these words have
 - a. similar meaning
 - b. opposite meaning
 - c. neither similar nor opposite meaning

The reliability of the WPT-R is quite impressive, especially considering the brevity of the instrument. Internal consistency reliabilities typically reach .90, while alternative-form reliabilities usually exceed .90. Normative data are available on hundreds of thousands of adults and hundreds of occupations. Regarding validity, if the WPT-R is considered a brief test of general mental ability, the findings are quite positive (**Dodrill & Warner, 1988**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib425>)). For example, Dodrill (1981 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib424>)) reports a correlation of .91 between scores on the original WPT and scores on the WAIS. This correlation is as high as that found between any two mainstream tests of general intelligence. Bell, Matthews, Lassister, and Leverett (2002

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib126>)) reported strong congruence between the WPT and the Kaufman Adolescent and Adult Intelligence Test in a sample of adults. Hawkins, Faraone, Pepple, Seidman, and Tsuang (1990

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib720>)) report a similar correlation ($r = .92$) between WPT and WAIS-R IQ for 18 chronically ill psychiatric patients. However, in their study, one subject was unable to manage the format of the WPT, suggesting that severe visuospatial impairment can invalidate the test.

Another concern about the Wonderlic is that examinees whose native language is not English will be unfairly penalized on the test (**Belcher, 1992**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib123>)). The Wonderlic is a speeded test. In fact, it has such a heavy reliance on speed that points are added for subjects aged 30 and older to compensate for the well-known decrement in speed that accompanies normal aging. However, no accommodation is made for nonnative English speakers who might also perform more slowly. One solution to the various issues of fairness cited would be to provide norms for untimed performance on the Wonderlic. However, the publishers have resisted this suggestion.

Bennett Mechanical Comprehension Test

In many trades and occupations, the understanding of mechanical principles is a prerequisite to successful performance. Automotive mechanics, plumbers, mechanical engineers, trade school applicants, and persons in many other “hands-on” vocations need to comprehend basic mechanical principles in order to succeed in their fields. In these cases, a useful instrument for occupational testing is the Bennett Mechanical Comprehension Test (BMCT). This test consists of pictures about which the examinee must answer straightforward questions. The situations depicted emphasize basic mechanical principles that might be encountered in everyday life. For example, a series of belts and flywheels might be depicted, and the examinee would be asked to discern the relative revolutions per minute of two flywheels. The test includes two equivalent forms (S and T).

The BMCT has been widely used since World War II for military and civilian testing, so an extensive body of technical and validity data exist for this instrument. Split-half reliability coefficients range from the .80s to the low .90s. Comprehensive normative data are provided for several groups. Based on a huge body of earlier research, the concurrent and predictive validity of the BMCT appear to be well established (**Wing, 1992**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1769>)). For example, in one study with 175 employees, the correlation between the BMCT and the DAT Mechanical Reasoning subtest was an impressive .80. An intriguing finding is that the test proved to be one of the best predictors of pilot success during World War II (**Ghiselli, 1966** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib584>)).

In spite of its psychometric excellence, the BMCT is in need of modernization. The test looks old and many items are dated. By contemporary standards, some BMCT items are sexist or potentially offensive to minorities (**Wing, 1992**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1769>)). The problem with dated and offensive test items is that they can subtly bias test scores. Modernization of the BMCT would be a straightforward project that could increase the acceptability of the test to women and minorities while simultaneously preserving its psychometric excellence.

Minnesota Clerical Test

The Minnesota Clerical Test (MCT), which purports to measure perceptual speed and accuracy relevant to clerical work, has remained essentially unchanged in format since its introduction in 1931, although the norms have undergone several revisions, most recently in 1979 (**Andrew, Peterson, & Longstaff, 1979** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib53>)). The MCT is divided into two subtests: Number Comparison and Name Comparison. Each subtest consists of 100 identical and 100 dissimilar pairs of digit or letter combinations (**Table 11.4**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch11lev1sec4#ch11tab4>)). The dissimilar pairs generally differ in regard to only one digit or letter, so the comparison task is challenging. The examinee is required to check only the identical pairs, which are randomly intermixed with dissimilar pairs. The score depends predominantly upon speed, although the examinee is penalized for incorrect items (errors are subtracted from the number of correct items).

TABLE 11.4 Items Similar to Those Found on the Minnesota Clerical Test

Number Comparison

1. 3496482 _____ 3495482
2. 17439903 _____ 17439903
3. 84023971 _____ 84023971
4. 910386294 _____ 910368294

Name Comparison

1. New York Globe _____ New York Globe
2. Brownell Seed _____ Brownel Seed
3. John G. Smith _____ John G Smith
4. Daniel Gregory _____ Daniel Gregory

The reliability of the MCT is acceptable, with reported stability coefficients in the range of .81 to .87 (**Andrew, Peterson, & Longstaff, 1979**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib53>)). The manual also reports a wealth of validity data, including some findings that are not altogether flattering. In these studies, the MCT was correlated with measures of job performance, measures of training outcome, and scores from related tests. The job performance of directory assistants, clerks, clerk-typists, and bank tellers was correlated significantly but not robustly with scores on the MCT. The MCT is also highly correlated with other tests of clerical ability.

Nonetheless, questions still remain about the validity and applicability of the MCT. Ryan (1985 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1424>)) notes that the manual lacks a discussion of the significant versus the nonsignificant validity studies. In addition, the MCT authors fail to provide detailed information concerning the specific attributes of the jobs, tests, and courses used as criterion measures in the reported validity studies. For this reason, it is difficult to surmise exactly what the MCT measures. Ryan (1985 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1424>)) complains that the 1979 norms are difficult to use because the MCT authors provide so little information on how the various norm groups were constituted. Thus, even though the revised MCT manual presents new norms for 10 vocational categories, the test user may not be sure which norm group applies to his or her setting. Because of the marked differences in performance between the norm groups, the vagueness of definition poses a significant problem to potential users of this test.

11.5 PERSONALITY TESTS

It is only in recent years, with the emergence of the “big five” approach to the measurement of personality and the development of strong measures of these five factors, that personality has proved to be a *valid* basis for employee selection, at least in some instances. In earlier times such as the 1950s into the 1990s, personality tests were used by many in a reckless manner for personnel selection:

Personality inventories such as the MMPI were used for many years for personnel selection—in fact, overused or misused. They were used indiscriminately to assess a candidate’s personality, even when there was no established relation between test scores and job success. Soon personality inventories came under attack. (Muchinsky, 1990)

In effect, for many of these earlier uses of testing, a consultant psychologist or human resource manager would look at the personality test results of a candidate and implicitly (or explicitly) make an argument along these lines: “In my judgment people with test results like this are [or are not] a good fit for this kind of position.” Sadly, there was little or no empirical support for such imperious conclusions, which basically amounted to a version of “because I said so.”

Certainly early research on personality and job performance was rather sobering for many personality scales and constructs. For example, Hough, Eaton, Dunnette, Kamp, and McCloy (1990 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib790>)) analyzed hundreds of published studies on the relationship between personality constructs and various job performance criteria. For these studies, they grouped the personality constructs into several categories (e.g., Extroversion, Affiliation, Adjustment, Agreeableness, and Dependability) and then computed the average validity coefficient for criteria of job performance (e.g., involvement, proficiency, delinquency, and substance abuse). Most of the average correlations were indistinguishable from zero! For job proficiency as the outcome criterion, the strongest relationships were found for measures of Adjustment and Dependability, both of which revealed correlations of $r = .13$ with general ratings of job proficiency. Even though statistically significant (because of the large number of clients amassed in the hundreds of studies), correlations of this magnitude are essentially useless, accounting for less than 2 percent of the variance.² (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch11lev1sec5#ch11fn02>) Specific job criteria such as delinquency (e.g., neglect of work duties) and substance abuse were better predicted in specific instances. For example, measures of Adjustment correlated $r = -.43$ with delinquency, and measures of Dependability correlated $r = -.28$ with substance abuse. Of course, the negative correlations indicate an inverse relationship: higher scores on Adjustment go along with lower levels of delinquency, and higher scores on Dependability indicate lower levels of substance abuse. Apparently, it is easier to predict specific job-related criteria than to predict general job proficiency.

Beginning in the 1990s, a renewed optimism about the utility of personality tests in personnel selection began to emerge (Behling, 1998

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib121>) ; Hurtz & Donovan, 2000 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib803>)). The reason for this change in perspective was the emergence of the Big Five framework for research on selection, and the development of robust measures of the five constructs confirmed by this approach such as the NEO Personality Inventory-Revised (Costa & McCrae, 1992

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib353>)). Evidence began to mount that personality—as conceptualized by the Big Five approach—possessed some utility for employee selection. The reader will recall from an earlier chapter that the five dimensions of this model

are Neuroticism, Extraversion, Openness to Experience, Conscientiousness, and Agreeableness. Shuffling the first letters, the acronym OCEAN can be used to remember the elements. In place of Neuroticism (which pertains to the negative pole of this factor), some researchers use the term Emotional Stability (which describes the positive pole of the same factor) so as to achieve consistency of positive orientation among the five factors.

A meta-analysis by Hurtz and Donovan (2000

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib803>) solidified Big Five personality factors as important tools in predicting job performance. These researchers located 45 studies using suitable measures of Big Five personality factors as predictors of job performance. In total, their data set was based on more than eight thousand employees, providing stable and robust findings, even though not all dimensions were measured in all studies. The researchers conducted multiple analyses involving different occupational categories and diverse outcome measures such as task performance, job dedication, and interpersonal facilitation. We discuss here only the most general results, namely, the operational validity for the five factors in predicting overall job performance. Operational validity refers to the correlation between personality measures and job performance, corrected for sampling error, range restriction, and unreliability of the criterion. Big Five factors and validity coefficients were as follows:

Conscientiousness	.26
Neuroticism	.13
Extraversion	.15
Agreeableness	.05
Openness to Experience	.04

Overall, Conscientiousness is the big winner in their analysis, although for some specific occupational categories, other factors were valuable (e.g., Agreeableness paid off for Customer Service personnel). Hurtz and Donovan (2000

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib803>) use caution and understatement to summarize the implications of their study:

What degree of utility do these global Big Five measures offer for predicting job performance? Overall, it appears that global measures of Conscientiousness can be expected to consistently add a small portion of explained variance in job performance across jobs and across criterion dimension. In addition, for certain jobs and for certain criterion dimensions, certain other Big Five dimensions will likely add a very small but consistent degree of explained variance. (p. 876)

In sum, people who describe themselves as reliable, organized, and hard-working (i.e., high on Conscientiousness) appear to perform better at work than those with fewer of these qualities.

For specific applications in personnel selection, certain tests are known to have greater validity than others. For example, the California Psychological Inventory (CPI) provides an accurate measure of managerial potential (Gough, 1984

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib629>), 1987

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib630>). Certain scales of the CPI predict overall performance of military academy students reasonably well (Blake, Potter, & Sliwak, 1993 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib161>)). The Inwald Personality Inventory is well validated as a preemployment screening test for law enforcement

(Chibnall & Detrick, 2003

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib302>) ; Inwald, 2008

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib808>)). The Minnesota Multiphasic Personality Inventory also bears mention as a selection tool for law enforcement (Selbom, Fischler, & Ben-Porath, 2007). Finally, the Hogan Personality Inventory (HPI) is well validated for prediction of job performance in military, hospital, and corporate settings (Hogan, 2002

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib757>)). The HPI was based upon the Big Five theory of personality (see **Topic 8A**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch08#ch08box1>) , Theories and the Measurement of Personality). This instrument has cross-validated criterion-related validities as high as .60 for some scales (Hogan, 1986

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib756>) ; Hogan & Hogan, 1986 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib756>)).

²The strength of a correlation is indexed by squaring it, which provides the proportion of variance accounted for in one variable by knowing the value of the other variable. In this case, the square of .13 is .0169 which is 1.69 percent.

11.6 PAPER-AND-PENCIL INTEGRITY TESTS

Several test publishers have introduced instruments designed to screen theft-prone individuals and other undesirable job candidates such as persons who are undependable or frequently absent from work (Cullen & Sackett, 2004

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib378>) ; Wanek, 1999 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1712>)). We will focus on issues raised by these tests rather than detailing the merits or demerits of individual instruments. **Table 11.5** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch11lev1sec6#ch11tab5>) lists some of the more commonly used instruments.

One problem with integrity tests is that their proprietary nature makes it difficult to scrutinize them in the same manner as traditional instruments. In most cases, scoring keys are available only to in-house psychologists, which makes independent research difficult. Nonetheless, a sizable body of research now exists on integrity tests, as discussed in the following section on validity.

An **integrity test** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm01#bm01gloss160>) evaluates attitudes and experiences relating to the honesty, dependability, trustworthiness, and pro-social behaviors of a respondent. Integrity tests typically consist of two sections. The first is a section dealing with attitudes toward theft and other forms of dishonesty such as beliefs about extent of employee theft, degree of condemnation of theft, endorsement of common rationalizations about theft, and perceived ease of theft. The second is a section dealing with overt admissions of theft and other illegal activities such as items stolen in the last year, gambling, and drug use. The most widely researched tests of this type include the Personnel Selection Inventory, the Reid Report, and the Stanton Survey. The interested reader can find addresses for the publishers of these and related instruments through Internet search.

TABLE 11.5 Commonly Used Integrity Tests

Overt Integrity Tests

Accutrac Evaluation System
Compuscan
Employee Integrity Index
Orion Survey
PEOPLE Survey
Personnel Selection Inventory
Phase II Profile
Reid Report and Reid Survey
Stanton Survey

Personality-Based Integrity Tests

Employment Productivity Index
Hogan Personnel Selection Series
Inwald Personality Inventory
Personnel Decisions, Inc., Employment Inventory
Personnel Reaction Blank

Note: Publishers of these tests can be easily found by using Google or another internet search engine.

Apparently, integrity tests can be easily faked and might, therefore, be of less value in screening dishonest applicants than other approaches such as background check. For example, Ryan and Sackett (1987 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1422>)) created a generic overt integrity test modeled upon existing instruments. The test contained 52 attitude and 11 admission items. In comparison to a contrast group asked to respond truthfully and another contrast group asked to respond as job applicants, subjects asked to “fake good” produced substantially superior scores (i.e., better attitudes and fewer theft admissions).

Validity of Integrity Tests

In a recent meta-analysis of 104 criterion-related validity studies, Van Iddekinge, Roth, Raymark, and Odle-Dusseau (2012) found that integrity tests were not particularly useful in predicting job performance, training performance, or work turnover (corrected r s of .15, .16, and .09, respectively). However, when counterproductive work behavior (CWB, e.g., theft, poor attendance, unsafe behavior, property destruction) was the criterion, the corrected r was a healthy .32. The correlation was even higher, $r = .42$, when based on self-reports of CWB as opposed to other reports or employee records. Overall, these findings support the value of integrity testing in personnel selection. Ones et al. (1993) requested data on integrity tests from publishers, authors, and colleagues. These sources proved highly cooperative: The authors collected 665 validity coefficients based upon 25 integrity tests administered to more than half a million employees. Using the intricate procedures of meta-analysis, Ones et al. (1993) computed an average validity coefficient of .41 when integrity tests were used to predict supervisory ratings of job performance. Interestingly, integrity tests predicted global disruptive behaviors (theft, illegal activities, absenteeism, tardiness, drug abuse, dismissals for theft, and violence on the job) better than they predicted employee theft alone. The authors concluded with a mild endorsement of these instruments:

When we started our research on integrity tests, we, like many other industrial psychologists, were skeptical of integrity tests used in industry. Now, on the basis of analyses of a large database consisting of more than 600 validity coefficients, we conclude that integrity tests have substantial evidence of generalizable validity.

This conclusion is echoed in a series of ingenious studies by Cunningham, Wong, and Barbee (1994 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib381>)). Among other supportive findings, these researchers discovered that integrity test results were correlated with returning an overpayment—even when subjects were instructed to provide a positive impression on the integrity test.

Other reviewers are more cautious in their conclusions. In commenting on reviews by the American Psychological Association and the Office of Technology Assessment, Camara and Schneider (1994 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib32>)) concluded that integrity tests do not measure up to expectations of experts in assessment, but that they are probably better than hit-or-miss, un-standardized methods used by many employers to screen applicants.

Several concerns remain about integrity tests. Publishers may release their instruments to unqualified users, which is a violation of ethical standards of the American Psychological Association. A second problem arises from the unknown base rate of theft and other undesirable behaviors, which makes it difficult to identify optimal cutting scores on integrity tests. If cutting scores are too stringent, honest job candidates will be disqualified unfairly. Conversely, too lenient a cutting score renders the testing pointless. A final concern is that situational factors may moderate the validity of these instruments. For

example, how a test is portrayed to examinees may powerfully affect their responses and therefore skew the validity of the instrument.

The debate about integrity tests juxtaposes the legitimate interests of business against the individual rights of workers. Certainly, businesses have a right not to hire thieves, drug addicts, and malcontents. But in pursuing this goal, what is the ultimate cost to society of asking millions of job applicants about past behaviors involving drugs, alcohol, criminal behavior, and other highly personal matters? Hanson (1991 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib696>)) has asked rhetorically whether society is well served by the current balance of power—in which businesses can obtain proprietary information about who is seemingly worthy and who is not. It is not out of the question that Congress could enter the debate. In 1988, President Reagan signed into law the Employee Polygraph Protection Act, which effectively eliminated polygraph testing in industry. Perhaps in the years ahead we will see integrity testing sharply curtailed by an Employee Integrity Test Protection Act. Berry, Sackett, and Wiemann (2007 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib152>)) provide an excellent review of the current state of integrity testing.

11.7 WORK SAMPLE AND SITUATIONAL EXERCISES

A **work sample** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm01#bm01gloss352>) is a miniature replica of the job for which examinees have applied. Muchinsky (2003 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1186>)) points out that the I/O psychologist's goal in devising a work sample is "to take the content of a person's job, shrink it down to a manageable time period, and let applicants demonstrate their ability in performing this replica of the job." Guion (1998 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib671>)) has emphasized that work samples need not include every aspect of a job but should focus upon the more difficult elements that effectively discriminate strong from weak candidates. For example, a position as clerk-typist may also include making coffee and running errands for the boss. However, these are trivial tasks demanding so little skill that it would be pointless to include them in a work sample. A work sample should test important job domains, not the entire job universe.

Campion (1972 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib263>)) devised an ingenious work sample for mechanics that illustrates the preceding point. Using the job analysis techniques discussed at the beginning of this topic, Campion determined that the essence of being a good mechanic was defined by successful use of tools, accuracy of work, and overall mechanical ability. With the help of skilled mechanics, he devised a work sample that incorporated these job aspects through typical tasks such as installing pulleys and repairing a gearbox. Points were assigned to component behaviors for each task. Example items and their corresponding weights were as follows:

Installing Pulleys and Belts

Scoring Weights

1. Checks key before installing against:

___ shaft	2
___ pulley	2
___ neither	0

Disassembling and Repairing a Gear Box

10. Removes old bearing with:

___press and driver	3
___bearing puller	2
___gear puller	1
___other	0

Pressing a Bushing into Sprocket and Reaming to Fit a Shaft

4. Checks internal diameter of bushing against shaft diameter:

___visually	1
___hole gauge and micrometers	3
___Vernier calipers	2
___scale	1
___does not check	0

Campion found that the performance of 34 male maintenance mechanics on the work sample measure was significantly and positively related to the supervisor's evaluations of their work performance, with

validity coefficients ranging from .42 to .66.

A situational exercise

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm01#bm01gloss293>) is approximately the white-collar equivalent of a work sample. Situational exercises are largely used to select persons for managerial and professional positions. The main difference between a situational exercise and a work sample is that the former mirrors only part of the job, whereas the latter is a microcosm of the entire job (Muchinsky, 1990). In a situational exercise, the prospective employee is asked to perform under circumstances that are highly similar to the anticipated work environment. Measures of accomplishment can then be gathered as a basis for gauging likely productivity or other aspects of job effectiveness. The situational exercises with the highest validity show a close resemblance with the criterion; that is, the best exercises are highly realistic (Asher & Sciarrino, 1974

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib65>) ; Muchinsky, 2003 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1186>)).

Work samples and situational exercises are based on the conventional wisdom that the best predictor of future performance in a specific domain is past performance in that same domain. Typically, a situational exercise requires the candidate to perform in a setting that is highly similar to the intended work environment. Thus, the resulting performance measures resemble those that make up the prospective job itself.

Hundreds of work samples and situational exercises have been proposed over the years. For example, in an earlier review, Asher and Sciarrino (1974

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib65>)) identified 60 procedures, including the following:

- Typing test for office personnel
- Mechanical assembly test for loom fixers
- Map-reading test for traffic control officers
- Tool dexterity test for machinists and riveters
- Headline, layout, and story organization test for magazine editors
- Oral fact-finding test for communication consultants
- Role-playing test for telephone salespersons
- Business-letter-writing test for managers

A very effective situational exercise that we will discuss here is the in-basket technique, a procedure that simulates the work environment of an administrator.

The In-Basket Test

The classic paper on the **in-basket test**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm01#bm01gloss152>) is the monograph by Frederiksen (1962

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib539>)). For this comprehensive study Frederiksen devised the Bureau of Business In-Basket Test, which consists of the letters, memoranda, records of telephone calls, and other documents that have collected in the in-basket of a newly hired executive officer of a business bureau. In this test, the candidate is instructed not to play a role, but to be himself.³ (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch11lev1sec7#ch11fn03>) The candidate is not to say what he would do, he is to do it.

The letters, memoranda, phone calls, and interviews completed by him in this simulated job environment constitute the record of behavior that is scored according to both content and style of the responses. *Response style* refers to how a task was completed—courteously, by telephone, by involving a superior, through delegation to a subordinate, and so on. *Content* refers to what was done, including making plans, setting deadlines, seeking information; several quantitative indices were also computed, including number of items attempted and total words written. For some scoring criteria such as imaginativeness—the number of courses of action which seemed to be good ideas—expert judgment was required.

Frederiksen (1962 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib539>)) administered his in-basket test to 335 subjects, including students, administrators, executives, and army officers. Scoring the test was a complex procedure that required the development of a 165-page manual. The odd-even reliability of the individual items varied considerably, but enough modestly reliable items emerged (rs of .70 and above) that Frederiksen could conduct several factor analyses and also make meaningful group comparisons.

When scores on the individual items were correlated with each other and then factor analyzed, the behavior of potential administrators could be described in terms of eight primary factors. When scores on these primary factors were themselves factor analyzed, three second-order factors emerged. These second-order factors describe administrative behavior in the most general terms possible. The first dimension is Preparing for Action, characterized by deferring final decisions until information and advice is obtained. The second dimension is simply Amount of Work, depicting the large individual differences in the sheer work output. The third major dimension is called Seeking Guidance, with high scorers appearing to be anxious and indecisive. These dimensions fit well with existing theory about administrator performance and therefore support the validity of Frederiksen's task.

A number of salient attributes emerged when Frederiksen compared the subject groups on the scorable dimensions of the in-basket test. For example, the undergraduates stressed verbal productivity, the government administrators lacked concern with outsiders, the business executives were highly courteous, the army officers exhibited strong control over subordinates, and school principals lacked firm control. These group differences speak strongly to the construct validity of the in-basket test, since the findings are consistent with theoretical expectations about these subject groups.

Early studies supported the predictive validity of in-basket tests. For example, Brass and Oldham (1976 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib210>)) demonstrated that performance on an in-basket test corresponded to on-the-job performance of supervisors if the appropriate in-basket scoring categories were used. Specifically, based on the in-basket test, supervisors who personally reward employees for good work, personally punish subordinates for poor work, set specific performance objectives, and enrich their subordinates' jobs are also rated by their superiors as being effective managers. The predictive power of these in-basket dimensions was significant, with a multiple correlation coefficient of .54 between predictors and criterion. Standardized in-basket tests can now be purchased for use by private organizations. Unfortunately, most of these tests are "in-house" instruments not available for general review. In spite of occasional cautionary reviews (e.g., Brannick et al., 1989 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib208>) ; Schroffell, 2012 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1474>)), the in-basket technique is still highly regarded as a useful method of evaluating candidates for managerial positions.

Assessment Centers

An assessment center is not so much a place as a process (Highhouse & Nolan, 2012). Many corporations and military branches—as well as a few progressive governments—have dedicated special sites to the application of in-basket and other simulation exercises in the training and selection of managers. The purpose of an **assessment center**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm01#bm01gloss18>) is to evaluate managerial potential by exposing candidates to multiple simulation techniques, including group presentations, problem-solving exercises, group discussion exercises, interviews, and in-basket techniques. Results from traditional aptitude and personality tests also are considered in the overall evaluation. The various simulation exercises are observed and evaluated by successful senior managers who have been specially trained in techniques of observation and evaluation. Assessment centers are used in a variety of settings, including business and industry, government, and the military. There is no doubt that a properly designed assessment center can provide a valid evaluation of managerial potential. Follow-up research has demonstrated that the performance of candidates at an assessment center is strongly correlated with supervisor ratings of job performance (Gifford, 1991

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib587>)). A more difficult question to answer is whether assessment centers are cost-effective in comparison to traditional selection procedures. After all, funding an assessment center is very expensive. The key question is whether the assessment center approach to selection boosts organizational productivity sufficiently to offset the expense of the selection process. Anecdotally, the answer would appear to be a resounding yes, since poor decisions from bad managers can be very expensive. However, there is little empirical information that addresses this issue.

Goffin, Rothstein, and Johnston (1996

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib601>)) compared the validity of traditional personality testing (with the Personality Research Form; Jackson, 1984b (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib818>)) and the assessment center approach in the prediction of the managerial performance of 68 managers in a forestry products company. Both methods were equivalent in predicting performance, which would suggest that the assessment center approach is not worth the (very substantial) additional cost. However, when both methods were used in combination, personality testing provided significant incremental validity over that of the assessment center alone. Thus, personality testing and assessment center findings each contribute unique information helpful in predicting performance.

Putting a candidate through an assessment center is very expensive. Dayan, Fox, and Kasten (2008 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib402>)) speak to the cost of assessment center operations by arguing that an employment interview and cognitive ability test scores can be used to cull the best and the worst applicants so that only those in the middle need to undergo these expensive evaluations. Their study involved 423 Israeli police force candidates who underwent assessment center evaluations after meeting initial eligibility. The researchers concluded in retrospect that, with minimal loss of sensitivity and specificity, nearly 20 percent of this sample could have been excused from more extensive evaluation. These were individuals who, based on interview and cognitive test scores, were nearly sure to fail or nearly certain to succeed.

³We do not mean to promote a subtle sexism here, but in fact Frederiksen (1962

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib539>)) tested a predominantly (if not exclusively) male sample of students, administrators, executives, and army officers.

11.8 APPRAISAL OF WORK PERFORMANCE

The appraisal of work performance is crucial to the successful operation of any business or organization. In the absence of meaningful feedback, employees have no idea how to improve. In the absence of useful assessment, administrators have no idea how to manage personnel. It is difficult to imagine how a corporation, business, or organization could pursue an institutional mission without evaluating the performance of its employees in one manner or another.

Industrial and organizational psychologists frequently help devise rating scales and other instruments used for performance appraisal (**Landy & Farr, 1983**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib948>)). When done properly, employee evaluation rests upon a solid foundation of applied psychological measurement—hence, its inclusion as a major topic in this text. In addition to introducing essential issues in the measurement of work performance, we also touch briefly on the many legal issues that surround the selection and appraisal of personnel. We begin by discussing the context of performance appraisal.

The evaluation of work performance serves many organizational purposes. The short list includes promotions, transfers, layoffs, and the setting of salaries—all of which may hang in the balance of performance appraisal. The long list includes at least 20 common uses identified by Cleveland, Murphy, and Williams (**1989**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1191>)). These applications of performance evaluation cluster around four major uses: comparing individuals in terms of their overall performance levels; identifying and using information about individual strengths and weaknesses; implementing and evaluating human resource systems in organizations; and documenting or justifying personnel decisions. Beyond a doubt, performance evaluation is essential to the maintenance of organizational effectiveness.

As the reader will soon discover, performance evaluation is a perplexing problem for which the simple and obvious solutions are usually incorrect. In part, the task is difficult because the criteria for effective performance are seldom so straightforward as “dollar amount of widgets sold” (e.g., for a salesperson) or “percentage of students passing a national test” (e.g., for a teacher). As much as we might prefer objective methods for assessing the effectiveness of employees, judgmental approaches are often the only practical choice for performance evaluation.

The problems encountered in the implementation of performance evaluation are usually referred to collectively as the criterion problem—a designation that first appeared in the 1950s (e.g., **Flanagan, 1956** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib509>) ; **Landy & Farr, 1983** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib948>)). The phrase **criterion problem**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm01#bm01gloss79>) is meant to convey the difficulties involved in conceptualizing and measuring performance constructs, which are often complex, fuzzy, and multidimensional. For a thorough discussion of the criterion problem, the reader should consult comprehensive reviews by Austin and Villanova (**1992**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib69>)) and Campbell, Gasser, and Oswald (**1996**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib260>)). We touch upon some aspects of the criterion problem in the following review.

11.9 APPROACHES TO PERFORMANCE APPRAISAL

There are literally dozens of conceptually distinct approaches to the evaluation of work performance. In practice, these numerous approaches break down into four classes of information: performance measures such as productivity counts; personnel data such as rate of absenteeism; peer ratings and self-assessments; and supervisor evaluations such as rating scales. Rating scales completed by supervisors are by far the preferred method of performance appraisal, as discussed later. First, we mention the other approaches briefly.

Performance Measures

Performance measures include seemingly objective indices such as number of bricks laid for a mason, total profit for a salesperson, or percentage of students graduated for a teacher. Although production counts would seem to be the most objective and valid methods for criterion measurement, there are serious problems with this approach (Guion, 1965). The problems include the following:

- The rate of productivity may not be under the control of the worker. For example, the fast-food worker can only sell what people order, and the assembly-line worker can only proceed at the same pace as coworkers.
- Production counts are not applicable to most jobs. For example, relevant production units do not exist for a college professor, a judge, or a hotel clerk.
- An emphasis upon production counts may distort the quality of the output. For example, pharmacists in a mail-order drug emporium may fill prescriptions with the wrong medicine if their work is evaluated solely upon productivity.

Another problem is that production counts may be unreliable, especially over short periods of time. Finally, production counts may tap only a small proportion of job requirements, even when they appear to be the definitive criterion. For example, sales volume would appear to be the ideal criterion for most sales positions. Yet, a salesperson can boost sales by misrepresenting company products. Sales may be quite high for several years—until the company is sued by unhappy customers. Productivity is certainly important in this example, but the corporation should also desire to assess interpersonal factors such as honesty in customer relations.

Personnel Data: Absenteeism

Personnel data such as rate of absenteeism provide another possible basis for performance evaluation. Certainly employers have good reason to keep tabs on absenteeism and to reduce it through appropriate incentives. Steers and Rhodes (1978

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1549>)) calculated that absenteeism costs about \$25 billion a year in lost productivity! Little wonder that absenteeism is a seductive criterion measure that has been researched extensively (Harrison & Hulin, 1989 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib711>)).

Unfortunately, absenteeism turns out to be a largely useless measure of work performance, except for the extreme cases of flagrant work truancy. A major problem is defining absenteeism. Landy and Farr (1983 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib948>)) list 28 categories of absenteeism, many of which are uncorrelated with the others. Different kinds of absenteeism include scheduled versus unscheduled, authorized versus unauthorized, justified versus unjustified, contractual versus noncontractual, sickness versus nonsickness, medical versus personal, voluntary versus involuntary, explained versus unexplained, compensable versus noncompensable,

certified illness versus casual illness, Monday/Friday absence versus midweek, and reported versus unreported. When is a worker truly absent from work? The criteria are very slippery.

In addition, absenteeism turns out to be an atrociously unreliable variable. The test-retest correlations (absentee rates from two periods of identical length) are as low as .20, meaning that employees display highly variable rates of absenteeism from one time period to the next. A related problem with absenteeism is that workers tend to underreport it for themselves and overreport it for others (Harrison & Shaffer, 1994

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib712>)). Finally, for the vast majority of workers, absenteeism rates are quite low. In short, absenteeism is a poor method for assessing worker performance, except for the small percentage of workers who are chronically truant.

Peer Ratings and Self-Assessments

Some researchers have proposed that peer ratings and self-assessments are highly valid and constitute an important complement to supervisor ratings. A substantial body of research pertains to this question, but the results are often confusing and contradictory. Nonetheless, it is possible to list several generalizations (Harris & Schaubroeck, 1988

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib710>) ; Smither, 1994):

- Peers give more lenient ratings than supervisors.
- The correlation between self-ratings and supervisor ratings is minimal.
- The correlation between peer ratings and supervisor ratings is moderate.
- Supervisors and subordinates have different ideas about what is important in jobs.

Overall, reviewers conclude that peer ratings and self-assessments may have limited application for purposes such as personal development, but their validity is not yet sufficiently established to justify widespread use (Smither, 1994).

Supervisor Rating Scales

Rating scales are the most common measure of job performance (Landy & Farr, 1983 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib948>) ; Muchinsky, 2003 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1186>)). These instruments vary from simple graphic forms to complex scales anchored to concrete behaviors. In general, supervisor rating scales reveal only fair reliability, with a mean interrater reliability coefficient of .52 across many different approaches and studies (Viswesvaran, Ones, & Schmidt, 1996 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1702>)). In spite of their weak reliability, supervisor ratings still rank as the most widely used approach. About three-quarters of all performance evaluations are based upon judgmental methods such as supervisor rating scales (Landy, 1985).

The simplest rating scale is the graphic rating scale, introduced by Donald Paterson in 1922 (Landy & Farr, 1983 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib948>)). A graphic rating scale

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm01#bm01gloss137>) consists of trait labels, brief definitions of those labels, and a continuum for the rating. As the reader will notice in **Figure 11.1** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch11lev1sec9#ch11fig1>) , several types of graphic rating scales have been used.

The popularity of graphic rating scales is due, in part, to their simplicity. But this is also a central weakness because the dimension of work performance being evaluated may be vaguely defined. Dissatisfaction with graphic rating scales led to the development of many alternative approaches to performance appraisal, as discussed in this section.

A critical incidents checklist

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm01#bm01gloss82>) is based upon actual episodes of desirable and undesirable on-the-job behavior (Flanagan, 1954 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib508>)). Typically, a checklist developer will ask employees to help construct the instrument by submitting specific examples of desirable and undesirable job behavior. For example, suppose that we intended to develop a checklist to appraise the performance of resident advisers (RAs) in a dormitory. Modeling a study by Aamodt, Keller, Crawford, and Kimbrough (1981 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib01>))), we might ask current dormitory RAs the following question:

Think of the best RA that you have ever known. Please describe in detail several incidents that reflect why this person was the best adviser. Please do the same for the worst RA you have ever known.

Based upon hundreds of nominated behaviors, checklist developers would then proceed to distill and codify these incidents into a smaller number of relevant behaviors, both desirable and undesirable. For example, the following items might qualify for the RA checklist:

- ☐ stays in dorm more than required
- ☐ breaks dormitory rules
- ☐ is fair about discipline
- ☐ plans special programs
- ☐ fails to discipline friends
- ☐ is often unfriendly
- ☐ shows concern about residents
- ☐ comes across as authoritarian

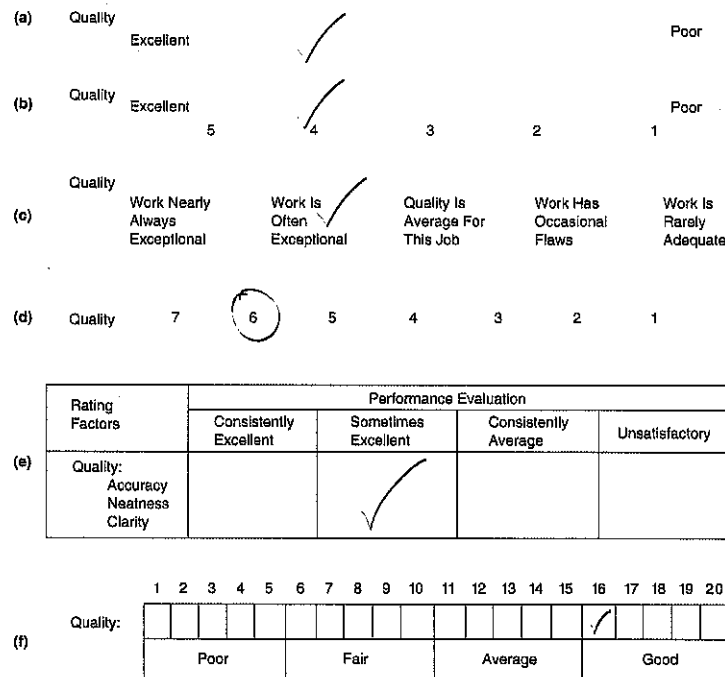


FIGURE 11.1 Examples of Graphic Rating Scales

Of course, the full checklist would be much longer than the preceding. The RA supervisor would complete this instrument as a basis for performance appraisal. If needed, an overall summary score can be derived from an appropriate weighting of individual items.

Another form of criterion-referenced judgmental measure is the **behaviorally anchored rating scale** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm01#bm01gloss32>) (BARS). The classic work on BARS dates back to Smith and Kendall (1963 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1513>)). These authors proposed a complex developmental procedure for producing criterion-referenced judgments. The procedure uses a number of experts to identify and define performance dimensions, generate behavior examples, and scale the behaviors meaningfully. Overall, the procedure is quite complex, time-consuming, and expensive. A number of variations and improvements have been suggested. An advantage to BARS and other behavior-based scales is their strict adherence to EEOC (Equal Employment Opportunity Commission) guidelines discussed later in this chapter. BARS and related approaches focus upon behaviors as opposed to personality or attitudinal characteristics. A behaviorally anchored scale for performance of college professors in posting office hours is depicted in **Figure 11.2** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch11lev1sec9#ch11fig2>). Of course, the comprehensive evaluation of a sales manager would include additional scales for other aspects of work.

Could be expected to post required and extra office hours the first week of the semester, maintain them without exception, and greet students in a friendly manner.	7
Could be expected to post required and extra office hours the first week of the semester, and maintain them without exception.	6
Could be expected to post required and extra office hours the first week of the semester, and maintain them most of the time.	5
Could be expected to post required office hours the first week of the semester, and maintain them most of the time.	4
Could be expected to post required office hours by mid-semester, and maintain them most of the time.	3
Could be expected to post required office hours with "push" from department chair, but would miss office hours without notice.	2
Could be expected to resist posting office hours and fail to maintain them.	1

FIGURE 11.2 Behaviorally Anchored Rating Scale for Posting and Maintaining Office Hours

Research on improving the accuracy of ratings with BARS is mixed. Some studies find fewer rating errors—especially a reduction in unwarranted leniency of evaluations—whereas other studies report no improvement with BARS compared to other evaluation methods (Murphy & Pardaffy, 1989 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1191>)). Overall, Muchinsky (2003 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1186>)) concludes that the BARS approach is not much better than graphic rating scales in reducing rating errors. Nonetheless, the scale development process of BARS may have indirect benefits in that supervisors are compelled to pay close attention to the behavioral components of effective performance.

A behavior observation scale

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm01#bm01gloss26>) (BOS) is a variation upon the BARS technique. The difference between the two is that the BOS approach uses a continuum from "almost never" to "almost always" to measure how often an employee performs the specific tasks on each behavioral dimension. As with the BARS technique, researchers question whether behavior observation scales are worth the extra effort (Guion, 1998 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib671>)).

A forced-choice scale

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm01#bm01gloss123>) is designed to eliminate bias and subjectivity in supervisor ratings by forcing a choice between options that are equal in social desirability. In theory, this approach makes it impossible for the supervisor to slant ratings in a biased or subjective manner. We will use the pathbreaking research by Sisson (1948 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1503>)) to illustrate the features of this approach. He developed a scale to evaluate Army officers that consisted of tetrads of behavioral descriptors. Each tetrad contained two positive items matched for social desirability and two negative items also matched for social desirability. The four items in each tetrad were topically related to a single performance dimension. Unknown to the supervisors who completed the rating scale, one of the two positive items was judged very descriptive of effective Army officers and the other judged less so. Likewise, one of the two negative items was judged more descriptive of ineffective Army officers and the other judged less so. Here is a sample tetrad (Borman, 1991):

Most DescriptiveLeast Descriptive

A. Cannot assume responsibility

B. Knows how and when to delegate authority

Most DescriptiveLeast Descriptive

C. Offers suggestions

D. Changes ideas too easily

Supervisors were asked to review the items in each tetrad and to check one item as most descriptive and one item as least descriptive of the officer being evaluated. A score of +1 was awarded for responding "most descriptive" to the positively keyed item (in this case, alternative B) or "least descriptive" to the negatively keyed item (in this case alternative A), whereas a score of -1 was awarded for responding "least descriptive" to the positively keyed item or "most descriptive" to the negatively keyed item. Responding to the nonkeyed items (alternatives C and D) as most or least descriptive earned a score of 0. Thus, each tetrad yielded a five-point continuum of scores: +2, +1, 0, -1, -2. The summary score used for performance appraisal consisted of the algebraic sum of the individual items.

The forced-choice approach has never really caught on, due largely to the effort required in scale construction. This is unfortunate because the method does effectively reduce unwanted bias. Borman (1991) refers to this approach as a "bold initiative" that produces a relatively objective rating scale.

11.10 SOURCES OF ERROR IN PERFORMANCE APPRAISAL

The most difficult problem in the assessment of job performance is the proper definition of appraisal criteria. If the supervisor is using a poorly designed instrument that does not tap the appropriate dimensions of job behavior, then almost by definition the performance appraisal will be inaccurate, incomplete, and erroneous. Undoubtedly, the failure to identify appropriate criteria for acceptable and unacceptable performance is a major source of error in performance appraisal. But it is not the only source. Even when supervisors have access to excellent, well-designed measures of performance appraisal, various sorts of subtle errors can creep in. We discuss three such additional sources of rating error: halo effect, rater bias, and criterion contamination.

Halo Effect

The tendency to rate an employee high or low on all dimensions because of a global impression is called **halo effect** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm01#bm01gloss142>). Research on the halo effect can be traced back to the early part of this century (**Thorndike, 1920** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1637>)). The most common halo effect is a positive halo effect. In this case, an employee receives a higher rating than deserved because the supervisor fails to be objective when rating specific aspects of the employee's behavior. A positive halo effect is usually based upon overgeneralization from one element of a worker's behavior. For example, an employee with perfect attendance may receive higher-than-deserved evaluations on productivity and work quality—even though attendance is not directly related to these job dimensions.

Smither (1998) lists the following approaches to control for halo effects:

- Provide special training for raters
- Supervise the supervisors during the rating
- Practice simulations before doing the ratings
- Keep a diary of information relevant to appraisal
- Provide supervisors with a short lecture on halo effects

Additional approaches to rater training are discussed by Goldstein (1991 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib612>)). An intriguing analysis of the nature and consequences of halo error can be found in Murphy, Jako, and Anhalt (1993 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1192>)). Contrary to the reigning prejudice against halo errors, these researchers conclude that the halo effect does not necessarily detract from the accuracy of ratings. They point out that a presumed halo effect is often the by-product of true overlap on the dimensions being rated. The debate over halo effect is not likely to be resolved anytime soon (**Arvey & Murphy, 1998** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib64>))).

Rater Bias

The potential sources of **rater bias**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm01#bm01gloss268>) are so numerous that we can only mention a few prominent examples here. Leniency or severity errors occur when a supervisor tends to rate workers at the extremes of the scale. Leniency may reflect social dynamics, as when the supervisor wants to be liked by employees. Leniency is also caused by extraneous factors such as the attractiveness of the employee. Severity errors refer to the practice of rating all aspects of

performance as deficient. In contrast, central tendency errors occur when the supervisor rates everyone as nearly average on all performance dimensions. Context errors occur when the rater evaluates an employee in the context of other employees rather than based upon objective performance. For example, the presence of a workaholic salesperson with extremely high sales volume might cause the sales supervisor to rate other sales personnel lower than deserved.

Recently, researchers have paid considerable attention to the possible biasing effects of whether a supervisor likes or dislikes a subordinate. Surprisingly, the trend of the findings is that supervisor affect (liking or disliking) toward specific employees does not introduce rating bias. In general, strong affect in either direction represents valid information about an employee. Thus, ratings of affect often correlate strongly with performance ratings, but this is because both are a consequence of how well or poorly the employee does the job (Ferris, Judge, Rowland, & Fitzgibbons, 1994

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib496>) ; Varma, DeNisi, & Peters, 1996 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1690>)).

Other forms of rater bias are discussed by Goldstein (1991

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib612>)) and Smither (1994).

Criterion Contamination

Criterion contamination

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm01#bm01gloss77>) is said to exist when a criterion measure includes factors that are not demonstrably part of the job (Borman, 1991; Harvey, 1991). For example, if a performance measure includes appearance, this would most likely be a case of criterion contamination—unless appearance is relevant to job success. Likewise, evaluating an employee on “dealing with the public” is only appropriate if the job actually requires the employee to meet the public. Goldstein (1992

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib613>)) outlines three kinds of criterion contamination:

- Opportunity bias occurs when workers have different opportunities for success, as when one salesperson is assigned to a wealthy neighborhood and others must seek sales in isolated, rural areas.
- Group characteristic bias is present when the characteristics of the group affect individual performance, as when workers in the same unit agree to limit their productivity to maintain positive social relations.
- Knowledge of predictor bias occurs when a supervisor permits personal knowledge about an employee to bias the appraisal, as when quality of the college attended by a new worker affects her evaluation.

Careful attention to job analysis as a basis for selection of appraisal criteria is the best way to reduce errors in performance appraisal. In addition, employers should follow certain guidelines in performance appraisal, as discussed in the following section.

Guidelines for Performance Appraisal

Performance appraisal is a formidable task. Not only must employers pay attention to the psychometric soundness of their approach, they must also design a practical system that meets organizational goals. For example, appraisal standards must be sufficiently difficult and detailed to ensure that organizational goals are accomplished. Another concern is that performance appraisal falls under the purview of Title VII of the Civil Rights Act of 1964. Hence, employers must develop fair systems that do not discriminate

on the basis of race, sex, and other protected categories. To complicate matters, these standards—soundness, practicality, legality—may conflict with one another. The practical approach may be neither psychometrically sound nor legal. Often, appraisal methods that show the best measurement characteristics (e.g., strong interrater reliability) will fail to assess the most important aspects of performance; that is, they are not practical. This is a familiar refrain within the measurement field. Too often, psychologists must choose between rigor and relevance, rarely achieving both at the same time. Finally, legal considerations must be considered when exploring the limits of performance appraisal.

Smither (1998) has published guidelines for developing performance appraisal systems that we paraphrase here:

- Base the performance appraisal upon a careful job analysis
- Develop specific, contamination-free criteria for appraisal from the job analysis
- Determine that the instrument used to rate performance is appropriate for the appraisal situation
- Train raters to be accurate, fair, and legal in their use of the appraisal instrument
- Use performance evaluations at regular intervals of six months to a year
- Evaluate the performance appraisal system periodically to determine whether it is actually improving performance

The training of raters is an especially important guideline. An appraisal system that seems perfectly straightforward to the employer could easily be misunderstood by an untrained rater, resulting in biased evaluations. Borman (1991) notes that two kinds of rater training are effective: rater error training, in which the trainer seeks simply to alert raters to specific kinds of errors (e.g., halo effect); and frame-of-reference training, in which the trainer familiarizes the raters with the specific content of each performance dimension. Research indicates that these kinds of training improve the accuracy of ratings.

Finally, we review an intriguing study conducted from an international perspective. Peretz and Fried (2012 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1281>)) remind us that cultural norms influence the nature, acceptability, and impact of different approaches to performance appraisal. They surveyed performance appraisal practices in 21 nations, obtained ratings on cultural norms for each nation, and determined their joint impact on organizational absenteeism and turnover. Specifically, the researchers collected data on personnel practices from thousands of organizations in these mainly European countries. Next, they obtained ratings for each country on four cultural practices: power distance (acceptance versus rejection of inequality), future orientation (present versus future orientation), person value (individualism versus collectivism), and uncertainty avoidance (acceptance versus avoidance of uncertainty). Each cultural norm was rated 1 to 7 for each nation based on an independent global data base. Then, they examined the joint impact of personnel practices and cultural norms on absenteeism and turnover. Their study is complex and detailed, beyond the scope of fine-tuned analysis here. In sum, they found that congruence between societal norms and personnel assessment methods tended to reduce turnover and/or absenteeism. One example is the use of the so-called 360-evaluation, in which performance appraisal is based on input from people at all levels who interact with the employee. This practice is more effective (leading to less absenteeism and turnover) in some cultures than others. Peretz and Fried (2012

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1281>)) found that personnel assessment systems with several sources of raters (e.g., supervisors, coworkers, and subordinates) were most acceptable to employees in companies located in societies with low power distance, high future orientation, and respect for individualism. In contrast, multiple sources of

assessment were not well received by employees working in collectivistic societies. It appears the best practices in personnel assessment depend upon the cultural context.

TOPIC 11B Assessment for Career Development in a Global Economy

11.11 Career Developments and the Functions of Work

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch11lev1sec11#ch11lev1sec11>)

11.12 Origins of Career Development Theories

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch11lev1sec12#ch11lev1sec12>)

11.13 Theory of Person-Environment Fit

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch11lev1sec13#ch11lev1sec13>)

11.14 Theory of Person-Environment Correspondence

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch11lev1sec14#ch11lev1sec14>)

11.15 Stage Theories of Career Development

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch11lev1sec15#ch11lev1sec15>)

11.16 Social Cognitive Approaches

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch11lev1sec16#ch11lev1sec16>)

11.17 O*NET in Career Development

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch11lev1sec17#ch11lev1sec17>)

11.18 Inventories for Career Assessment

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch11lev1sec18#ch11lev1sec18>)

Prior to the 1700s, agrarian economies dominated cultural and economic life in the Western world. Vocational opportunities for most people remained limited to farming, crafts, labor, and small businesses. The modern vision that individuals could pursue dozens or hundreds of careers likely did not exist for the masses who scrambled simply to survive (Zinn, 1995). With the advent of the first industrial revolution in the 1700s, including the invention of the steam engine and other labor saving devices, the need for human labor diminished rapidly. In parallel, the vocational world expanded substantially, offering upward mobility to some of the working class and poor. Gradually, the concept of career identity emerged in the public consciousness.

Career identity is now recognized as essential to personhood and vital to a sense of well-being. When we meet someone for the first time, our natural inclination is to ask, or at least to wonder, "What do you do for a living?" The values, political views, and personal qualities of the individual are important, too, but how the individual contributes to society is typically the first thing we want to know. An occupational title communicates an abundance of information, including personality characteristics, economic class, and social standing (Andersen & Vandehey, 2011

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib49>)).

Work and career are so central to personal well-being that unemployment, especially when prolonged, consistently causes a wide range of physical, psychological, and spiritual maladies. These include:

... economic hardship, loss of health insurance, foreclosure, and mental health problems. The mental health problems include depression and anxiety, feelings of hopelessness and shame, and familial tension and conflict (**Jones & Barber, 2012**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib848>) , p. 18).

A meta-analysis of 104 empirical studies revealed that the negative impact of unemployment is buffered by the availability of coping resources (e.g., family and financial support) and, conversely, made worse by work-role centrality (e.g., the belief that work is central to one's life and satisfaction) (**McKee-Ryan, Song, Wanberg, & Kinicki, 2005**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1103>)).

Except in a few totalitarian states where occupational access is rigidly controlled by the ruling elite, individuals usually have some degree of latitude in finding their own way to a vocation. They also possess some capacity to change occupations in their lifetimes. Even though the widely cited assertion that the average individual will switch careers seven times has no factual basis, nonetheless, career change likely is more common now than in years past (**Bialik, 2010**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib156>)). Also, initial career choice for the young adult remains a vexing issue for many, especially with the continual emergence of substantially new vocations. The advent of new vocations is driven by technological innovations and the aging of the population. A few examples of new careers include cloud computing expert, market research data analyst, and corporate listening officer (*Forbes* magazine, May 5, 2011).

The need for flexibility in career development originates, in part, from the globalization of the world economies, spelled out in the provocative best seller, *The World Is Flat* (**Friedman, 2009**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib547>)). Information technology is now instantly available to everyone, linking knowledge centers into a single worldwide network, creating a more level economic playing field, and requiring corporations to restructure as new opportunities emerge. One concrete example of the new, flat world: For the previous edition of this textbook, the editorial production and composition services were completed by the skilled and efficient employees of a dynamic company located in India. After a few phone calls and email exchanges of PDF files with the author, the text was ready for printing in the United States in a matter of weeks.

In summary, psychologists who provide career guidance will need new approaches to assessment that are sensitive to the need for transition planning in a rapidly changing and increasingly competitive global economy. But practitioners need to avoid the "Test and Tell" trap:

Clients often come to career counseling assuming an expert will administer some test that tells the client "the answer" as to what occupation is "the right one." The client's expectation for "test and tell" sets the stage for the client and the counselor to depend on a limited, structured approach (**Andersen & Vandehey, 2011**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib49>) , p. 10).

The problem with this method is that the counselor will fail to discern the unique needs of the client in a developmentally sensitive context. Guidance will be far more effective if the practitioner slows the process down and provides the opportunity for mutual exploration. In other words, career guidance is a tactic of assessment in the broader sense, not a limited method of testing in the narrow sense.

Assessment for career development requires knowledge of theories of career development, sensitivity to issues of diversity, and understanding of information resources. Thus, before turning to a survey of suitable instruments, we begin with a brief review of prominent career development theories. We start with a simple but provocative question pursued by Blustein (2006 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib168>)), "What is work for?"

11.11 CAREER DEVELOPMENT AND THE FUNCTIONS OF WORK

For some people, gainful employment provides more than just a means to pay for food and housing. Psychologists who provide assessment for career development need to keep in mind the multiple functions of work, reviewed here. Yet, it is also true that many people, perhaps the majority, do not have access to the educational and employment opportunities that would allow them to develop a work vision or to realize a career dream.

Since recorded time, humanity has been plagued by various forms of structural barriers based on race, culture, immigration status, religion, gender, age, sexual orientation, and social class that have had a differential impact on individuals. Our belief is that counselors need to be fully cognizant of how these barriers affect clients so that they are able to provide maximally effective interventions that do not inadvertently blame the victims of social oppression (**Blustein, Kenna, Gill, & DeVoy, 2008** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib169>) , p. 297).

It bears repeating that discrimination continues to obstruct career potential for minorities. A subtle racism on the part of employers and agencies often is the source. Many studies could be cited to buttress this point as a global issue. For reasons of space, we offer just two examples. A recent study from Great Britain confirms that ethnic minorities experience an "ethnic penalty" with higher unemployment rates, greater concentrations in dead-end assembly line jobs, and lower earnings than Whites, even for the same job (**Bell & Casebourne, 2008**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib124>)). Immigrants to Great Britain likewise face career barriers. When able to find work, it is typically in just a few industries such as catering, language translation, shop work, and clerical jobs. Professional employment was notably lacking, despite previous experience (**Bloch, 2002** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib163>)).

Unfortunately, most theories of career development do not acknowledge the profound challenges faced by low income individuals, minorities, and immigrants. The psychology-of-working viewpoint provided by Blustein and his collaborators is an exception. These researchers provide a meta-theoretical perspective that can be used alongside traditional models of career development. We begin with a summary of their model.

According to Blustein and colleagues (**2008**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib169>)), work can fulfill any or all of three sets of needs:

- *Survival and Power:* These are the foundational reasons that most people work, namely, to meet basic subsistence needs such as food, clothing, and shelter. In varying degrees, work also provides access to economic and social power. Specifically, those with financial resources are more likely to prevail and to get their way in the wider community. Money talks.
- *Social Connection:* Work is the place where many of our vital human connections are formed. Deep friendships are forged and sometimes maintained over a lifetime. The quality of these relationships has the potential to enhance performance when coworkers are positive and supportive, or to create great stress when colleagues are abrasive and conflict-prone.
- *Self-Determination:* For some individuals, work is also a means of self-actualization and personal fulfillment. Everyone is familiar with those fortunate individuals who love what they do and are privileged to be paid for it, too. But Blustein et al. (**2008** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib169>)) remind us

that many workers do not have the opportunity to select a career that provides for creative and fulfilling self-expression.

In addition to discrimination, structural barriers often prevent career development among minorities. For example, African Americans may lack relevant social networks, lack public transit for employment, and lack savings needed to relocate for available work (**Weller & Fields, 2011**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1748>)). Further, unemployment is itself a serious structural barrier. In 2011, unemployment among African Americans was about 16 percent, double that of Whites. These data do not include those who have quit looking for work, or who are chronically underemployed. Being out of work tends to become a vicious, self-perpetuating cycle, with the unemployed individual losing work skills with each passing month, further reducing employment prospects.

11.12 ORIGINS OF CAREER DEVELOPMENT THEORIES

Implicitly or explicitly, practitioners make use of a theoretical framework in their practice of assessment in career counseling. Thus, we provide a short review of essential viewpoints here. We begin with an historical note, acknowledging the seminal contributions of Frank Parsons, considered by many the founder of the field of career guidance. In 1909, he published *Choosing a Vocation*, a practical manual for providing career direction to young men and women. Parsons (1909

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1267>)) advocated making a career choice based on matching personal traits with job factors:

In the wise choice of a vocation there are three broad factors: (1) a clear understanding of yourself, your aptitudes, abilities, interests, ambitions, resources, limitations, and their causes; (2) a knowledge of the requirements and conditions of success, advantages and disadvantages, compensation, opportunities, and prospects in different lines of work; (3) true reasoning on the relations of these two groups of facts (p. 5).

Parsons provided a 116-item questionnaire to survey the accomplishments, interests, and aptitudes of the client. This was followed by a lengthy, penetrating interview designed to illuminate aspects of social presentation and personal character (e.g., "Do you smile naturally and easily?" "Is your handshake warm and cordial?" "Are you careful about voice modulation?" "Are you honest, truthful, and candid?" "Are you industrious, hard-working, and persistent?" "Do you welcome people of different creed or political faith?"). His manual also provided an extensive analysis of the qualities needed for success in dozens of vocations. Consultation with each client continued over a span of several weeks. The task of the counselor was to match the traits of the client with the requirements of specific lines of work. Effectively, this was an early, rudimentary form of the method advocated by John Holland and others, known as person-environment fit.

11.13 THEORY OF PERSON-ENVIRONMENT FIT

Over 50 years ago, John Holland (1959

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib762>) established the framework for a sophisticated theory of vocational choice that has engendered more research than any other approach in the field. From the beginning, he also constructed and validated assessment tools that embodied the practical application of his model, known as Person-Environment Fit. He proposed that personality traits/interests tend to cluster into a small number of vocationally relevant patterns, called types. For each personality type, there is also a corresponding work environment best suited to that type. According to Holland, there are six types: Realistic, Investigative, Artistic, Social, Enterprising, and Conventional. Each type corresponds to both a set of personality traits/interests and also to a set of environmental work demands. **Figure 11.3**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch11lev1sec13#ch11fig3>) depicts this approach, sometimes known as the **RIASEC model**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm01#bm01gloss280>), in reference to the first letters of the six types. The types are idealizations that few people (or work environments) fit completely. The RIASEC personality patterns are summarized in **Table 11.6**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch11lev1sec13#ch11tab6>), and corresponding work environments are found in **Table 11.7**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch11lev1sec13#ch11tab7>).

Regarding the six personality types, it is rare that an individual is a “pure” representation of only one type. Instead, most individuals reveal a preferred type, but display some resemblance to a secondary and a tertiary type as well. For example, someone who was very strong on the Investigative dimension (likes to analyze) might reveal a secondary emphasis for the Social aspect (enjoys helping others), and a lesser emphasis on the Artistic type (reveals a creative element). Using the first letters of these three types in descending order of emphasis, we arrive at the **Holland code** for the individual, namely, **ISA**. We will say more about Holland codes when we discuss assessment tools such as the *Self-Directed Search* developed for this purpose. For now it will suffice to know that excellent tools exist for the empirically validated assessment of the six types.

Consistency and differentiation are two concepts important in the Holland approach. Referring to the hexagonal model depicted in **Figure 11.3**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch11lev1sec13#ch11fig3>), adjacent personality types bear greater similarity to one another than types that are separated on the figure. For example, the Realistic and Conventional types (side by side) are somewhat similar, whereas the Realistic and Social types (across the hexagon) are quite different or inconsistent. Thus, a client whose Holland code was **RCE** (adjacent codes) would be considered more consistent than a client whose code was **REA** (separated codes). This is relevant to assessment and career guidance because work environments tend to possess consistency in regard to types. It is easier for clients to find person–environment fit when they possess consistency, too.

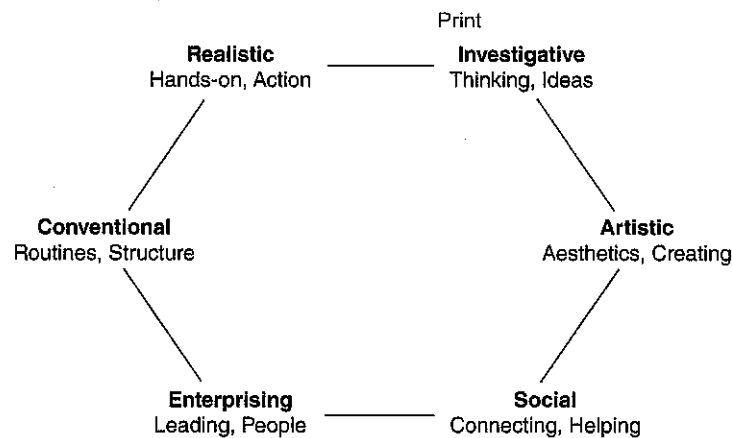


FIGURE 11.3 Holland's Hexagonal Model of Personality Types and Occupational Themes

TABLE 11.6 RIASEC Personality Types

Realistic individuals are practical, conservative, and value tangible rewards. They like to work with tools, machines, and things. They usually avoid interaction with others.

Investigative persons show a strong analytical bent. They value knowledge and like to explore, understand, and predict natural and social phenomena. They tend to avoid selling or persuading others.

Artistic individuals are unconventional and enjoy the creative expression of ideas and emotions. They value musical, literary, or artistic endeavors. They avoid conformity to standards and routine activities.

Social persons possess empathy and strong social skills. They value interpersonal contact, teaching, and serving others. They typically avoid technical or mechanical activities.

Enterprising individuals are good at persuading and directing others. They value social status and material accomplishment. They typically avoid intellectual or abstract topics.

Conventional persons like to maintain order and establish routines. They value financial success and power in social or business settings. They usually avoid ambiguous or unstructured situations.

Source: Based on Holland, J. L. (1985

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib765>)). *Vocational Preference Inventory (VPI) manual—1985 edition*. Odessa, FL: Psychological Assessment Resources.

TABLE 11.7 RIASEC Work Environments

Realistic work environments require hands-on involvement, physical movement, mechanical skill, and technical competencies. Pragmatic problem solving is needed. Typical vocations include auto repair, cook, drywall installer, machinist, taxi driver, and umpire.

Investigative settings require the use of abstract thinking and creative abilities. The focus is a rational approach and ideas, not people. Typical positions include architect, arson investigator, pharmacist, physician, psychologist, and software engineer.

Artistic environments require the creative application of artistic forms. These settings demand prolonged work and place a premium on access to intuition and emotional life. Typical vocations include actor, composer, graphic designer, model, photographer, and reporter.

Social environments involve an interest in caring for people and the ability to discern and influence their behavior. These work settings require good social skills and the ability to deal with emotionally laden interactions. Typical positions include clergy, teacher, emergency medical technician, marriage therapist, psychiatric aide, and waitperson.

Enterprising work environments involve the influence of others through verbal skills. These roles require self-confidence and leadership capacities for directing and controlling the activity of others. Typical vocations include bartender, real estate agent, construction manager, first-line supervisor, police detective, and travel agent.

Conventional work environments require the methodical, routine, and concrete processing of words and mathematical symbols. The key to these settings is repetitive application of established clerical procedures. Typical settings include bank teller, bookkeeper, court recorder, insurance underwriter, office clerk, and shipping clerk.

Source: Based on Holland, J. L. (1985

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib765>)). *Vocational Preference inventory (VPI) manual—1985 edition*. Odessa, FL: Psychological Assessment Resources.

Differentiation refers to the relative strength of the first, second, and third personality types of the Holland code. A client with strong differentiation will reveal a marked preference for his or her first category, and less interest in the second and third categories. A client with weak differentiation might demonstrate scores that are nearly tied on the top three categories of the Holland code. This could indicate a difficulty committing to one kind of work environment. Most work environments require some degree of differentiation. Hence, the undifferentiated client may struggle to find a satisfying work environment.

Holland's theoretical approach has been so influential that nearly every assessment tool in the field of career guidance makes reference to his six personality types. But the simple elegance of this approach is also a potential weakness. The assessment tools that embody the Holland model typically list suitable occupations and rule out nonmatching environments. Counselors and clients can foreclose on further exploration. It is easy to fall into the "test and tell" trap.

11.14 THEORY OF PERSON-ENVIRONMENT CORRESPONDENCE

The theory of Person-Environment Correspondence (PEC) evolved from the Theory of Work Adjustment (TWA). First envisioned in the 1950s, TWA arose as a basis for conducting research on the work adjustment of vocational rehabilitation clients. Soon it became clear that TWA applied to situations other than rehabilitation, and that the approach was a specific case of a more general method, which came to be known as Person-Environment Correspondence or PEC (**Dawis, 2002** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib399>)).

PEC bears modest similarity to the personenvironment approach advocated by Holland and colleagues. The central point of similarity is that, in determining suitable careers, both theories compare the attributes of individuals with the qualities needed in occupations (**Dawis, 1996** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib398>) ; Dawis & Lofquist, 1991). One difference is that PEC places greater emphasis on individual abilities and their match to the ability patterns required by specific occupations. Ability is different from skill level, which can be acquired with preparation. Ability refers to aptitude, indicating the level of mastery an individual can achieve with suitable training and experience. Another difference is that PEC places greater weight on individual values and their correspondence to the value fulfillments provided by specific occupations (**Dawis, 2002** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib399>) ; **Eggerth, 2008** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib450>)).

PEC theory identifies six crucial values that need to be considered in assessment and counseling for career development. These values are as follows:

- *Achievement*—the importance of using one's abilities and having a feeling of accomplishment
- *Altruism*—the importance of harmony with, and being of service to, others
- *Autonomy*—the importance of being independent and having a sense of control
- *Comfort*—the importance of feeling comfortable and not being stressed
- *Safety*—the importance of stability, order, and predictability
- *Status*—the importance of recognition and being in a dominant position (**Dawis, 2002** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib399>) , p. 446).

This list is not comprehensive and it is likely that additional values will emerge with further research. Of course, correspondence between personal values held by the client and the potential for their fulfillment in an occupation is central to work satisfaction and productivity.

PEC theory is rich in complexity because it has evolved over more than five decades; we can only provide a few highlights. The central principle is that the more closely the rewards of the job or the organization correspond to the core values of the individual, the more likely it is that he or she will find satisfaction with the position. But PEC also invokes cognitive, personality, and environmental styles in its understanding of work adjustment. For example, environmental styles include celerity, pace, rhythm, and endurance required to complete the job, which are each assessed on a continuum (**Dawis, 1996** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib398>)):

- *Celerity* refers to the quickness of response that is needed in responding to job demands. For example, emergency room personnel often need to respond very quickly, whereas a diamond cutter would be foolhardy to do so.
- *Pace* refers to the level of effort needed in responding to the environment. A position such as office clerical worker might require modest effort in comparison to firefighter, where periods of

intense effort will be encountered.

- *Rhythm* refers to whether the pattern of responding is steady, cyclical, or erratic. An example of a steady environment would be telephone operator, whereas a police officer might work in an erratic environment, facing hours of boredom interrupted with occasional bursts of fear.
- *Endurance* refers to whether the duration of responding to environmental demands is brief or protracted. A position requiring less endurance might be financial advisor, whereas a computer software engineer employed under deadline would need to keep working, day and night, until the project is finished.

Andersen and Vandehey (2012) provide a useful illustration of how these environmental styles play out for specific occupations:

Two examples demonstrating differing styles are an emergency room and a gemsmith. An emergency room requires cyclical, intense work periods as well as down times. Medical personnel need high celerity (be fast) with a high level of effort (pace). Also, some surgeries could last up to 16 hours, requiring high endurance. By contrast, a gemsmith is ill advised to be fast when cutting gems, and the celerity requirements are low. In addition, several outstanding gems may be worth more money than many poorly cut stones (low pace). The work environment has a steady rhythm and probably requires varying amounts of endurance, depending upon the stone size and complexity of the cuts (p. 47).

Of course, these four dimensions also manifest as measurable personality styles. In the world of career counseling, a mismatch between these two broad factors (environmental style required by a job, personality style preferred by the client) often is a precipitating referral issue.

Dawis and colleagues offer 17 testable propositions derived from PEC and provide a wealth of supporting research (**Dawis, 2002**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib399>) ; **Dawis & Lofquist, 1984** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib400>)). For example, one proposition is:

Proposition III: P's satisfaction is a function of the correspondence of E's reinforcers to P's values, provided that P's abilities correspond to E's ability requirements.

Put simply, a person's satisfaction with a job is a function of the match of the available environmental reinforcers with the values of the individual, provided that his or her abilities correspond to those required by the position. This is an empirically testable hypothesis that has stood up well in research studies (**Dawis, 2002**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib399>)).

11.15 STAGE THEORIES OF CAREER DEVELOPMENT

Beginning in the 1950s, Donald Super and colleagues developed an influential stage theory in the field of career guidance and development (**Super, 1953**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1601>) , 1994

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1603>)). His approach departs from the trait-factor method preferred by many in the field, and embraces a more flexible, holistic, life span perspective. The essentials of the theory were stated with elegant simplicity in his first and most widely cited article, "A theory of vocational development" (**Super, 1953**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1601>)). Later papers provided additional details to the original framework (**Super, Savickas, & Super, 1996**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1604>)).

Super acknowledged the obvious fact that people differ in their abilities, interests, and personalities, but also believed that most people were qualified for several occupations, not just a few positions.

Individuals and occupations were each flexible enough to "allow both some variety of occupations for each individual and some variety of individuals in each occupation" (**Super, 1953**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1601>) , p. 189).

He argued that the individual self-concept evolves with time and experience, so that vocational choice and adjustment are continuous and lifelong processes. He envisioned five occupational life stages: growth, exploration, establishment, maintenance, and decline. These stages are sometimes known as a career ladder (**Super et al., 1996**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1604>)).

The growth stage extends into the teenage years and involves the observation of adult behavior and the exploration of fantasies and interests. The exploration stage was subdivided into fantasy, tentative, and realistic phases, as the young adult tries out one or more lines of training or education toward an eventual career. The establishment stage begins around age 25 or 30, and was subdivided into the trial and stabilization phases. Vocational development tasks encountered in this stage include the assimilation of organizational climate, the consolidation of positive relationships with coworkers, and the advancement of career responsibilities through promotion (**Super, 1990**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1602>)).

In the maintenance stage of middle age, the individual may need to innovate, update skills, or face career stagnation. Additionally, some persons ask: "Should I remain in this career?" If the answer is "No" then the individual would reenter the exploration and establishment stages before attaining the maintenance stage. The last stage, decline, is hypothesized to occur in old age and may require possible specialization, disengagement, or retirement.

The stage development theory proposed by Super provides a useful reminder that career development does not end in young adulthood but extends throughout the life span. However, the theory was based on career development as found in the dominant culture of his time which was mainly white and often middle class or higher. In a changing global economy, some of the developmental stages no longer seem as relevant. In particular, the maintenance phase is difficult for many to sustain because of the need for frequent career transitions (**Friedman, 2009**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib547>)). Super died in 1994. Toward the end of his career, he acknowledged new realities:

Work and occupation provide a focus for personality organization for most men and women, although for some individuals this focus is peripheral, incidental, or even nonexistent. Then other foci such as leisure activities and homemaking, may be central. Social traditions such as sex-role stereotyping and modeling, racial and ethnic biases, and the opportunity structure, as well as individual differences are important determinants of preferences for such roles as worker, student, leisurite, homemaker, and citizen (**Super et al., 1996**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1604>) , p. 126).

The brief mention of “opportunity structure” is important to underscore, in light of the Great Recession experienced worldwide in the early part of the twenty-first century.

11.16 SOCIAL COGNITIVE APPROACHES

Social cognitive approaches to career development acknowledge that people learn and develop attitudes about work within a social context through observation and modeling of behavior. Prominent exemplars of this approach include Gottfredson (2005

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib627>)), Lent, Brown, and Hackett (2000 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib969>)), and Krumboltz (2009 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib925>)).

In our coverage here, we summarize the recent views of John Krumboltz because of their direct relevance to matters of assessment. Krumboltz (2009

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib925>)) calls his approach the **Happenstance Learning Theory (HLT)**. In brief:

HLT posits that human behavior is the product of countless numbers of learning experiences made available by both planned and unplanned situations in which individuals find themselves. The learning outcomes include skills, interests, knowledge, beliefs, preferences, sensitivities, emotions, and future actions (p. 135).

The theory is practical and compassionate in style, attempting to explain how and why each person follows a unique path, and describing how counselors can facilitate development. In regard to the how and why of behavior, Krumboltz surveys genetic influences, learning experiences, environmental conditions, parents and caretaker influences, peer groups, and structured educational settings. He concludes by noting that “Social justice is not equally distributed among humans on our planet.” He argues powerfully that practitioners have a responsibility to help overcome social injustice. The proper uses of assessment might be a small part of the solution.

HLT is based on four premises (Krumboltz, 2009

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib925>)):

- The goal of career counseling is to help clients learn to take actions to achieve more satisfying career and personal lives—not to make a single career decision.

Krumboltz notes that the future is uncertain for everyone, especially in the world of work, where new careers emerge and old ones die out. In his view, making a single career decision is potentially foolhardy. A more tentative, exploratory approach is to be preferred.

- Assessments are used to stimulate learning—not to match personal characteristics with occupational characteristics.

For example, in regard to interest assessment, Krumboltz contends that the goal is to help clients find attractive activities to explore now. In regard to happenstance, it is his experience that helping clients commit to new actions often will open up unexpected opportunities. A similar argument holds for personality assessment, which can be used to stimulate discussion about alternative settings for the client, and to identify areas of needed change (e.g., assertiveness training for an introverted client). It may also prove helpful to identify dysfunctional career beliefs by using the *Career Beliefs inventory* (Krumboltz & Vosvick, 1996 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib924>)), which is discussed later in this topic.

Krumboltz (1993

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib923>) , 1996

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib924>)) has been

critical of many interest inventories because most clients have little or no experience with the topics being assessed. Instead of marking items as like, dislike, or indifferent, he playfully suggests that the response options should be “I don’t know yet,” “I haven’t tried that yet,” or “I’d like to learn more about that before I answer” (Kumboltz, 1996

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib924>) , p. 57). He also finds fault with these instruments because they focus excessively on cognitive matching of client to work environments, and overlook the emotional problems, including dysfunctional career beliefs, that hamper career development.

- Clients learn to engage in exploratory actions as a way of generating beneficial unplanned events—not to plan all their actions in advance. The statement that “chance favors only the prepared mind” is attributed to Louis Pasteur (1822-1895), the French biologist and chemist. But the statement can be applied to career development as well. Kumboltz asserts that the goal of the counselor is to help clients engage in activities that are likely to generate unplanned events, and to prepare clients to benefit from these happenstance occurrences. An example might be encouraging an unemployed client to join a health club as a means of exploring her interests in yoga. At the club she befriends a bank manager who is impressed by her winsome personality, which leads to a job interview and a new career endeavor.
- The success of counseling is assessed by what the client accomplishes in the real world outside the counseling session—not by what takes place during counseling.

HLT is an action-based theory. The task of the counselor is to collaboratively identify things that the client can do outside of the consultation that will promote new learning and new opportunities. A simple example is asking the client to commit to one action step between appointments (e.g., ask three people how they came to be working in their current job) and to report back by email how things went.

11.17 O*NET IN CAREER DEVELOPMENT

The Occupational Information Network or O*NET is the primary source of occupational information in the United States. O*NET is sponsored by the U.S. Department of Labor and is free and open to anyone in the world who has an Internet connection. This is a rich and sophisticated database that includes detailed information on nearly 1,000 specific occupations. For each occupation, the website lists the knowledge, skills, and abilities needed. Personality qualities needed, education required, technology needs, and typical salary also are given.

The website provides several assessment tools for career exploration, including a number of instruments that can be self-administered. For example, the O*NET Interest Profiler is an online test consisting of 60 occupational activities that are rated on a five-point scale from *strongly dislike* to *strongly like*. The test not only yields a score for each of the six RIASEC dimensions, but also links to a user-friendly list of specific occupations suited to the preparation level selected by the examinee. Further, these occupations are individually rated for employment outlook, environmental or “green” appeal, and apprenticeship needed.

11.18 INVENTORIES FOR CAREER ASSESSMENT

One guiding motif in this topic is that successful assessment for career guidance requires ongoing interaction with clients. Career counseling extends well beyond mere testing. Avoiding the “test and tell” trap is vital. Even so, the use of appropriate assessment tools can be helpful, sometimes even essential. The number of instruments available for career assessment is huge, and new tools emerge every year. We survey a number of widely used tests here, to provide a sense of the diversity available. We begin with a specialized tool designed to challenge mal-adaptive career beliefs.

Career Beliefs Inventory

Krumboltz (1991) created the Career Beliefs Inventory to identify and measure attitudes and beliefs that might block career development. In his work with clients, he often noted that people firmly hold to self-limiting beliefs that prevent them from finding a satisfying job or career. Examples of such beliefs include:

- I don't have enough confidence to try that
- I don't have the skills needed for that position
- I can't do that because I don't have any experience
- I'm really dumb when it comes to that kind of activity
- It would involve too much risk to go in that direction
- That kind of work wouldn't give me any satisfaction

The Career Beliefs Inventory (CBI) was designed to increase the awareness of clients to underlying career beliefs and to gauge the potential influence of these beliefs on occupational choice and life satisfaction.

The CBI can be taken individually or administered in a group setting to persons in grade 8 or higher. The paper-and-pencil test can be hand-scored, but computer-scoring is preferable because it yields an elegant 12-page report. Hand scoring is also confusing and likely to introduce errors.

The 96 test items, all in Likert format, are grouped into 25 scales organized under the following five headings:

- Your Current Career Situation. Four Scales: Employment Status, Career Plans, Acceptance of Uncertainty, and Openness.
- What Seems Necessary for Your Happiness. Five scales: Achievement, College Education, Intrinsic Satisfaction, Peer Equality, and Structured Work Environment.
- Factors that Influence Your Decisions. Six scales: Control, Responsibility, Approval of Others, Self-other Comparisons, Occupation/College Variation, and Career Path Flexibility.
- Changes You Are Willing to Make. Three scales: Post-training Transition, Job Experimentation, and Relocation.
- Effort You Are Willing to Initiate. Seven scales: Improving Self, Persisting While Uncertain, Taking Risks, Learning Job Skills, Negotiating/Searching, Overcoming Obstacles, and Working Hard.

Standardization of the CBI is based on more than 7,500 individuals in the United States and Australia. The sample was reasonably diverse, with age range of 12 to 75, including junior high, high school, and college students, as well as adults, both employed and unemployed. Initial test-retest reliability data for the CBI are mixed, with one month reliabilities ranging from .30s to the .70s for the high school sample. Internal consistencies were likewise modest, with coefficients mainly in the range of .40 to .50. This might

be due to the small number of items for some scales, as few as two items for several scales. Fuqua and Newman (1994 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib554>)) recommend that the CBI could be improved if additional items were added to some of the scales.

Walsh (1996 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1709>)) supplemented the original standardization sample for the CBI with nearly 600 additional participants. She reported more promising results, with internal consistencies ranging from the low .30s to the high .80s, with a mean coefficient alpha of .57 for the CBI scale scores. Regarding validity, results of factor analyses did find reproducible clusters of beliefs, but these did not correspond to the scale clusters provided in the CBI reports. She suggests that the practical application of the CBI might rest with exploring client beliefs at the level of the individual items (Walsh, Thompson, & Kapes, 1996 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1709>)).

In a study of convergent validity correlating CBI results with data from four other personality and vocational inventories, Holland, Johnston, Asama, and Polys (1993 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib770>)) reported at least moderate construct validity for most of the CBI scales. They concluded that the test seems to be measuring variance in career variables not assessed by other instruments. In addition, significant correlation of some CBI scales with the State-Trait Anxiety Inventory indicated that certain self-limiting and irrational beliefs caused emotional discomfort.

11.19 INVENTORIES FOR INTEREST ASSESSMENT

In most applications of psychological testing, the goals of assessment are reasonably clear. For example, intelligence testing helps predict school performance; aptitude testing foretells potential for accomplishment; and personality testing provides information about social and emotional functioning. But what is the purpose of interest assessment? Why would a psychologist recommend it? What can a client expect to gain from a survey of his or her interests?

Interest assessment promotes two compatible goals: life satisfaction and vocational productivity. It is nearly self-evident that a good fit between individual interests and chosen vocation will help foster personal life satisfaction. After all, when work is interesting we are more likely to experience personal fulfillment as well. In addition, persons who are satisfied with their work are more likely to be productive. Thus, employees and employers both stand to gain from the artful application of interest assessment. Several useful instruments exist for this purpose, and we will review the most widely used interest inventories later.

In the selection of employees, the consideration of personal interests may be of great practical significance to employers and, therefore, circumstantially relevant to the job candidates as well. We may sketch out a rough equation as follows: productivity = ability X interest. In other words, high ability in a specific field does not guarantee success; neither does high interest level. The best predictions are possible when both variables are considered together. Thus, employers have good reason to determine whether a potential employee is well matched to the position; the employee should like to know as well.

Working from the Holland RIASEC model described earlier, Ny, Su, Rounds, and Drasgow (2012) recently completed an intriguing quantitative summary of 60 years of research on the relationship between vocational interests, person-environment fit, and job performance. Their review was based on 568 correlations from published empirical studies. The basic premise of their survey was that:

Holland's theory suggests that the similarities between an individual's interest profile and the profile of his or her occupation should predict tenure and performance in academic and work domains (p. 387).

This is exactly what their analyses revealed. For the employment studies reviewed, the correlations between "fit" (congruence between an individual's Holland code and the code of his/her chosen occupation) and job performance ranged from .21 to .30, depending on the inventory used and the characteristics of the study. The same pattern emerged in the academic samples. The correlations between "fit" (congruence between a student's Holland code and the code of his/her chosen major) and grades were mainly in the range of .27 to .31. In other words, when employees or students possess interest patterns that match the expectations of their job or major, they are more likely to be productive in their work or studies.

We turn now to a critical examination of major interest tests. The four instruments chosen for review include:

- The Strong Interest Inventory-Revised (SII-R), the latest revision of the well-known Strong Vocational Interest Blank (SVIB)
- The Vocational Preference Inventory (VPI), a useful inventory that embodies the RIASEC model of John Holland

- The Self-Directed Search (SDS), a self-administered and self-scored guide to exploring career options
- The Campbell Interest and Skill Survey (CISS), an appealing test that is simple in format but sophisticated in execution

Strong Interest Inventory-Revised (SII-R)

The Strong Interest Inventory-Revised (SII-R) is the latest revision of the Strong Vocational Interest Blank (SVIB), one of the oldest and most prominent instruments in psychological testing (**Donnay, Thompson, Morris, & Schaubhut, 2004**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib433>)). We can best understand the SII-R by studying the history of its esteemed predecessor, the SVIB. In particular, we need to review the guiding assumptions used in the construction of the SVIB that have been carried over into the SII-R.

The first edition of the SVIB appeared in 1927, eight years after E. K. Strong formulated the essential procedures for measuring occupational interests while attending a seminar at the Carnegie Institute of Technology (**Campbell, 1971**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib256>) ; **Strong, 1927** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1592>)). In constructing the SVIB, Strong employed two little-used techniques in measurement. First, the examinee was asked to express liking or disliking for a large and varied sample of occupations, educational disciplines, personality types, and recreational activities. Second, the responses were empirically keyed for specific occupations. In an empirical key, a specific response (e.g., liking to roller skate) is assigned to the scale for a particular occupation only if successful persons in that occupation tend to answer in that manner more often than comparison subjects.

Although Strong did not express his underlying assumptions in a simple and straightforward manner, it is clear that the theoretical foundation for the SVIB derives from a typological, trait-oriented conception of personality. Tzeng (**1987**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1673>)) has identified the following basic assumptions in the development and application of the SVIB:

- Each occupation has a desirable pattern of interests and personality characteristics among its workers. The ideal pattern is represented by successful people in that occupation.
- Each individual has relatively stable interests and personality traits. When such interests and traits match the desirable interest patterns of the occupation the individual has a high probability to enter that occupation and be more likely to succeed in it.
- It is highly possible to differentiate individuals in a given occupation from others-in-general in terms of the desirable patterns of interests and traits for that occupation.

Strong constructed the scales of his inventory by contrasting the responses of several specific occupational criterion groups with those of a people-in-general group. The subjects for each criterion group were workers in that occupation who were satisfied with their jobs and who had been so employed for at least three years. The items that differentiated the two groups, keyed in the appropriate direction, were selected for each occupational scale. For example, if members of a specific occupational group disliked "buying merchandise for a store" more often than people-in-general, then that item (keyed in the dislike direction) was added to the scale for that occupation.

The first SVIB consisted of 420 items and a mere handful of occupational scales (**Strong, 1927** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1592>)). Separate editions for men and women followed shortly. The inventory has undergone numerous revisions over the years (**Tzeng, 1987** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1673>)), culminating in the modern instrument known as the Strong Interest Inventory-Revised (**Campbell, 1974** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib257>) ; **Hansen, 1992** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib693>) ; **Hansen & Campbell, 1985** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib694>) ; **Donnay, et al., 2004** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib433>)).

Although the Strong Interest Inventory (SII-R) was fashioned according to the same philosophy as the SVIB, the latest revision departs from its predecessors in a number of ways. The SII-R was developed with the following goals in mind:

- Shorten the instrument
- Add current occupations
- Increase the level of business, technology, and teamwork measures
- Broaden the assessment of work and leisure activities
- Reflect the diversity of the U.S. workforce in the samples obtained

The SII-R consists of 291 items answered in a 5-point Likert format, with options of *Strongly Like*, *Like*, *Indifferent*, *Dislike*, *Strongly Dislike*. The standardization sample ($N = 2,250$) consists of an equal number of employed men and women from the U.S. workforce. The sample is restricted to employed persons because the main purpose of the test is to determine interest patterns within occupational groups. Racial and ethnic groups accurately represent the U.S. population and constitute 30 percent of the sample.

Test results are organized in six sections. At the most global level are the six General Occupational Theme scores, namely, Realistic, Investigative, Artistic, Social, Enterprising, and Conventional. These scores are based on the theoretical analysis of Holland (**1966**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib763>) , **1985** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib765>)), whose work was discussed earlier. Each theme score pertains to a major interest area that describes both a work environment and a type of person. For example, persons scoring high on the Realistic theme are generally quite robust, have difficulty expressing their feelings, and prefer to work outdoors with heavy machinery.

The 30 Basic Interest Scales are found within the general theme scores. These identify specific interest domains, indicating areas likely to be stimulating and rewarding to the client. Examples of these scales include Counseling and Helping, Visual Arts and Design, Marketing and Advertising, Finance and Investing, Medical Science, and Mechanics and Construction. The interest scales are empirically derived and consist of substantially intercorrelated items.

The most detailed results consist of 130 Occupational Scales, with separate normative data for each gender. Scores on these scales indicate the similarity of people of the client's gender who have been working in, and are satisfied with, the listed occupation. Each scale produced at least a one standard deviation separation between the occupational sample and the reference sample, supporting the distinctiveness of specific career paths (**Donnay et al., 2004** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib433>)).

The SII-R also yields five Personal Style Scales. These are designed to measure preferences for broad styles of living and working. These scales assist in vocational guidance by showing the level of comfort with distinctive styles. The five style scales are as follows:

1. *Work Style*, on which a high score indicates a preference to work with people and a low score signifies an interest in ideas, data, and things;
2. *Learning Environment*, on which a high score indicates a preference for academic learning environments and a low score indicates a preference for more applied learning activities;
3. *Leadership Style*, on which a high score indicates comfort in taking charge of others and a low score indicates uneasiness; and
4. *Risk Taking/Adventure*, on which a high score indicates a preference for risky and adventurous activities as opposed to safe and predictable activities; and
5. *Team Orientation*, on which a high score indicates a preference for collaboration and working on teams as opposed to working independently.

The personal style scales each have a mean of 50 and a standard deviation of 10. Note that these are truly bipolar scales for which each pole is distinct and meaningful.

The SII-R can only be scored by prepaid answer sheets or booklets that are mailed or faxed to the publisher, or through purchase of a software system that provides on-site scoring for immediate results. The results consist of a lengthy printout that is organized according to several themes. All scores are expressed as standard scores with a mean of 50 and an SD of 10.

Evaluation of the SII-R

The SII-R represents the culmination of over 70 years of study, involving literally thousands of research reports and hundreds of thousands of respondents. In evaluating this instrument, we can only outline basic trends in the research, referring the reader to other sources for details (**Bailey, Larson, Borgen, & Gasser, 2008** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib76>) ; **Savickas, Taber, & Spokane, 2002**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1441>) ; **Hansen, 1992** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib693>) ; **Hansen & Campbell, 1985** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib694>)).

We should also point out that evaluations of the reliability and validity of the SII-R are based in part upon its similarity to the SII and SVIB, for which a huge amount of technical data exists.

Based upon test-retest studies, the reliability of the Strong has proved to be exceptionally good in the short run, with one- and two-week stability coefficients for the occupational scales generally in the .90s. When the test-retest interval is years or decades, the correlations drop to the .60s and .70s for the occupational scales, except for respondents who were older (over age 25) upon first testing. For younger respondents first tested as adolescents, the median test-retest correlation after 15 years is around .50 (**Lubinski, Benbow, & Ryan, 1995**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1011>)). But for older respondents, first tested after the age of 25, the median test-retest correlation 10 to 20 years later is a phenomenal .80 (**Campbell, 1971**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib256>)). Apparently, by the time we pass through young adulthood, personal interests become extremely stable. The questions on the SII-SVIB capture that stability in the occupational scores, providing support for the trait conception of personality upon which these instruments were based.

The validity of the Strong is premised largely on the ability of the initial occupational profile to predict the occupation eventually pursued. Strong (1955

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1593>) reported that the chances were about two in three that people would be in occupations predicted by high occupational scale scores, and about one in five that respondents would be in occupations for which they had shown little interest when tested. Although other researchers have quibbled with the exact proportions

(Dolliver, Irvin, & Bigley, 1972

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib427>), it is clear that the SII-SVIB has impressive hit rates in predicting occupational entry. The instrument functions even better in predicting the occupations that an examinee will *not* enter. In a recent study, Donnay and Borgen (1996 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib432>)) provide evidence for construct validity by demonstrating strong overall differentiation between 50 occupational groups on the SII:

The big picture is that people in diverse occupations show large and predictable differences in likes and dislikes, whether in terms of vocational interests or in terms of personal styles. And the Strong provides valid, structural, and comprehensive measures of these differences. (p. 290)

The SII-R is used mainly with high school and college students and adults seeking vocational guidance or advice on continued education. Because most students' interests are undeveloped and unstabilized prior to age 13 or 14, the test is not recommended for use below high-school level. As evident in the reliability data reported, the SII-R becomes increasingly valuable with older subjects, and it is not unusual to see middle-aged persons use the results of this instrument for guidance in career change.

Vocational Preference Inventory

The Vocational Preference Inventory is an objective, paper-and-pencil personality interest inventory used in vocational and career assessment (Holland, 1985c

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib765>)). The VPI measures 11 dimensions, including the six personality-environment themes of Realistic, Investigative, Artistic, Social, Enterprising, and Conventional, and five additional dimensions of Self-Control, Masculinity/Femininity, Status, In-frequency, and Acquiescence. The test items consist of 160 occupational titles toward which the examinee expresses a feeling by marking *y* (yes) or *n* (no). The VPI is a brief test (15 to 30 minutes) and is intended for persons 14 years and older with normal intelligence.

As noted previously, Holland proposes that personality traits tend to cluster into a small number of vocationally relevant patterns, called types. For each personality type there is also a corresponding work environment best suited to that type. According to Holland, there are six types: Realistic, Investigative, Artistic, Social, Enterprising, and Conventional. This is sometimes known as the **RIASEC model**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm01#bm01gloss280>), in reference to the first letters of the six types.

Test-retest reliability coefficients for the six major scales range from .89 to .97. VPI norms are based upon large convenience samples of college students and employed adults from earlier VPI editions. The characteristics of the standardization sample are not well defined, which makes the norms somewhat difficult to interpret (Rounds, 1985).

The validity of the VPI is essentially tied to the validity of Holland's (1985a

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib766>)) hexagonal model of vocational interests. Literally hundreds of studies have examined this model from different

perspectives. We will cite trends and representative studies. The reader is referred to Holland (1985c (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib768>)) and Walsh and Holland (1992 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1711>)) for more details.

Several VPI studies have investigated a key assumption of Holland's theory—that individuals tend to move toward environments that are congruent with their personality types. If this assumption is correct, then the real-world match between work environments and personality types of employees should be substantial. We should expect to find that Realistic environments have mainly Realistic employees, Social environments have mainly Social employees, and so on. Research on this topic has followed a straightforward methodology: Subjects are tested with the VPI and classified by their Holland types (using up to six letters); the work environments of the subjects are then independently classified by an appropriate environmental measure; finally, the degree of congruence between persons and environments is computed. In better studies, a correction for chance agreement is also applied.

Using his hexagonal model, Holland has developed occupational codes as a basis for classifying work environments (Gottfredson & Holland, 1989

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib626>) ; Holland, 1966

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib763>) , 1978

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib764>) , 1985c

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib768>)). For example, landscape architect is coded as RIA (Realistic, Investigative, Artistic) because this occupation is known to be a technical, skilled trade (Realistic component) that requires scientific skills (Investigative component) and also demands artistic aptitude (Artistic component). The Realistic component is listed first because it is the most important for landscape architect, whereas the Investigative and Artistic components are of secondary and tertiary importance, respectively. Some other occupations and their codes are taxi driver (RSE), mathematics teacher (ISC), reporter (ASE), police officer (SRE), real estate appraiser (ECS), and secretary (CSA). In a similar manner, Holland has also worked out codes for different college majors.

One approach to congruence studies is to compare VPI results of students or workers with the Holland codes that correspond to their college majors or occupations. For example, VPI Holland codes for a sample of police officers should consist mainly of profiles that begin with S and should contain a larger-than-chance proportion of specifically SRE profiles. Furthermore, the degree of congruence should be related to the degree of expressed satisfaction with that line of work or study.

Research with college students provides strong support for the congruence prediction: Students tend to select and enter college majors that are congruent with their primary personality types (Holland, 1985a (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib766>) ; Walsh & Holland, 1992 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1711>)). Thus, Artistic types tend to major in art, Investigative types tend to major in biology, and Enterprising types tend to major in business, to cite just a few examples. These results provide strong support for the VPI and the theory upon which it is based.

This short review has barely touched the surface of supportive validity studies with the VPI. Walsh and Holland (1992 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1711>)) cite several additional lines of research that buttress the validity of this test. But not all studies of the VPI affirm its validity. Furnham, Toop, Lewis, and Fisher (1995

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib557>)) failed to find a relationship between personality–environment (P-E) “fit” and job satisfaction, a key theoretical underpinning of the test. According to Holland's theory, the better the P-E fit, the greater should be job

satisfaction. In three British samples, the relationships were weak or nonexistent, suggesting that the VPI does not “travel well” in cultures outside of the United States.

Self-Directed Search

Holland has always shown a keen interest in the practical applications of his research on vocational development. Consistent with this interest, he developed the Self-Directed Search, a highly practical, brief test that is appealing in its simplicity (**Holland, 1985a**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib766>) , b

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib767>)). As the name suggests, the Self-Directed Search is designed to be a self-administered, self-scored, and self-interpreted test of vocational interest. The SDS measures the six RIASEC vocational themes described previously.

The SDS consists of dichotomous items that the examinee marks “like” or “dislike” (or “yes” or “no”) in four sections: (1) Activities (six scales of 11 items each); (2) Competencies (six scales of 11 items each); (3) Occupations (six scales of 14 items each); and (4) Self-Estimates (two sets of six ratings). For each section, the face-valid items are grouped by RIASEC themes. For each theme, the total number of “like” and “yes” answers is combined with the self-estimates of ability to come up with a total theme score. The SDS takes 30 to 50 minutes for completion and is intended for persons 15 years and older.

The RIASEC themes on the SDS showed test-retest reliabilities that range from .56 to .95 and internal consistencies that range from .70 to .93. Norms for SDS scales and codes are reported for pooled convenience samples of 4,675 high school students, 3,355 college students, and 4,250 employed adults ages 16 through 24 (**Holland, 1985a**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib766>) , b

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib767>)). However, SDS results are typically interpreted in an individualized, ipsative manner (“Is this occupation a good fit for this client?”), so normative data are of limited relevance.

The SDS is available in a hand-scored paper-and-pencil version and a computerized version as well. Unfortunately, the paper-and-pencil version is prone to a 16 percent clerical error rate when used by high school students (**Holland, 1985a**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib766>) , b

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib767>)). The user-friendly microcomputer test is probably the preferred version because of the ease of administration and the error-free scoring and interpretation.

When a subject takes the SDS, the three highest theme scores are used to denote a summary code. For example, a person whose three highest scores were on Investigative, Artistic, and Realistic would have a summary code of IAR. In a separate booklet distributed with the test—the *Occupations Finder*—the examinee can look up his or her summary code and find a list of occupations that provide the best “fit.” For example, an examinee with an IAR summary code would learn that he or she most closely resembles persons in the following occupations: anthropologist, astronomer, chemist, pathologist, and physicist. The test booklet contains additional information, which helps the examinee explore relevant career options.

The SDS serves a very useful purpose in providing a quick and simple format for prompting young persons to examine career alternatives. By eliminating the time-consuming process of administration, scoring, interpretation, and counselor feedback, the test makes it possible for a wide audience to receive an introductory level of career counseling. Holland (**1985a**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib766>) , b (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib767>)) proposes that the SDS is appropriate for up to 50 percent of students and adults who might desire career guidance. Presumably, the other 50 percent would find the SDS an insufficient basis for career exploration. Holland (1985a (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib766>) , b (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib767>)) rightfully warns users to consider many sources of information in career choice and not to rely too heavily on test scores per se. Levinson (1990 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib973>)) discusses the integration of SDS data with other psychoeducational data to make specific vocational recommendations for high school students.

LaBarbera (2005 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib938>)) illustrates the potential application of the SDS in a study of 463 physician assistants (PAs) known to be well satisfied with their work. The PAs are medical professionals who provide care under the supervision of a licensed physician. This is a demanding profession with well defined duties that include many of the same functions provided by a general practitioner. Who is a good candidate for this up-and-coming profession in high demand? LaBarbera (2005 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib938>)) determined that the Holland profile was a distinctive SIR for men, especially those with interests in surgery, whereas the profile for women maintained the first two letters (SI) but yielded a muddle for the third theme. This is valuable information for prospective students and career counselors.

The validity of the SDS is linked to the validity of the hexagonal model of personality and environments upon which the test is based. One aspect of validity, then, is whether the model makes predictions that are confirmed by SDS results in the real world. In general, the results from over 400 studies support the construct validity of the SDS (Dumenci, 1995

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib439>) ; Holland, 1985a (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib766>) , b (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib767>) , 1987 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib769>)).

One approach to construct validity is to determine whether the relationships among SDS scales make theoretical sense. One tenet of construct validity is that similar scales should reveal stronger relationships, dissimilar scales weaker relationships. For example, it is not hard to imagine one person combining Artistic and Investigative themes in personality and work environment. After all, these themes are mildly similar, so we would predict a moderately positive correlation between them. This is exactly what Holland (1985a

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib766>) , b (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib767>)) found. In a general reference sample of 175 women aged 26 to 65 years, scores on these two themes correlated modestly, $r = .26$, as would be predicted. Further, unrelated themes like Investigative and Enterprising (which bear little in common) should reveal a weak correlation. The value turned out to be a negligible $r = -.02$. Overall, the various correlations among the six themes of the SDS make theoretical sense, which supports the construct validity of the test.

The predictive validity of the SDS has been investigated in several dozen studies, which are summarized by Holland (1985a (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib766>) , b (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib767>) , 1987

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib769>)). The typical methodology for these studies is that SDS high-point codes for large samples of students are compared with the first letter of their occupational choices (or aspirations) one to three years later. Overall, the findings indicate that the SDS has moderate to high predictive efficiency, depending upon the age of the sample (hit rates go up with age), the length of the time interval (hit rates go down with time), and the specific category predicted (hit rates are better for Investigative and Social predictions) (**Gottfredson & Holland, 1975** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib625>)).

Campbell Interest and Skill Survey

The Campbell Interest and Skill Survey (CISS; **Campbell, Hyne, & Nilsen, 1992**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib258>)) is a newer measure of self-reported interests and skills. The test is designed to help individuals make better career choices by describing how their interests and skills match the occupational world. The primary target population for the CISS is students and young adults who have not entered the job market, but the test is also suitable for older workers who are considering a change in careers. The test is appropriate for persons 15 years of age and older with a sixth-grade reading level, although younger children can be tested in exceptional circumstances.

The CISS consists of 200 interest items and 120 skill items. The interest items include occupations, school subjects, and varied working activities that the examinee rates on a six-point scale from strongly like to strongly dislike. The interest items resemble the following:

- A pilot, flying commercial aircraft
- A biologist, working in a research lab
- A police detective, solving crimes

The skill items include a list of activities that the examinee rates on a six-point scale from expert (widely recognized as excellent in this area) to none (have no skills in this area). The skill items resemble the following:

- Helping a family resolve its conflicts
- Making furniture, using woodworking and power tools
- Writing a magazine story

CISS results are scored on several different kinds of scales: Orientation Scales, Basic Interest and Skill Scales, Occupational Scales, Special Scales, and Procedural Checks. All scale scores are reported as *T* scores, normed to a population average of 50, with a standard deviation of 10.

The Orientation Scales serve to organize the CISS profile—the interest, skill, and occupational scales are reported under the appropriate Orientations. The seven Orientations are as follows (**Campbell et al., 1992** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib258>) , pp. 2-3):

- *Influencing*—influencing others through leadership, politics, public speaking, and marketing
- *Organizing*—organizing the work of others, managing, and monitoring financial performance
- *Helping*—helping others through teaching, healing, and counseling
- *Creating*—creating artistic, literary, or musical productions, and designing products or environments
- *Analyzing*—analyzing data, using mathematics, and carrying out scientific experiments
- *Producing*—producing products, using “hands-on” skills in farming, construction, and mechanical crafts

- *Adventuring*—adventuring, competing, and risk taking through athletic, police, and military activities

There are 29 pairs of Basic Scales, each pair consisting of parallel interest and skill scales. The Basic Scales are clustered within the seven Orientations, based upon their intercorrelations. For example, the Helping Orientation contains the following Basic Scales, each with separate interest and skill components: Adult Development, Counseling, Child Development, Religious Activities, and Medical Practice.

The 58 pairs of Occupational Scales, each with separate interest and skill components, provide feedback on the degree of similarity between the examinee and satisfied workers in that occupation. These scales were constructed empirically by contrasting the responses of happily employed persons in specific occupations with responses of a general reference sample drawn from the working population at large.

In addition to Basic and Occupational Scales, the CISS incorporates three special scales: Academic Focus, a measure of interest and confidence in intellectual, scientific, and literary activities; Extraversion, a measure of social extraversion; and Variety, a measure of the examinee's breadth of interests and skills. Finally, the CISS reports a variety of Procedural Checks to detect possible problems in test taking such as random responding or excessive omissions.

Overall, the reliability of CISS scales is exceptionally strong. For example, coefficient alpha for the Orientation Scales is typically in the high .80s, and three-month test-retest reliabilities for 324 respondents are in the mid- to high .80s. Similar findings for reliability are reported for the Basic and Occupational Scales. Norms for the CISS are based upon 5,000 subjects spread over the 58 occupations. The authors report extensive validity data for the Occupational Scales, including sample means for each occupational sample as well as lists of the three highest- and lowest-scoring occupations for each scale (Campbell et al., 1992

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib258>)). These data document that the scales do discriminate between occupations in an effective and meaningful way. For example, the average *T* score on accountant by accountants is 75.8. Statisticians, bookkeepers, and financial planners achieve the next three highest scores for this scale, with average *T* scores in the low 60s. Commercial artists, professors, and social workers obtain the three lowest scores, with average *T* scores around 40. Because these results fit well with our expectations about occupational interest and skill patterns, they provide support for the validity of the CISS.

Independent correlational studies also support the validity of the CISS. For example, in a sample of 221 college students, Hansen (2007

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib692>)) correlated CISS Skill Scale scores with SII scores and found strong evidence for convergent and discriminant validity (i.e., strong correlations with similar scales, negligible correlations with dissimilar scales). In a sample of 118 adults, Savickas et al. (2002

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1441>)) correlated scores from individual occupational scales of the CISS with scores from the scales of other mainstream instruments such as the Strong Interest Inventory. They also found strong support for both convergent validity (i.e., modest correlations for same-named pairs of scales) and discriminant validity (i.e., negligible correlations for unlike pairs of scales). In a sample of 128 college students, Hansen and Neuman (1999 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib695>)) confirmed the concurrent validity of the CISS by finding a good fit between occupational scale scores and students' chosen majors. The fit was considered "excellent" or "moderately good" for more than 70 percent of the students. Boggs (1999

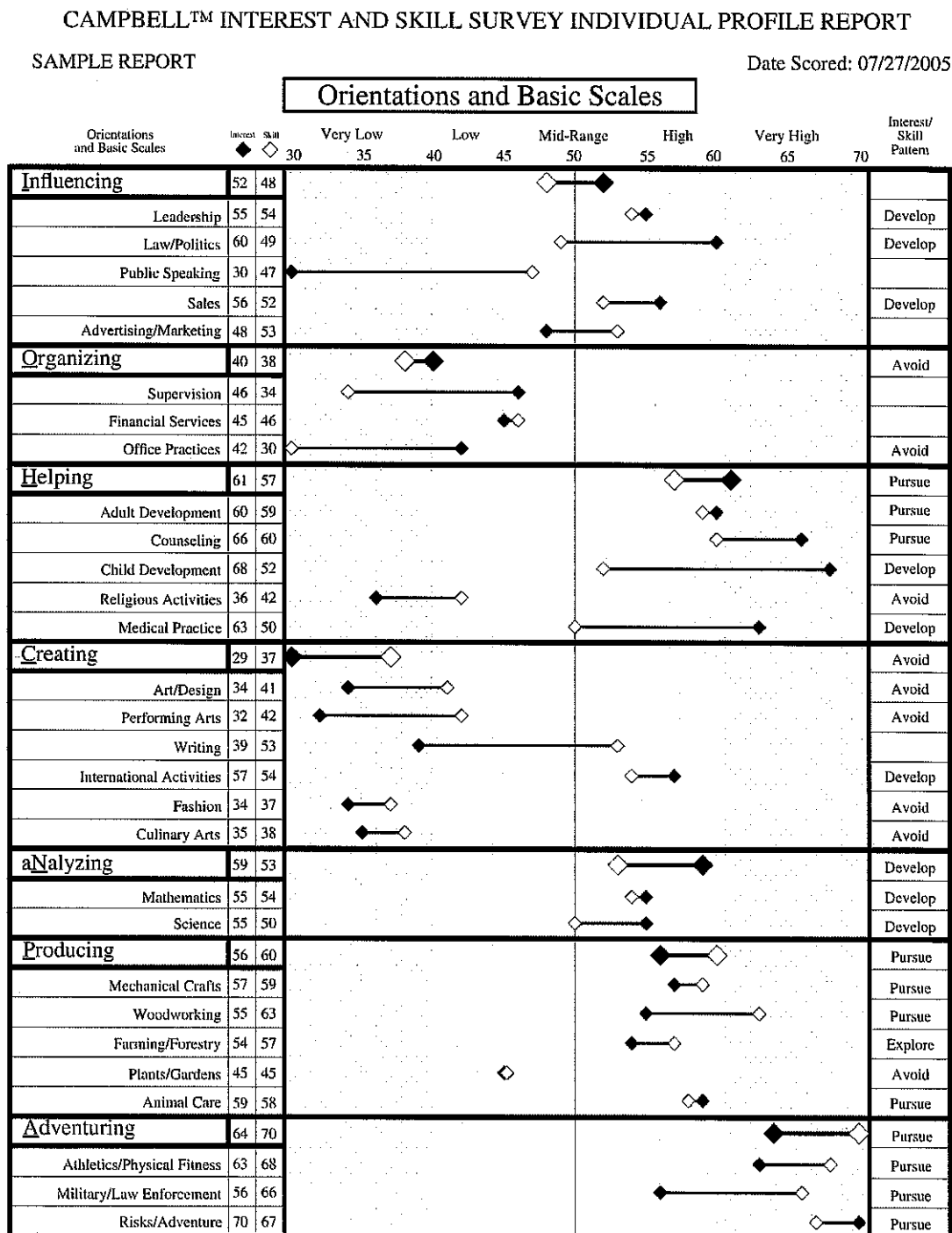
(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib175>)) provides a review

and critique of the CISS. Campbell (2002

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib255>) presents the history and development of the instrument.

This instrument will almost certainly receive increased attention in the years ahead. One noteworthy feature of the CISS is the comprehensiveness and clarity of the profile report form. The report consists of 11 user-friendly pages. We have reprinted two pages in **Figure 11.4**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch11lev1sec19#ch11fig4>) for illustrative purposes. This format is preferable to the detail-rich but eye-straining graphs encountered with many instruments. The CISS promises to rival the Strong Interest Inventory for vocational guidance of young adults.



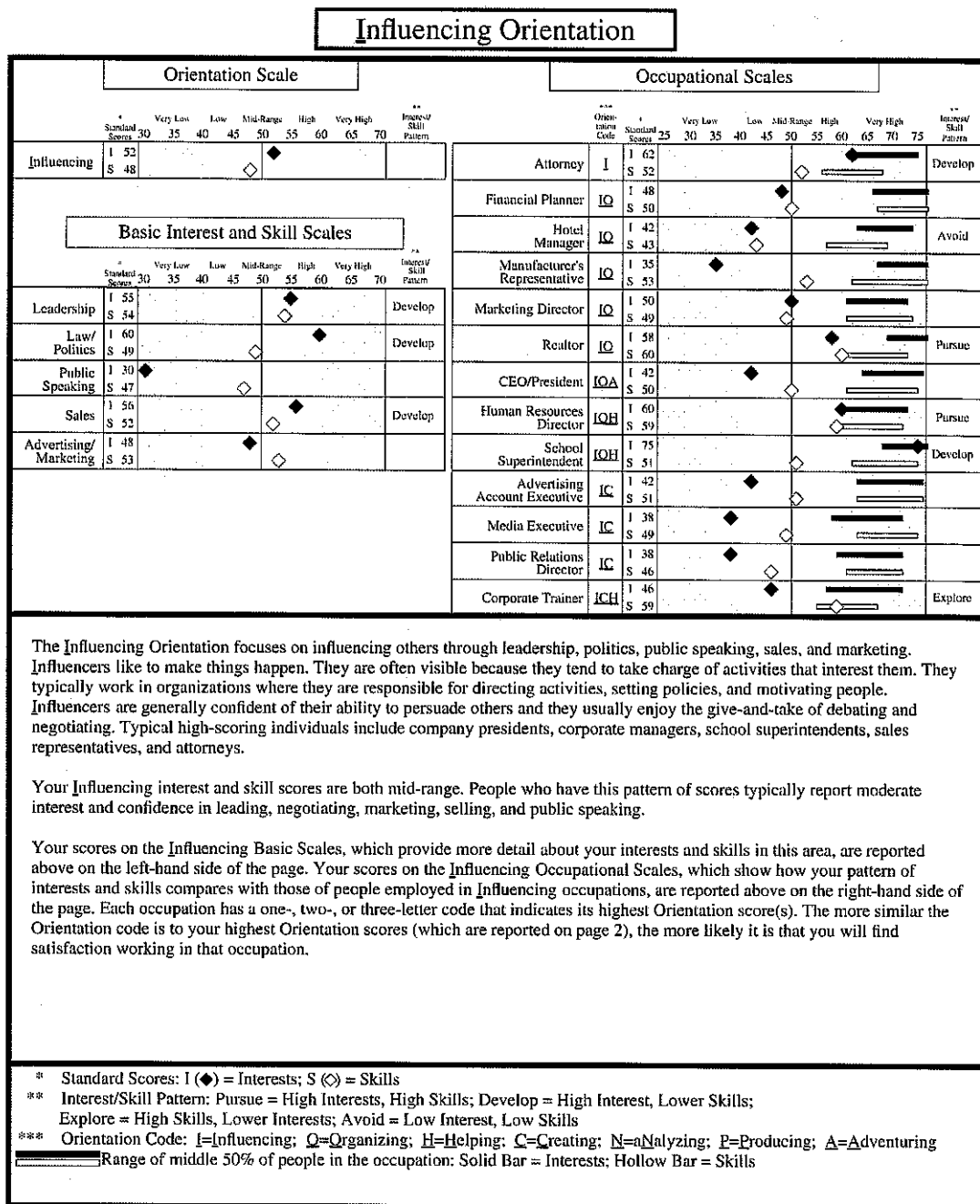


FIGURE 11.4 Representative Sections from the Campbell Interest and Skill Survey

Note: The full profile consists of an 11-page printout.

Source: From Campbell Interest and Skill Survey (CISS). Copyright © 1997 David Campbell, Ph.D. Reproduced with permission of the publisher NCS Pearson, Inc. All rights reserved. "Campbell" and "CISS" are trademarks, in the US and/or other countries, of Pearson Education, INC. or its affiliates.



Minnesota Multiphasic
Personality Inventory-2
Restructured Form®

SAMPLE REPORT

Case Description: Mr. C Police Candidate Interpretive Report

Mr. C is a 34-year-old, married male candidate for an entry-level police officer position in a large urban agency. His background revealed some relatively minor juvenile conduct problems. After graduating high school, he married at age 18 but divorced within three years. He then enlisted in the military and was deployed twice to Iraq. Since his discharge four years prior to this evaluation, Mr. C has worked exclusively in construction jobs. The background investigation revealed no adult history of job terminations or legal conflicts. Personal and developed references described him as introverted.

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. C'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/mmapi2rf.

Copyright © 2015 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. 9594 01/15



Minnesota Multiphasic
Personality Inventory-2
Restructured Form®

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

MMPI-2-RF®

Police Candidate Interpretive Report

David M. Corey, PhD, & Yossef S. Ben-Porath, PhD

ID Number:	Mr. C
Age:	34
Gender:	Male
Marital Status:	Married
Years of Education:	14
Date Assessed:	11/11/2013



Copyright © 2014 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.

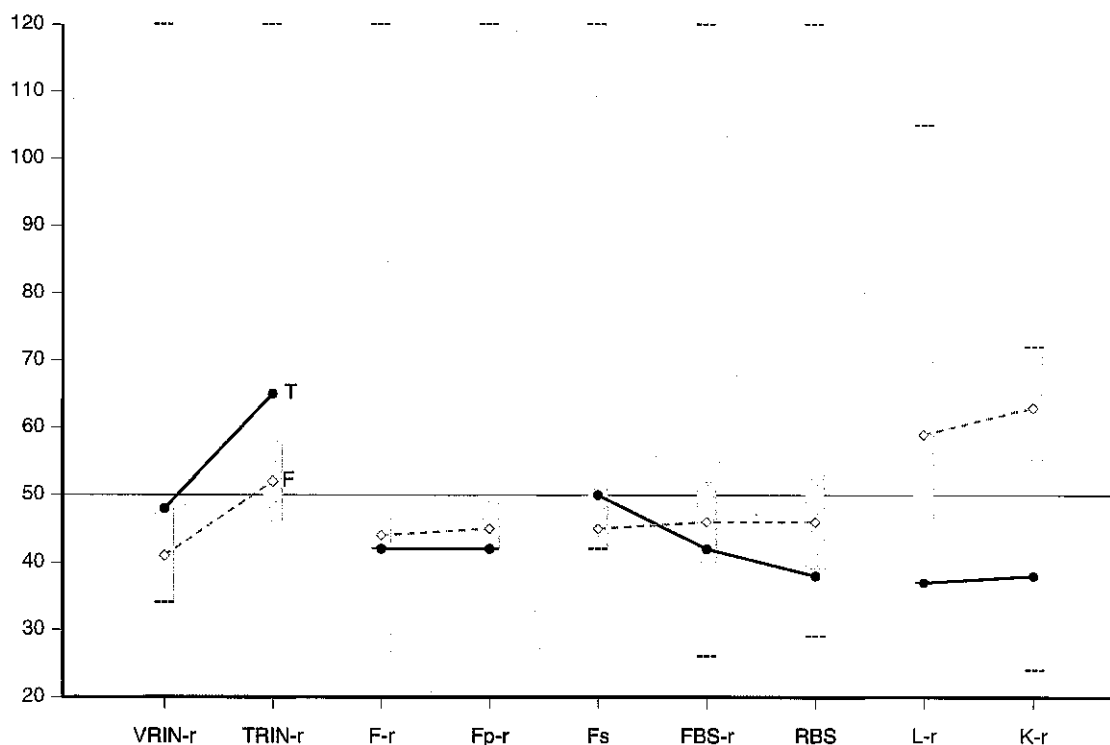
MMPI-2-RF, the MMPI-2-RF logo, and Minnesota Multiphasic Personality Inventory-2-Restructured Form are registered trademarks of the University of Minnesota. Pearson, the PSI logo, and PsychCorp are trademarks in the U.S. and/or other countries of Pearson Education, Inc., or its affiliate(s).

TRADE SECRET INFORMATION

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

[3.0 / 56 / 0.0.0]

MMPI-2-RF Validity Scales



Raw Score:	3	13	0	0	1	5	2	0	4
T Score:	48	65 T	42	42	50	42	38	37	38
Response %:	100	100	100	100	100	100	100	100	100
Cannot Say (Raw):	0								
					Percent True (of items answered):				42%

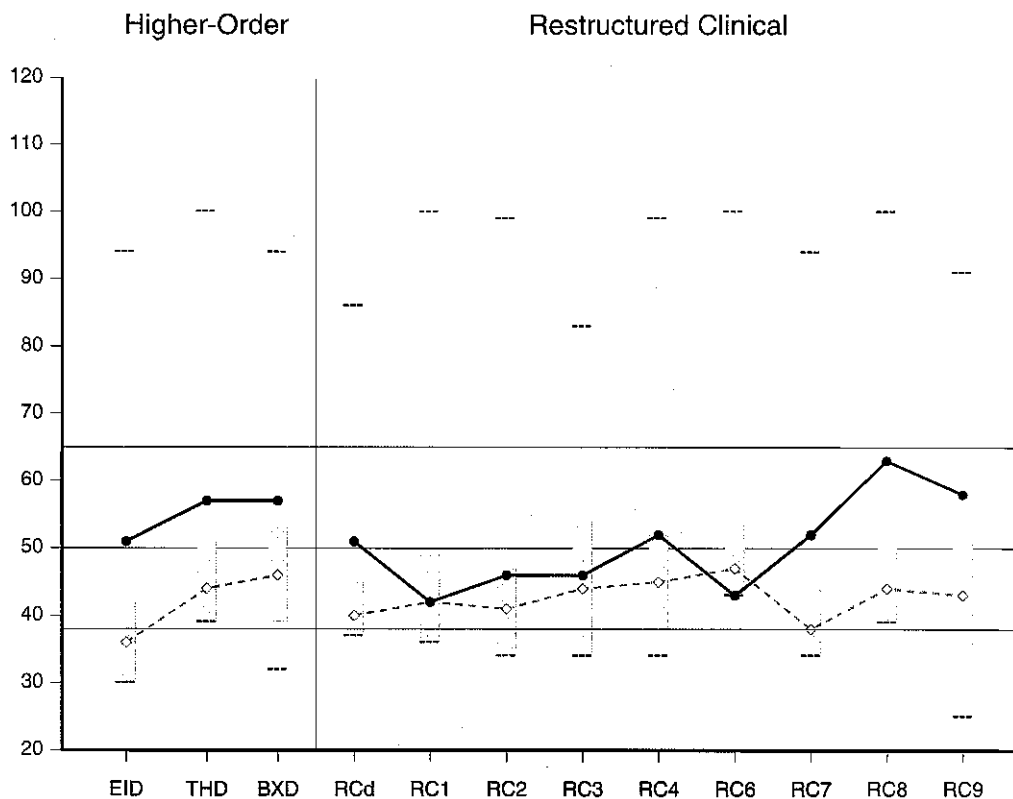
Comparison Group Data: Police Officer Candidate (Men and Women), N = 2,074

Mean Score (◇--◇):	41	52 F	44	45	45	46	46	59	63
Standard Dev (±1 SD):	7	6	4	5	6	6	7	13	8
Percent scoring at or below test taker:	92	99	75	78	92	32	20	4	0.7

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:	10	3	9	4	1	3	4	5	0	7	5	16
T Score:	51	57	57	51	42	46	46	52	43	52	63	58
Response %:	100	100	100	100	100	100	100	100	100	100	100	100

Comparison Group Data: Police Officer Candidate (Men and Women), N = 2,074

Mean Score (◇---◇):	36	44	46	40	42	41	44	45	47	38	44	43
Standard Dev (±1 SD):	6	7	7	5	7	6	10	7	7	6	7	8
Percent scoring at or below test taker:	98	96	96	97	68	88	65	87	75	97	98	97

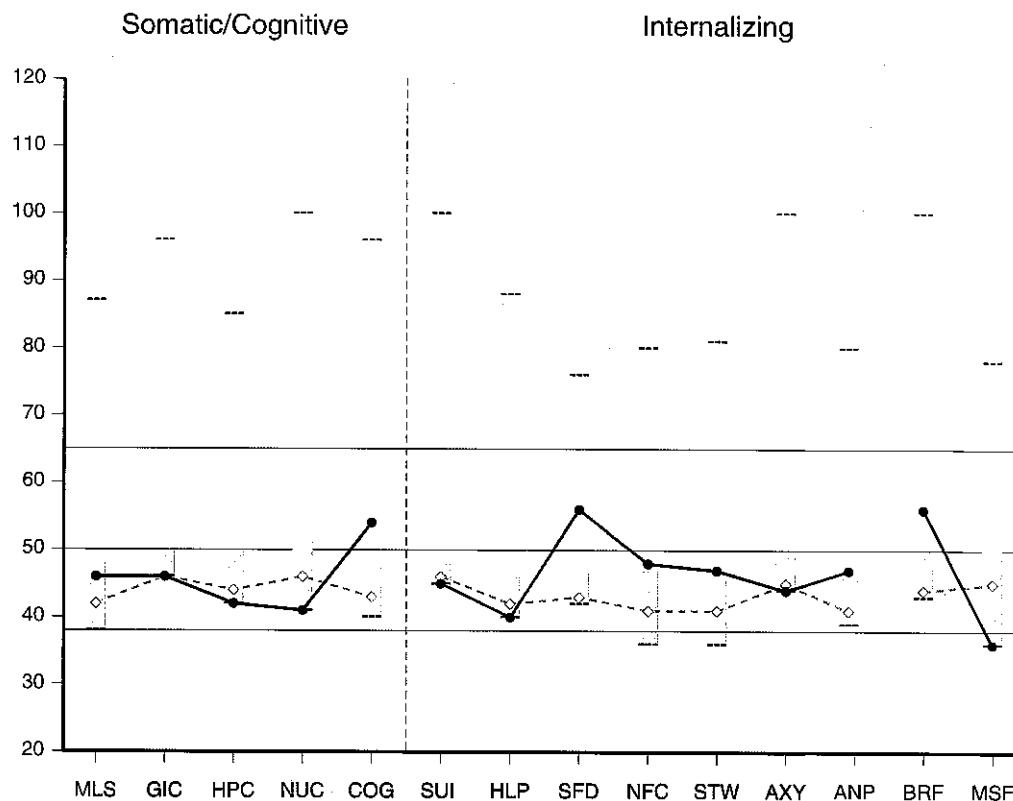
The highest and lowest T scores possible on each scale are indicated by a "---"; MMPI-2-RF T scores are non-gendered.

EID Emotional/Internalizing Dysfunction
THD Thought Dysfunction
BXD Behavioral/Externalizing 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

MMPI-2-RF Somatic/Cognitive and Internalizing Scales



Raw Score:	1	0	0	0	2	0	0	2	2	2	0	1	1	0
T Score:	46	46	42	41	54	45	40	56	48	47	44	47	56	36
Response %:	100	100	100	100	100	100	100	100	100	100	100	100	100	100

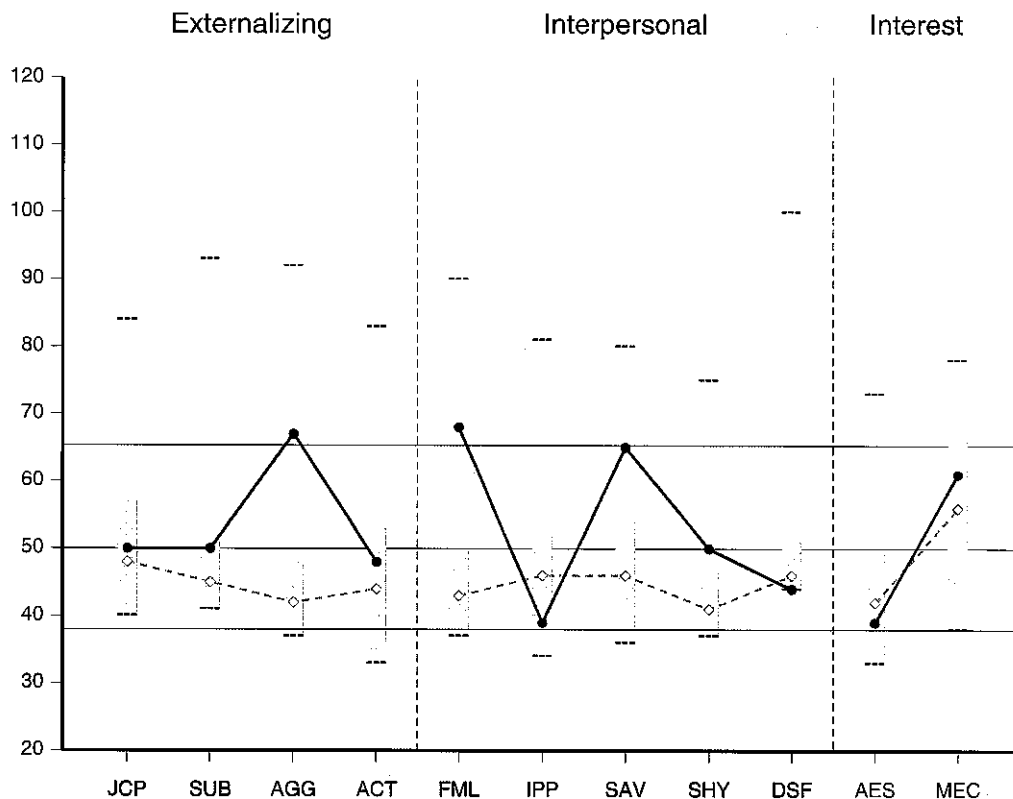
Comparison Group Data: Police Officer Candidate (Men and Women), N = 2,074

Mean Score (◇---◇):	42	46	44	46	43	46	42	43	41	41	45	41	44	45
Standard Dev (±1 SD):	6	4	6	7	5	2	4	4	6	6	4	5	5	8
Percent scoring at or below test taker:	87	95	83	65	97	99.3	88	99.1	91	91	94	94	98	25

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	1	5	3	6	1	7	3	0	1	5
T Score:	50	50	67	48	68	39	65	50	44	39	61
Response %:	100	100	100	100	100	100	100	100	100	100	100

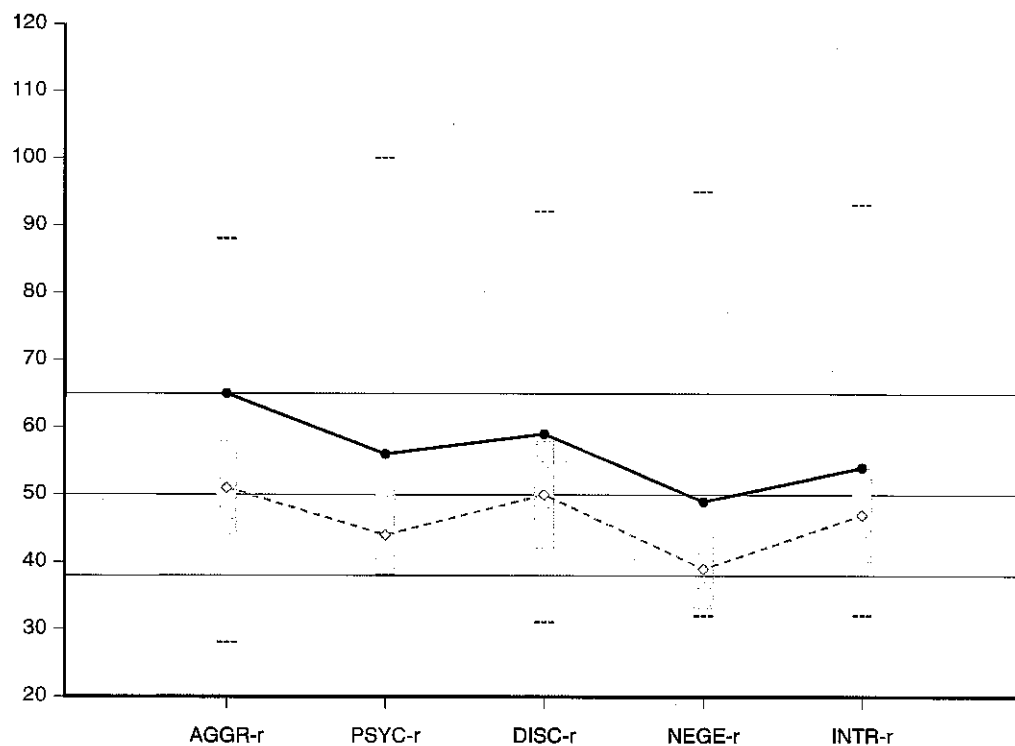
Comparison Group Data: Police Officer Candidate (Men and Women), N = 2,074

Mean Score (◇---◇):	48	45	42	44	43	46	46	41	46	42	56
Standard Dev (±1 SD):	9	6	6	9	7	6	8	6	5	8	11
Percent scoring at or below test taker:	76	90	99.9	79	99.9	20	98	95	90	57	73

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:	13	3	10	6	8
T Score:	65	56	59	49	54
Response %:	100	100	100	100	100

Comparison Group Data: Police Officer Candidate (Men and Women), N = 2,074

Mean Score (◇---◇):	51	44	50	39	47
Standard Dev (±1 SD):	7	7	8	6	7
Percent scoring at or below test taker:	98	95	93	95	89

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	0	48	65 T			
	CNS	VRIN-r	TRIN-r			
Over-Reporting	42	42		50	42	38
	F-r	Fp-r		Fs	FBS-r	RBS
Under-Reporting	37	38				
	L-r	K-r				

SUBSTANTIVE SCALES

Somatic/Cognitive Dysfunction	42	46	46	42	41	54
	RC1	MLS	GIC	HPC	NUC	COG
Emotional Dysfunction	51	51	45	40	56	48
	EID	RCd	SUI	HLP	SFD	NFC
		46	54			
		RC2	INTR-r			
		52	47	44	47	56
		RC7	STW	AXY	ANP	BRF
						36
						MSF
						49
						NEGE-r
Thought Dysfunction	57	43				
	THD	RC6				
		63				
		RC8				
		56				
		PSYC-r				
Behavioral Dysfunction	57	52	50	50		
	BXD	RC4	JCP	SUB		
		58	67	48	65	59
		RC9	AGG	ACT	AGGR-r	DISC-r
Interpersonal Functioning	68	46	39	65	50	44
	FML	RC3	IPP	SAV	SHY	DSF
Interests	39	61				
	AES	MEC				

Scale scores shown in bold font are interpreted in the report.

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 in the context of preemployment psychological evaluations of police and other law enforcement officer candidates. It focuses on identifying problems; it does not convey potential strengths. The information it contains should be considered in the context of the test taker's background, the demands of the position under consideration, the clinical interview, findings from supplemental tests, and other relevant information.

The interpretive statements in the Protocol Validity section of the report are based on T scores derived from the general MMPI-2-RF normative sample, as well as scores obtained by the multisite sample of 2,074 individuals that make up the Police Officer Candidate comparison group.

The interpretive statements in the Clinical Findings and Diagnostic Considerations sections of the report are based on T scores derived from the general MMPI-2-RF normative sample. Following recommended practice, only T scores of 65 and higher are considered clinically significant. Scores at this clinical level are generally rare among police officer candidates.

Statements in the Comparison Group Findings and Job-Relevant Correlates sections are based on comparisons with scores obtained by the Police Officer Candidate comparison group. Statements in these sections may be based on T scores that, although less than 65, are nevertheless uncommon in reference to the comparison group.

Sources for interpretive statements in all sections are listed in the Endnotes section of this report. See User's Guide for the MMPI-2-RF Police Candidate Interpretive Report for detailed information on report features.

SYNOPSIS

This is a valid MMPI-2-RF protocol. Scores on the substantive scales indicate clinically significant behavioral and interpersonal dysfunction. Behavioral-externalizing problems relate to aggression. Interpersonal difficulties include family problems and social avoidance.

Comparison group findings point to additional possible concerns about self-doubt, odd perceptions and beliefs, and excitation.

Possible job-relevant problems are identified in the following domains: Emotional Control and Stress Tolerance, Routine Task Performance, Decision-Making and Judgment, Feedback Acceptance, Assertiveness, Social Competence and Teamwork, Integrity, Conscientiousness and Dependability, Substance Use, and Impulse Control.

PROTOCOL VALIDITY

Content Non-Responsiveness

There are no problems with unscorable items in this protocol. The test taker responded relevantly to the items on the basis of their content.

Over-Reporting

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

Under-Reporting

The candidate's scores show no evidence of under-reporting, indicating a cooperative test-taking approach.

The test taker claimed no uncommon virtues¹. This very rare pattern of responding is found in only 4.0% of the Police Officer Candidate comparison group members.

In addition, he reported being much less well-adjusted than members of the general population². Less than 1% of the Police Officer Candidate comparison group reported this low level of psychological adjustment. As detailed later in this report, his scores on the substantive scales do indeed raise significant concerns about the candidate's psychological adjustment.

CLINICAL FINDINGS

Clinical-level symptoms, personality characteristics, and behavioral tendencies of the test taker are described in this section and organized according to an empirically guided framework. (Please see Chapter 8, Yossef S. Ben-Porath, Interpreting the MMPI-2-RF, for details.) 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 accessed with the annotation features of this report.

The test taker reports engaging in physically aggressive, violent behavior and losing control³, and is indeed likely to have a history of violent behavior toward others⁴.

The test taker reports conflictual family relationships and lack of support from family members⁵. He is indeed likely to have family conflicts and to experience poor family functioning⁶, to have strong negative feelings about family members⁷, and to blame family members for his difficulties⁷. He reports not enjoying social events and avoiding social situations⁸. He is likely to be introverted⁹, to have difficulty forming close relationships¹⁰, and to be emotionally restricted¹¹.

There are no indications of clinically significant somatic, cognitive, emotional, or thought dysfunction in this protocol.

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:

Behavioral-Externalizing Disorders

- Disorders associated with interpersonally aggressive behavior such as intermittent explosive disorder¹²

Interpersonal Disorders

- Disorders associated with social avoidance such as avoidant personality disorder¹³

COMPARISON GROUP FINDINGS

*This section describes the MMPI-2-RF substantive scale findings in the context of the Police Officer Candidate comparison group. Specific sources for each statement can be accessed with the annotation features of this report. **Job-related correlates of these results, if any, are provided in the subsequent Job-Relevant Correlates section.***

Emotional/Internalizing Problems

The test taker reports a comparatively high level of self-doubt for a police officer candidate¹⁴. Only 2.8% of comparison group members convey this or a greater lack of confidence.

Unusual Thoughts, Perceptions, and Beliefs

The test taker reports a comparatively high level of unusual thinking for a police officer candidate¹⁵. Only 8.7% of comparison group members convey such thoughts at this or a higher level. More specifically, he reports a relatively high level of odd perceptions and thoughts for a police officer candidate¹⁶. Only 3.0% of comparison group members convey this or a greater level of unusual experiences.

Behavioral/Externalizing Problems

The test taker reports a comparatively large number of behavioral problems for a police officer candidate¹⁷. Only 7.3% of comparison group members convey this or a greater level of behavioral difficulties. More specifically, he reports a relatively high level of excitability for a police officer candidate¹⁸. Only 4.8% of comparison group members convey this or a greater level of stimulation and irritable temperament or disinhibition. In particular, his responses indicate a level of physically violent behavior that may be incompatible with public safety requirements for behavioral control¹². This level of aggression is very uncommon among police officer candidates. Only 0.6% of comparison group members give evidence of this or a greater level of physically violent behavior.

Interpersonal Problems

The test taker's responses indicate a level of family problems that may be incompatible with public safety requirements for good interpersonal functioning¹⁹. This level of family conflict is very uncommon among police officer candidates. Only 0.5% of comparison group members give evidence of this or a

greater level of family problems. His responses also indicate a level of social avoidance that may impede conformance with public safety requirements for good interpersonal functioning²⁰. This level of socially avoidant behavior is very uncommon among police officer candidates. Only 4.0% of comparison group members demonstrate this or a greater level of social avoidance.

JOB-RELEVANT CORRELATES

Job-relevant personality characteristics and behavioral tendencies of the test taker are described in this section and organized according to ten problem domains commonly identified in the professional literature as relevant to police officer candidate suitability. (Please see User's Guide for MMPI-2-RF Police Candidate Interpretive Report for details.) Statements that begin with "Compared with other police officer candidates" are based on correlations with other self-report measures obtained in police officer candidate samples that included individuals who were subsequently hired as well as those who were not. Statements that begin with "He is more likely than most police officers or trainees" are based on correlations with outcome data obtained in samples of hired candidates during academy or field training, probation, and/or the post-probation period. Specific sources for each statement can be accessed with the annotation features of this report.

Emotional Control and Stress Tolerance Problems

Compared with other police officer candidates, the test taker is more likely to worry about problems and be uncertain about how to deal with them²¹ and to become impatient with others over minor infractions²².

He is more likely than most police officers or trainees to exhibit difficulties performing under stressful conditions²³.

Routine Task Performance Problems

The test taker is more likely than most police officers or trainees to exhibit difficulties carrying out tasks under non-stressful conditions²⁴.

Decision-Making and Judgment Problems

Compared with other police officer candidates, the test taker is more likely to have thoughts, perceptions, and/or experiences that are rarely reported²⁵.

Feedback Acceptance Problems

The test taker is more likely than most police officers or trainees to exhibit difficulties accepting and responding to constructive performance feedback²⁶.

Assertiveness Problems

Compared with other police officer candidates, the test taker is more likely to avoid situations that others generally view as benign and non-intimidating²⁷ and to be ill at ease in dealing with others²⁸.

He is more likely than most police officers or trainees to exhibit difficulties in demonstrating a command presence and controlling situations requiring order or resolution²⁹.

Social Competence and Teamwork Problems

Compared with other police officer candidates, the test taker is more likely to have a history of problems getting along with others²²; to be opinionated and outspoken²²; and to be demanding²². He is also more likely to have a limited social support network²⁸ and to have difficulty trusting others²².

He is more likely than most police officers or trainees to exhibit difficulties reading people, listening to others, and adapting his language and approach to the requirements of the situation²³. He is also more likely to exhibit difficulties stemming from rude and/or overbearing behavior that results in complaints from the public³⁰ and cooperating with peers and/or supervisors³¹.

Integrity Problems

Compared with other police officer candidates, the test taker is more likely to have skeptical and/or antisocial views of the world²² and to believe that life is unfair and that exploiting opportunities for personal gain is justified²².

He is more likely than most police officers or trainees to exhibit difficulties leading to integrity violations³² and sustained internal affairs investigations³².

Conscientiousness and Dependability Problems

The test taker is more likely than most police officers or trainees to exhibit difficulties reliably attending court³³; with punctuality and attendance²³; and with reliable work behavior and dependable follow-through³⁴.

Substance Use Problems

Compared with other police officer candidates, the test taker is more likely to have a history of substance use problems³⁵.

Impulse Control Problems

Compared with other police officer candidates, the test taker is more likely to behave impulsively or without adequate consideration of the consequences or implications of his actions²².

He is more likely than most police officers or trainees to exhibit difficulties reacting to situations with the proper degree of emotional and behavioral restraint and control, and avoiding impulsive and/or unnecessarily risky behavior³⁶.

ITEM-LEVEL INFORMATION

Unscorable Responses

The test taker produced scorable responses to all the MMPI-2-RF items.

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 Police Officer Candidate (Men and Women) comparison group (CG) that answered each item in the keyed direction are provided in parentheses following the item content.

Aggression (AGG, T Score = 67)

- 23. Item Content Omitted. (True; NS 39.0%, CG 7.8%)
- 26. Item Content Omitted. (True; NS 19.9%, CG 3.3%)
- 84. Item Content Omitted. (True; NS 12.1%, CG 1.8%)
- 316. Item Content Omitted. (True; NS 45.1%, CG 30.7%)
- 337. Item Content Omitted. (True; NS 50.2%, CG 17.4%)



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.

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 Police Officer Candidate (Men and Women) comparison group (CG) that answered each item in the keyed direction are provided in parentheses following the item content.

Thought Dysfunction (THD, T Score = 57)

- 12. Item Content Omitted. (True; NS 22.2%, CG 5.8%)
- 199. Item Content Omitted. (True; NS 12.1%, CG 6.7%)
- 330. Item Content Omitted. (True; NS 15.2%, CG 3.5%)

Behavioral/Externalizing Dysfunction (BXD, T Score = 57)

- 61. Item Content Omitted. (False; NS 61.6%, CG 58.7%)
- 84. Item Content Omitted. (True; NS 12.1%, CG 1.8%)
- 107. Item Content Omitted. (True; NS 47.3%, CG 35.5%)
- 131. Item Content Omitted. (True; NS 43.3%, CG 32.9%)
- 156. Item Content Omitted. (True; NS 59.8%, CG 45.9%)
- 205. Item Content Omitted. (True; NS 13.0%, CG 8.6%)
- 226. Item Content Omitted. (True; NS 21.5%, CG 38.1%)
- 237. Item Content Omitted. (False; NS 27.4%, CG 17.2%)
- 316. Item Content Omitted. (True; NS 45.1%, CG 30.7%)

Aberrant Experiences (RC8, T Score = 63)

- 12. Item Content Omitted. (True; NS 22.2%, CG 5.8%)
- 32. Item Content Omitted. (True; NS 21.1%, CG 15.0%)
- 199. Item Content Omitted. (True; NS 12.1%, CG 6.7%)
- 257. Item Content Omitted. (True; NS 12.4%, CG 4.3%)
- 330. Item Content Omitted. (True; NS 15.2%, CG 3.5%)

Hypomanic Activation (RC9, T Score = 58)

- 26. Item Content Omitted. (True; NS 19.9%, CG 3.3%)
- 39. Item Content Omitted. (True; NS 51.0%, CG 42.1%)
- 61. Item Content Omitted. (False; NS 61.6%, CG 58.7%)
- 72. Item Content Omitted. (True; NS 81.5%, CG 52.6%)
- 84. Item Content Omitted. (True; NS 12.1%, CG 1.8%)
- 97. Item Content Omitted. (True; NS 50.5%, CG 24.4%)
- 107. Item Content Omitted. (True; NS 47.3%, CG 35.5%)
- 118. Item Content Omitted. (True; NS 57.4%, CG 55.1%)
- 131. Item Content Omitted. (True; NS 43.3%, CG 32.9%)
- 143. Item Content Omitted. (True; NS 27.5%, CG 20.5%)
- 207. Item Content Omitted. (True; NS 66.9%, CG 43.2%)
- 219. Item Content Omitted. (True; NS 51.5%, CG 35.1%)
- 244. Item Content Omitted. (True; NS 56.9%, CG 84.9%)
- 305. Item Content Omitted. (True; NS 37.6%, CG 59.4%)
- 316. Item Content Omitted. (True; NS 45.1%, CG 30.7%)
- 337. Item Content Omitted. (True; NS 50.2%, CG 17.4%)

Self-Doubt (SFD, T Score = 56)

- 89. Item Content Omitted. (True; NS 35.9%, CG 6.1%)
- 232. Item Content Omitted. (True; NS 21.9%, CG 2.6%)

Family Problems (FML, T Score = 68)

- 19. Item Content Omitted. (False; NS 17.0%, CG 7.8%)
- 58. Item Content Omitted. (True; NS 57.3%, CG 37.4%)
- 103. Item Content Omitted. (True; NS 38.6%, CG 7.6%)

- 180. Item Content Omitted. (True; NS 30.4%, CG 17.4%)
- 215. Item Content Omitted. (True; NS 23.8%, CG 5.2%)
- 307. Item Content Omitted. (True; NS 19.1%, CG 3.6%)

Social Avoidance (SAV, T Score = 65)

- 47. Item Content Omitted. (False; NS 57.1%, CG 49.5%)
- 57. Item Content Omitted. (False; NS 22.1%, CG 15.1%)
- 109. Item Content Omitted. (False; NS 40.0%, CG 24.6%)
- 153. Item Content Omitted. (False; NS 25.9%, CG 29.6%)
- 201. Item Content Omitted. (False; NS 24.8%, CG 12.0%)
- 222. Item Content Omitted. (False; NS 19.6%, CG 6.2%)
- 278. Item Content Omitted. (True; NS 28.3%, CG 13.8%)

Critical Follow-up Items

This section contains a list of items to which the test taker responded in a manner warranting follow-up. The items were identified by police officer screening experts as having critical content. Clinicians are encouraged to follow up on these statements with the candidate by making related inquiries, rather than reciting the item(s) verbatim. Each item is followed by the candidate's response, the percentage of Police Officer Candidate comparison group members who gave this response, and the scale(s) on which the item appears.

- 26. Item Content Omitted. (True; 3.3%; RBS, RC9, AGG, AGGR-r)
- 84. Item Content Omitted. (True; 1.8%; BXD, RC9, AGG, AGGR-r)
- 257. Item Content Omitted. (True; 4.3%; VRIN-r, RC8, COG)
- 318. Item Content Omitted. (True; 2.8%; VRIN-r, RC7, ANP)
- 322. Item Content Omitted. (True; 4.0%; TRIN-r, EID, RC7)
- 330. Item Content Omitted. (True; 3.5%; THD, RC8, PSYC-r)
- 337. Item Content Omitted. (True; 17.4%; VRIN-r, RC9, AGG)

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 Test Response: L-r=37
- ² Test Response: K-r=38
- ³ Test Response: AGG=67
- ⁴ Correlate: AGG=67, Ref. 10
- ⁵ Test Response: FML=68
- ⁶ Correlate: FML=68, Ref. 2, 10
- ⁷ Correlate: FML=68, Ref. 10
- ⁸ Test Response: SAV=65
- ⁹ Correlate: SAV=65, Ref. 1, 10
- ¹⁰ Correlate: SAV=65, Ref. 4, 10
- ¹¹ Correlate: SAV=65, Ref. 10
- ¹² Inference: AGG=67
- ¹³ Correlate: SAV=65, Ref. 11
- ¹⁴ Test Response: SFD=56
- ¹⁵ Test Response: THD=57
- ¹⁶ Test Response: RC8=63
- ¹⁷ Test Response: BXD=57
- ¹⁸ Test Response: RC9=58
- ¹⁹ Inference: FML=68
- ²⁰ Inference: SAV=65
- ²¹ Correlate: FML=68, Ref. 9
- ²² Correlate: RC9=58, Ref. 3, 9
- ²³ Correlate: SAV=65, Ref. 5, 8
- ²⁴ Correlate: RC8=63, Ref. 5, 7; FML=68, Ref. 5, 8
- ²⁵ Correlate: RC8=63, Ref. 3, 9
- ²⁶ Correlate: AGG=67, Ref. 6
- ²⁷ Correlate: SAV=65, Ref. 9
- ²⁸ Correlate: SAV=65, Ref. 3, 9
- ²⁹ Correlate: SAV=65, Ref. 5
- ³⁰ Correlate: RC9=58, Ref. 9; AGG=67, Ref. 8
- ³¹ Correlate: AGG=67, Ref. 6, 7, 8
- ³² Correlate: RC8=63, Ref. 9
- ³³ Correlate: RC8=63, Ref. 7
- ³⁴ Correlate: SAV=65, Ref. 5, 6
- ³⁵ Correlate: RC9=58, Ref. 3, 5, 9
- ³⁶ Correlate: RC9=58, Ref. 9

RESEARCH REFERENCE LIST

1. 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)
2. Block, A. R., Ben-Porath, Y. S., & Marek, R. J. (2013). Psychological risk factors for poor outcome of spine surgery and spinal cord stimulator implant: A review of the literature and their assessment with the MMPI-2-RF. *The Clinical Neuropsychologist*, 27, 81-107. doi: [10.1080/13854046.2012.721007](https://doi.org/10.1080/13854046.2012.721007)
3. Detrick, P., Ben-Porath, Y.S., & Sellbom, M. (under review). Associations between MMPI-2-RF (Restructured Form) and Inwald Personality Inventory (IPI) scale scores in a law enforcement pre-employment screening sample.
4. Forbey, J. D., Lee, T. T. C., & Handel, R. W. (2010). Correlates of the MMPI-2-RF in a college setting. *Psychological Assessment*, 22, 737-744. doi: [10.1037/a0020645](https://doi.org/10.1037/a0020645)
5. Tarescavage, A. M., Brewster, J., Corey, D. M., & Ben-Porath, Y. S. (in press). Use of pre-hire MMPI-2-RF police candidate scores to predict supervisor ratings of post-hire performance. *Assessment*.
6. Tarescavage, A. M., Corey, D. M., & Ben-Porath, Y. S. (2014). Minnesota Multiphasic Personality Inventory-2-Restructured Form predictors of police officer problem behavior. *Assessment*. doi: [10.1177/1073191114534885](https://doi.org/10.1177/1073191114534885)
7. Tarescavage, A. M., Corey, D. M., & Ben-Porath, Y. S. (under review). A prorating method for estimating MMPI-2-RF scores from MMPI responses: Examination of score fidelity and illustration of empirical utility in the PERSEREC police integrity study sample.
8. Tarescavage, A. M., Corey, D. M., Gupton, H. M., & Ben-Porath Y.S. (under review). Criterion validity and practical utility of the Minnesota Multiphasic Personality Inventory-2-Restructured Form in assessments of police officer candidates.
9. Tarescavage, A. M., Fischler, G., Cappo, B., Hill, D., Corey, D. M., & Ben-Porath, Y. S. (under review). Minnesota Multiphasic Personality Inventory-2-Restructured Form (MMPI-2-RF) predictors of police officer problem behavior and collateral self-report test scores.
10. 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.
11. 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)

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.



Minnesota Multiphasic
Personality Inventory-2
Restructured Form®

SAMPLE REPORT

Case Description: Ms. D **Police Candidate Interpretive Report**

Ms. D is a 25-year-old, single female who applied to a small rural police department for an entry-level police officer position. Her background revealed significant conflicts among her family members, and between her and her parents, although she denied this in the interview. The background also noted that she was placed on academic probation for underage drinking at her small religious college. During the interview, she denied any abuse of alcohol or other drugs, and she presented as defensive and reticent.

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 Ms. D'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 © 2015 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. 9594 01/15



Minnesota Multiphasic
Personality Inventory-2
Restructured Form®

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

MMPI-2-RF®

Police Candidate Interpretive Report

David M. Corey, PhD, & Yossef S. Ben-Porath, PhD

ID Number:	Ms. D
Age:	25
Gender:	Female
Marital Status:	Never Married
Years of Education:	12
Date Assessed:	04/07/2013



Copyright © 2014 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.

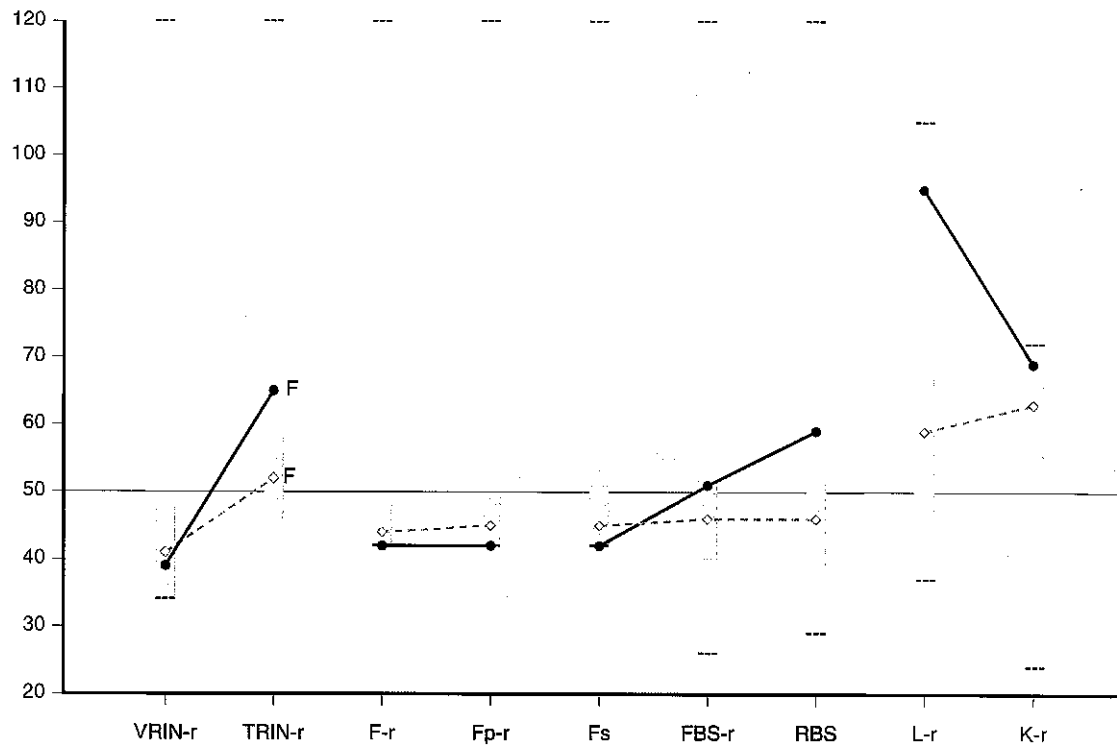
MMPI-2-RF, the MMPI-2-RF logo, and Minnesota Multiphasic Personality Inventory-2-Restructured Form are registered trademarks of the University of Minnesota. Pearson, the PSI logo, and PsychCorp are trademarks in the U.S. and/or other countries of Pearson Education, Inc., or its affiliate(s).

TRADE SECRET INFORMATION

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

[3.0 / 56 / 0.0.0]

MMPI-2-RF Validity Scales



Raw Score:	1	9	0	0	0	8	7	12	13
T Score:	39	65 F	42	42	42	51	59	95	69
Response %:	100	100	100	100	100	100	100	100	100
Cannot Say (Raw):	0								
								Percent True (of items answered):	24 %

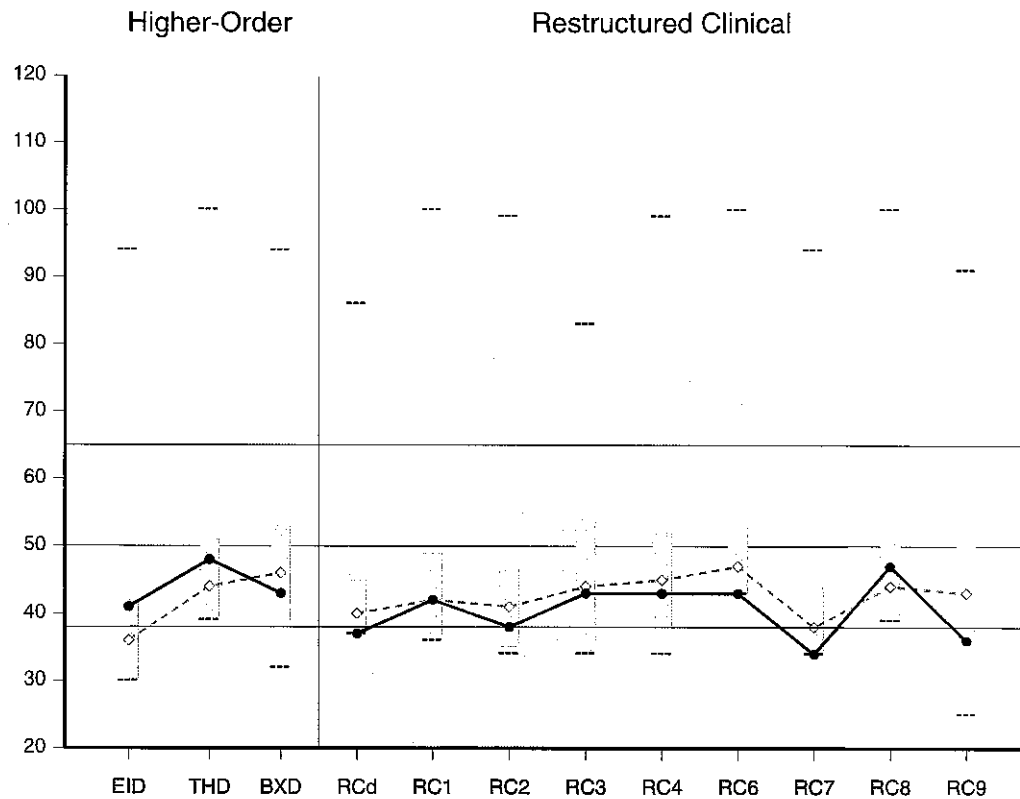
Comparison Group Data: Police Officer Candidate (Men and Women), N = 2,074

Mean Score (◇--◇):	41	52 F	44	45	45	46	46	59	63
Standard Dev (±1 SD):	7	6	4	5	6	6	7	13	8
Percent scoring at or below test taker:	59	99	75	78	70	89	97	99.9	82

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:	4	1	3	0	1	1	3	2	0	0	1	4
T Score:	41	48	43	37	42	38	43	43	43	34	47	36
Response %:	100	100	100	100	100	100	100	100	100	100	100	100

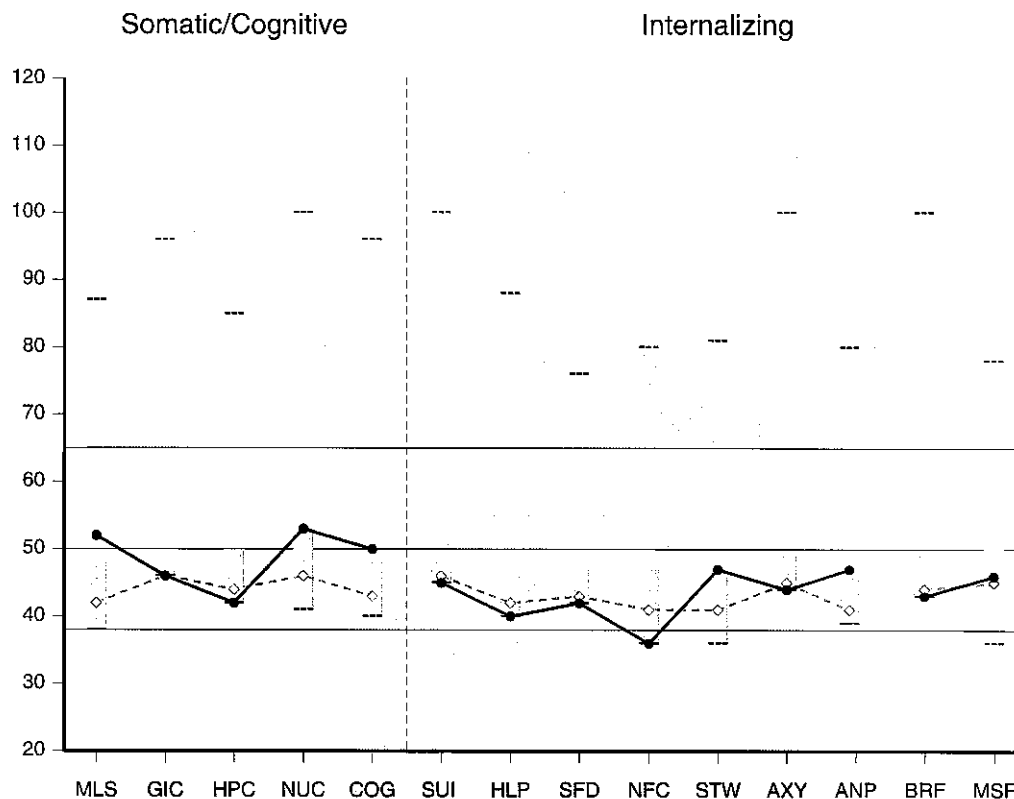
Comparison Group Data: Police Officer Candidate (Men and Women), N = 2,074

Mean Score (---◇):	36	44	46	40	42	41	44	45	47	38	44	43
Standard Dev (±1 SD):	6	7	7	5	7	6	10	7	7	6	7	8
Percent scoring at or below test taker:	83	82	43	62	68	56	57	51	75	45	79	21

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	1	0	0	0	0	2	0	1	0	2
T Score:	52	46	42	53	50	45	40	42	36	47	44	47	43	46
Response %:	100	100	100	100	100	100	100	100	100	100	100	100	100	100

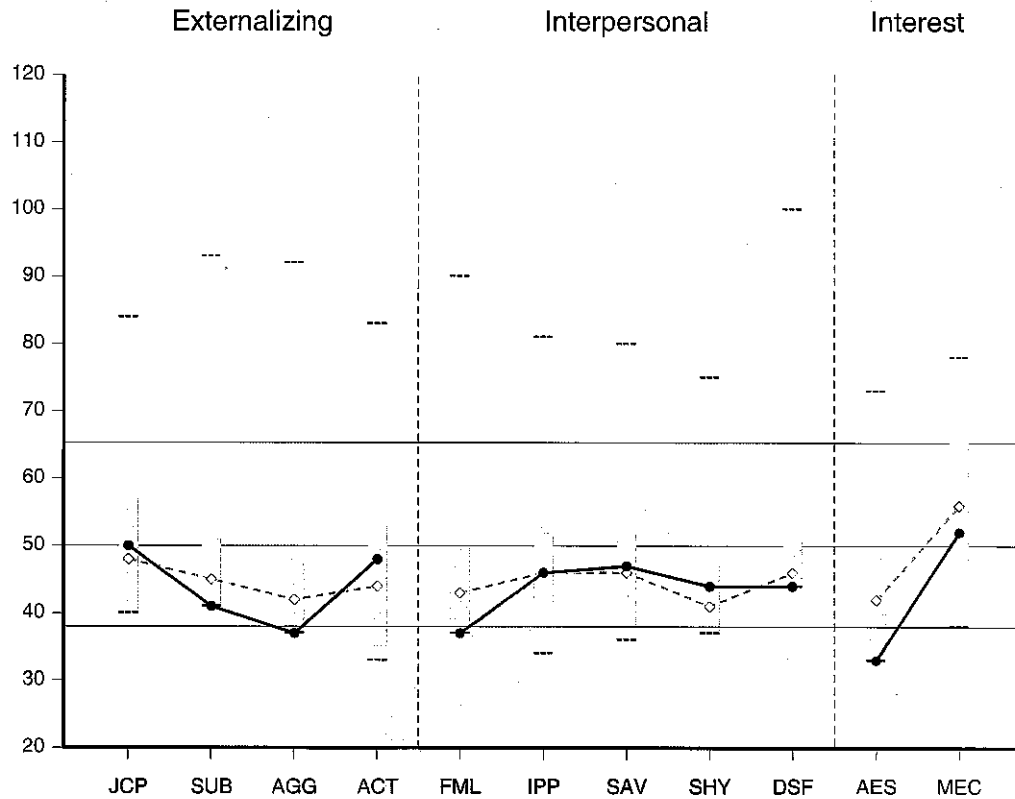
Comparison Group Data: Police Officer Candidate (Men and Women), N = 2,074

Mean Score (◇---◇):	42	46	44	46	43	46	42	43	41	41	45	41	44	45
Standard Dev (±1 SD):	6	4	6	7	5	2	4	4	6	6	4	5	5	8
Percent scoring at or below test taker:	96	95	83	90	93	99.3	88	90	50	91	94	94	88	64

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	0	0	3	0	3	2	1	0	0	3
T Score:	50	41	37	48	37	46	47	44	44	33	52
Response %:	100	100	100	100	100	100	100	100	100	100	100

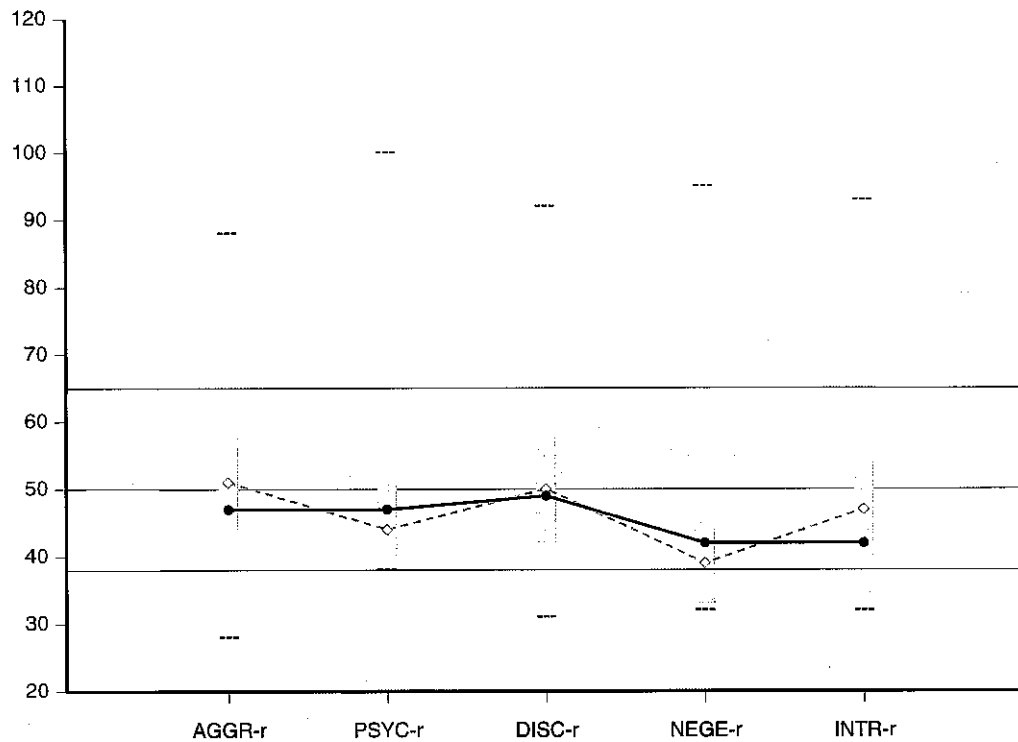
Comparison Group Data: Police Officer Candidate (Men and Women), N = 2,074

Mean Score (◊---◊):	48	45	42	44	43	46	46	41	46	42	56
Standard Dev (±1 SD):	9	6	6	9	7	6	8	6	5	8	11
Percent scoring at or below test taker:	76	65	56	79	47	65	65	80	90	26	46

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:	8	1	6	3	3
T Score:	47	47	49	42	42
Response %:	100	100	100	100	100

Comparison Group Data: Police Officer Candidate (Men and Women), N = 2,074

Mean Score ($\diamond$ - - - $\diamond$):	51	44	50	39	47
Standard Dev (± 1 SD):	7	7	8	6	7
Percent scoring at or below test taker:	39	80	55	79	34

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	0	39	65 F
	CNS	VRIN-r	TRIN-r
Over-Reporting	42	42	42 51 59
	F-r	Fp-r	Fs FBS-r RBS
Under-Reporting	95	69	
	L-r	K-r	

SUBSTANTIVE SCALES

Somatic/Cognitive Dysfunction	42	52	46	42	53	50
	RC1	MLS	GIC	HPC	NUC	COG
Emotional Dysfunction	41	37	45	40	42	36
	EID	RCd	SUI	HLP	SFD	NFC
		38	42			
		RC2	INTR-r			
		34	47	44	47	43
		RC7	STW	AXY	ANP	BRF
						46
						MSF
						42
						NEGE-r
Thought Dysfunction	48	43				
	THD	RC6				
		47				
		RC8				
		47				
		PSYC-r				
Behavioral Dysfunction	43	43	50	41		
	BXD	RC4	JCP	SUB		
		36	37	48	47	49
		RC9	AGG	ACT	AGGR-r	DISC-r
Interpersonal Functioning	37	43	46	47	44	44
	FML	RC3	IPP	SAV	SHY	DSF
Interests	33	52				
	AES	MEC				

Scale scores shown in bold font are interpreted in the report.

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 in the context of preemployment psychological evaluations of police and other law enforcement officer candidates. It focuses on identifying problems; it does not convey potential strengths. The information it contains should be considered in the context of the test taker's background, the demands of the position under consideration, the clinical interview, findings from supplemental tests, and other relevant information.

The interpretive statements in the Protocol Validity section of the report are based on T scores derived from the general MMPI-2-RF normative sample, as well as scores obtained by the multisite sample of 2,074 individuals that make up the Police Officer Candidate comparison group.

The interpretive statements in the Clinical Findings and Diagnostic Considerations sections of the report are based on T scores derived from the general MMPI-2-RF normative sample. Following recommended practice, only T scores of 65 and higher are considered clinically significant. Scores at this clinical level are generally rare among police officer candidates.

Statements in the Comparison Group Findings and Job-Relevant Correlates sections are based on comparisons with scores obtained by the Police Officer Candidate comparison group. Statements in these sections may be based on T scores that, although less than 65, are nevertheless uncommon in reference to the comparison group.

Sources for interpretive statements in all sections are listed in the Endnotes section of this report. See User's Guide for the MMPI-2-RF Police Candidate Interpretive Report for detailed information on report features.

SYNOPSIS

Scores on the MMPI-2-RF validity scales raise concerns about the possible impact of under-reporting on the validity of this protocol.

PROTOCOL VALIDITY

Content Non-Responsiveness

There are no problems with unscorable items in this protocol. The test taker responded relevantly to the items on the basis of their content.

Over-Reporting

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

Under-Reporting

The test taker presented herself in an extremely positive light by denying an extraordinarily large number of minor faults and shortcomings that most people acknowledge¹. This level of virtuous self-presentation is very uncommon even among individuals with a background stressing traditional values². It is also very uncommon among police officer candidates. Only 1.1% of the comparison group members claimed this many or more uncommon virtues. Any absence of elevation on the substantive scales is uninterpretable. Elevated scores on the substantive scales may underestimate the problems assessed by those scales³. The candidate's responses may be a result of unintentional (e.g., very naïve) or intentional under-reporting. One way to distinguish between the two is to compare her responses to items with historical content against available collateral information (e.g., background information, interview data). Following are the test taker's responses to items with potentially verifiable historical content:

- 19. Item Content Omitted (True)
- 38. Item Content Omitted (True)
- 49. Item Content Omitted (False)
- 66. Item Content Omitted (False)
- 141. Item Content Omitted (False)
- 173. Item Content Omitted (False)
- 205. Item Content Omitted (False)
- 223. Item Content Omitted (False)
- 312. Item Content Omitted (False)



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.

Corroborated evidence of intentional under-reporting may be incompatible with the integrity requirements of the position. In addition, this level of virtuous self-presentation may reflect uncooperativeness that precludes a reliable determination of the candidate's suitability. Corroborating evidence in support of this possibility may be found in other test data, the clinical interview, or background information.

The candidate's virtuous self-presentation may reflect an overly rigid orientation to matters of morality and/or an inability to self-examine that may impair her effectiveness as a law enforcement officer. This can be explored through interview and collateral sources.

In addition, she presented herself as very well-adjusted⁴. This reported level of psychological adjustment is relatively rare in the general population but more common among police officer candidates.

CLINICAL FINDINGS

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

There are no indications of clinically significant somatic, cognitive, emotional, thought, or behavioral dysfunction in this protocol. However, because of indications of under-reporting described earlier, such problems cannot be ruled out.

DIAGNOSTIC CONSIDERATIONS

No specific psychodiagnostic recommendations are indicated by this MMPI-2-RF protocol.

COMPARISON GROUP FINDINGS AND JOB-RELEVANT CORRELATES

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

The test taker's scores on the substantive scales are all within normal limits for the general population and for police officer candidates. However, as indicated earlier, in light of evidence of considerable under-reporting³, these results do not rule out the possibility that psychological problems will impede the candidate's ability to perform the duties of a police officer.

ITEM-LEVEL INFORMATION

Unscorable Responses

The test taker produced scorable responses to all the MMPI-2-RF items.

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 her T score on that scale is 65 or higher.

The test taker has not produced an elevated T score (≥ 65) on any of these scales.

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 her 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 Police Officer Candidate (Men and Women) comparison group (CG) that answered each item in the keyed direction are provided in parentheses following the item content.

Uncommon Virtues (L-r, T Score = 95)

- 16. Item Content Omitted (False; NS 16.7%, CG 37.3%)
- 45. Item Content Omitted (False; NS 40.0%, CG 52.0%)
- 95. Item Content Omitted (False; NS 5.8%, CG 29.6%)
- 127. Item Content Omitted (False; NS 2.7%, CG 11.1%)



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.

- 154. Item Content Omitted (False; NS 33.1%, CG 51.2%)
- 182. Item Content Omitted (True; NS 33.6%, CG 72.4%)
- 183. Item Content Omitted (False; NS 9.6%, CG 5.8%)
- 211. Item Content Omitted (False; NS 10.9%, CG 29.2%)
- 241. Item Content Omitted (False; NS 14.0%, CG 28.2%)
- 268. Item Content Omitted (True; NS 18.1%, CG 20.0%)
- 298. Item Content Omitted (False; NS 27.2%, CG 65.8%)
- 325. Item Content Omitted (False; NS 5.4%, CG 7.6%)



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 Follow-up Items

This section contains a list of items to which the test taker responded in a manner warranting follow-up. The items were identified by police officer screening experts as having critical content. Clinicians are encouraged to follow up on these statements with the candidate by making related inquiries, rather than reciting the item(s) verbatim. Each item is followed by the candidate's response, the percentage of Police Officer Candidate comparison group members who gave this response, and the scale(s) on which the item appears.

The test taker did not respond to any critical follow-up items in the keyed direction.

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 Test Response: L-r=95

² Correlate: L-r=95, Ref. 1

³ Correlate: L-r=95, Ref. 2, 3, 4

⁴ Test Response: K-r=69

RESEARCH REFERENCE LIST

1. Bridges, S. A., & Baum, L. J. (2013). An examination of the MMPI-2-RF L-r scale in an outpatient protestant sample. *Journal of Psychology and Christianity*, 32, 115-123.
2. Forbey, J. D., Lee, T. T. C., Ben-Porath, Y. S., Arbisi, P. A., & Gartland, D. (2013). Associations between MMPI-2-RF validity scale scores and extra-test measures of personality and psychopathology. *Assessment*. doi: [10.1177/1073191113478154](https://doi.org/10.1177/1073191113478154)
3. Sellbom, M., & Bagby, R. M. (2008). The validity of the MMPI-2-RF (Restructured Form) L-r and K-r scales in detecting under-reporting in clinical and non-clinical samples. *Psychological Assessment*, 20, 370-376. doi: [10.1037/a0012952](https://doi.org/10.1037/a0012952)
4. 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.

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.

For this discussion, you will use the PSY640 Checklist for Evaluating Tests (Links to an external site.) document to compare two assessment instruments used in industrial and organizational (I-O) psychology assessment. Based on the information in your text and assigned readings, select and evaluate two psychological tests used in industrial-organizational psychological assessment. You may not evaluate any of the tests you evaluated in the Week Four Applications in Personality Testing discussion. In addition to the text, locate a minimum of two appropriate scholarly and/or peer-reviewed sources to aid you in the analysis of the psychometric properties of the instruments based on published data. In your initial post, provide the names of the two tests you evaluated, and attach your completed PSY640 Checklist for Evaluating Tests document (Links to an external site.). You must maintain the original format of the document and include the textbook and two additional scholarly and/or peer-reviewed sources in the references section.

PLEASE USE TWO PEER
REVIEWED SOURCES FOR THIS
DISCUSSION THE CHECKLIST
IS ATTACHED PLEASE DO
NOT USE PERSONALITY ASSESSMENT.

PSY640 CHECKLIST FOR EVALUATING TESTS

Test Name and Versions	
Assessment One	Assessment Two

Purpose(s) for Administering the Tests	
Assessment One	Assessment Two
Characteristic(s) to be Measured by the Tests (skill, ability, personality trait)	
Assessment One	Assessment Two
Target Population (education, experience level, other background)	
Assessment One	Assessment Two

Test Characteristics		
	Assessment One	Assessment Two
1. Type (paper-and-pencil or computer): Alternate forms available:		
2. Scoring method (computer or manually):		
3. Technical considerations: a) Reliability: $r =$ b) Validity: $r =$ c) Reference/norm group: d) Test fairness evidence: e) Adverse impact evidence: f) Applicability (indicate any special groups):		
4. Administration considerations:		
5. Administration time:		
6. Materials needed (include start-up, operational, and scoring costs):		
7. Facilities needed:		
8. Staffing requirements:		
9. Training requirements:		
10. Other considerations (consider clarity, comprehensiveness, and utility):		
11. Test manual information:		
12. Supporting documents		

available from the publisher:		
13. Publisher assistance:		
14. Independent reviews:		

Overall Evaluation (One to two sentences providing your conclusions about the test you evaluated)	
Assessment One	Assessment Two
Name of Test:	Name of Test:

References

List references in APA format as outlined by the Ashford Writing Center.