

Conducting an Employment Selection Validation Study at a Food Processing Plant

Mark A. Johnson
James P. Jolly
Idaho State University

ABSTRACT

Two consultants were hired to design and implement an employee selection system for a sugar processing plant. They conducted job analysis, collected performance appraisal data (criteria), worked to establish the validity of the test battery, and performed other activities related to completing a professional test validation study. The case ends with the plant manager enthusiastically announcing to the consultants and departmental managers that testing would begin at the packing plant in ten days, that other company facilities would soon follow suit, and that the corporate office was very impressed with the near term implementation of testing at all of its processing plants. The consultants informed the plant manager that his plan was premature and a lot more analysis was required before testing could begin at the current location and that extending the testing to the other plants was beyond the scope of the current validation study. The manager told the consultants “we need to make this happen” and later asked “can you make this happen?”

Background

In a November 2006 meeting of the employee selection project committee, Brandon Smith, Plant Manager at Northwestern Sugar, announced to the others in the room, “Great, we will begin testing applicants in 10 days. Also, we have been talking with corporate and they are very interested in initiating the testing battery at our other four plants soon after we get started here. We have really impressed them by getting this program online.” Brian Higgins, a university professor, and Dennis Hill, a doctoral student, consultants to the committee, were in shock. They weren’t sure how to respond. “Where the hell did that come from and how did they jump to this conclusion at this point in the project?” Higgins thought to himself.

Higgins and Hill had been working on the development of an employee selection system for Northwestern Sugar for ten months. The project had come about almost by accident, spurred by a comment Smith had made earlier that year in January during a management-training program Higgins and Hill were conducting at a major Western university.

In that training program, Higgins and Hill had been discussing personnel selection and the importance of establishing the job relatedness of tests when one of the trainees, Brandon Smith, exclaimed, “Testing is illegal! There was this Griggs case that disallows testing because of discrimination against minorities.”

Higgins and Hill glanced at one another in a way that communicated to the other, “Oh no, here we go again,” but fought back their smiles, not wanting to embarrass Smith. The two had heard this sort of comment before. Unfortunately, the manager’s assertions were consistent with the perceptions and understandings of many practicing managers, but they were neither accurate nor reflective of current employment law.

Higgins and Hill then discussed the current legal status of employment testing with Smith and the other trainees. The two trainers also met with Smith after that day’s management program to discuss testing and validation processes in more detail. They subsequently were asked to help Smith and his company, Northwestern Sugar, with their employment testing program.

Northwestern Sugar

Northwestern Sugar was a privately held corporation that processed sugar beets and packaged the sugar obtained from the beets. The firm employed more than 2,000 employees at its five sugar processing/packaging plants. The plants were located in three Northwestern states. Operating employees were represented by the Teamsters. Employee relations were generally good.

The February Meeting: To Test, or Not to Test?

In February of 2006, just one month after the training program, Higgins and Hill were sitting with Smith in a conference room at Northwestern’s largest plant. Also in attendance were the Packaging Department Manager, Charlie Jackson, and Assistant Manager, Dan Berns. The purpose of the meeting was to plan a strategy for implementing a selection-testing program for hiring the plant’s operating employees. After further discussions with Higgins and Hill, Smith had done an about-face from his initial assertion that testing was illegal. He now realized that employment testing could be conducted without violating employment laws.

Smith went on to discuss a major reason for Northwestern’s interest in testing. Due to increased competition and changes in the marketplace, the company’s management realized a need for significant changes to some of their human resource philosophies and practices. Gone were the days of merely hiring “hands”, a term familiar to most people in the area, which was dominated by ranches and farms. Now the company needed a more comprehensive approach to hiring and retaining employees. Northwestern needed employees with wider ranging and more substantial skills and they needed systems in place to attract, select, and retain more skilled workers. As an example of the company’s changing needs, Smith described the company’s plan to implement an Enterprise Resource Planning (ERP) system. The effects of this information system would pervade every aspect of the company’s operations, going all the way down to the plant floor. “Traditionally, the work requirements at our plant demanded more physical effort than brains,” Smith commented, “but now, well, a lot has changed and a lot more is going to change with the implementation of the ERP system.

Our plant workers will have to use the computer systems to record operations, to receive information, and to monitor the processes. This is unlike anything they have had to do in the past.” Smith continued:

We need workers who have the aptitudes and skills to effectively run these systems. Moreover, we have had a problem finding capable workers to promote into our higher skilled jobs and, eventually, into supervisory and managerial positions. Our current selection processes do not provide us with enough capable plant floor workers and this in turn limits our selection pool for promotions. For years we were bound to adhere to our promotion-from-within system. We had a very formalized bidding system, and seniority was the most important criterion. Over the last few years we have increased the importance of hiring the most qualified internal candidate, but seniority is still given a lot of weight. I sense that we will continue the trend toward increasing the skill requirements for bidding and promotions and rely less and less on seniority.

Smith paused for a moment as if to catch his breath and to give everyone a chance to digest what he had said. He continued:

The bottom line is that we need not only to hire more qualified people into our entry-level positions but also we need to simultaneously improve the level of all our workers in the higher job classifications. What we need is a testing program that will improve the average aptitudes and performance of workers, which will improve the quality of our internal ranks for promotion and enable us to objectively screen for higher classification and supervisory positions.

Current Selection Process and Job Paths

As the meeting between Higgins, Hill, and the three Northwestern Sugar managers continued, Higgins asked the managers to describe the current hiring processes. Dan Berns, the Assistant Packaging Manager, detailed the process.

First, we list our entry-level openings with the local state employment office. There, interested candidates complete our job application forms and the state office refers those meeting our specifications to us. The specs are pretty low: All we require is that they be able to read and write well enough to complete our application forms. Heck, some of the time we aren’t even sure they can do that. At one point we found out that applicants were being allowed to take their applications home, and it turns out that some who were illiterate were having family members or friends help them fill out the forms. At any rate, the employment office refers those applicants who meet the minimum qualifications to us. Our Employment Specialist interviews candidates at the state employment office and makes decisions based on the application form information and screening interviews.

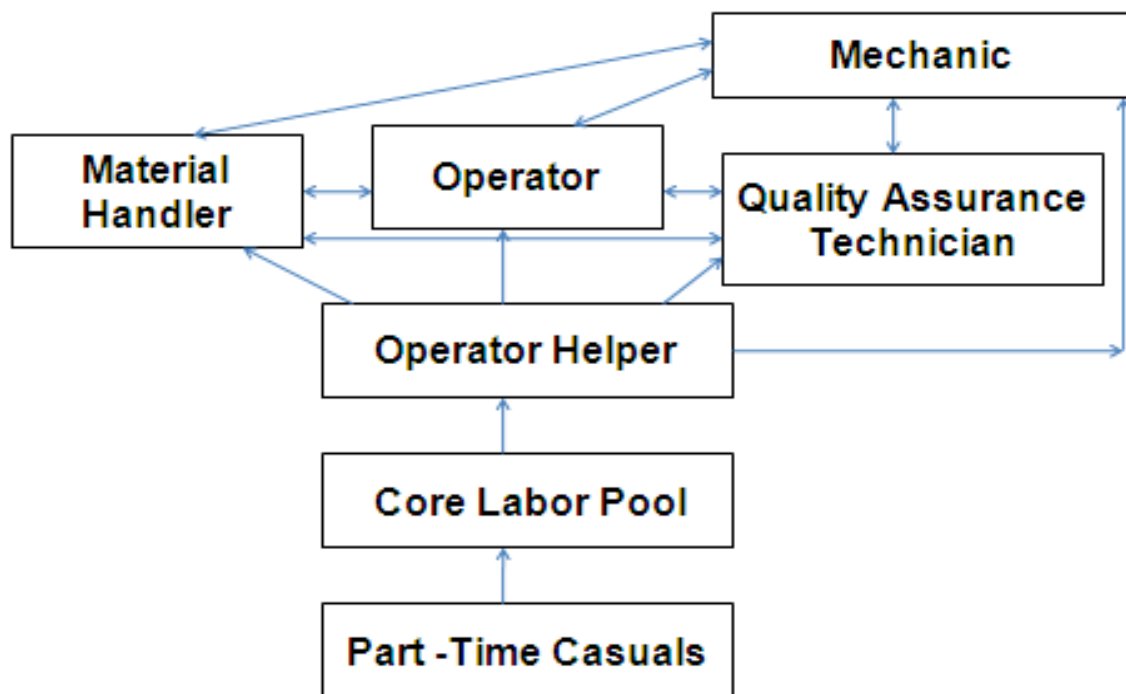
Berns continued, "In the past, we thought about using high school diplomas as a job requirement but the Human Resource Department people were quick to tell us that the EEOC scrutinized this practice quite closely. Evidently, plenty of people who graduate don't have the skills we need and there are many people who don't have degrees that do have those skills. HR said that when they looked closely at our own workers, this held true." Higgins chimed in, "That sounds right. Using high school graduation as a job requirement for operating employees such as yours has often resulted in adverse impact, an unfavorable disproportional effect for a protected group such as African Americans or Hispanics. In fact, the Griggs case that Smith, Hill, and I talked about during our management training program dealt with this very issue. Selection techniques that result in adverse impact against a protected group must be shown to be job related. This applies to specific screening methods such as requirements for a high school diploma."

Higgins then asked the managers to describe the job paths for new hires. Collectively they described how new hires begin in the classification "Casual Employees" who work part-time, and often on an as-needed basis. These workers typically perform some sanitation/laborer functions, cleaning equipment and the plant. Others may work on the processing and packaging lines, performing routine types of work such as loading cardboard cartons onto the equipment, monitoring the equipment that filled boxes or bulk bins with product, or removing jammed raw product or containers from equipment.

From the Casual pool, workers become part of the Core Labor pool, working more regular hours – some even working full-time. Core Labor workers perform many of the same types of tasks that they did when they were classified as Casuals. In addition, many of the Core Labor workers fill in for the Operator Helper positions, working with Operators and other Operator Helpers on the various processing and packaging lines.

From the Core Labor pool, workers first bid into the Operator Helper job. Operator Helpers assist line Operators on the processing and packaging lines. Just about every higher job grade worker (Operator, Material Handler, Quality Assurance Technician, and Mechanic) in the processing and packaging units has worked as an Operator Helper and many of them eventually become Operators. Operator Helpers can bid into Operator, Material Handler, Quality Assurance Technician, and Mechanic jobs. Material Handlers operate forklifts and other lift equipment and are permitted to bid into Operator, Quality Assurance Technician, and Mechanic jobs. Similarly, Quality Assurance Technicians may bid into Operator, Material Handlers, and Mechanic jobs. Mechanic jobs are the highest paid jobs and workers who get into these jobs do not tend to leave them unless they advance into supervisory jobs. Also, most workers who bid into the Operator, Material Handler, and Quality Assurance Technician jobs tend to stay in those except for those who are selected as Mechanics or advance into supervisory jobs. (The case authors provide Exhibit 1 to illustrate the typical job path for processing and packaging workers).

Exhibit 1 – Common Job Paths



As previously noted, supervisors are selected from the rank and file. The most senior bidder who meets the job qualifications is promoted. Historically, questions about what constitutes “qualified” have been raised. There had been little consistency in applying this criterion because no objective standards had been developed. As a result, supervisors simply made a judgment as to whether the most senior bidder was qualified to advance; this, of course, opened the door to allegations of favoritism and bias. If Higgins and Hill could help Northwestern develop comprehensive, accurate job descriptions and objective tests of key skills and abilities related to those job descriptions then workers and the union would see promotion practices as less arbitrary and capricious, in addition to the obvious benefit of having more qualified workers at entry and advanced levels.

Limited Internal Pool

Berns went on to describe retention and turnover issues. “We lose perhaps 50 percent of our Casuals before they have a chance to advance into our Core Labor pool. Some workers just can’t wait two to six months for the opportunity to become a Core Labor worker. Many Casuals, and usually the best ones, have other opportunities elsewhere.” Berns continued, “Effectively, we lose the cream of our Casual employees. Consequently, when the remaining workers bid into the Core Labor pool, the average quality of these workers is lower than our entry-level hire quality. This in turn eventually affects the average quality of our hiring pool into the Operator Helper, Quality Assurance Technician, Material Handler, and Mechanic jobs. We of course also

experience turnover in those areas as well,” Berns stated using a tone that demonstrated his frustration with the system. “And this diminished pool has been our sole input into the supervisory ranks. Well, you tell me--how sound a process is this?” He went on to add that even many of the workers realized the resulting problem with the caliber of some of the workforce. Many of the best workers were frustrated with processes that “stick us with co-workers whose only qualification is that they can fog a mirror”.

The Wage System

Higgins asked the managers for a copy of the collective bargaining agreement. He took a minute or so to browse the document and made some comments as he looked through it. He noted that the agreement contained information about a number of provisions including seniority, filling job vacancies, lines of progression to be followed when filling bids, and wage brackets along with their corresponding wage rates. Higgins then asked the managers to describe the competitiveness of the wage system. After hearing that the negotiated rates met or exceeded the rates of similar jobs at other organizations in the relevant labor market, Higgins asked the managers whether the current rate structure, although locally competitive at the time, would be high enough to attract the higher-level candidates that they hoped to hire and retain in the future with a testing battery. “I guess not,” and “probably not,” were the responses that Higgins heard. “Effectively you want to change your recruiting pool by raising it a notch or two and cutting out the low end of the current labor pool,” Higgins noted. “Yes, yes, that’s about it,” Smith replied.

At that point, Higgins emphasized that Northwestern would have to rethink its wages in view of its strategy to improve the skill levels of its operating and supervisory workforce. “These things often have to come with a price tag,” Higgins commented. “You need to realize that the selection system, though very important, is only one variable associated with your overall Human Resources system. You just can’t look at selection in isolation.” Smith agreed, but indicated that compensation issues would have to wait for now, saying “We have a pretty full plate trying to deal with these selection issues.”

Types of Validity

Higgins and Hill then took the opportunity to describe the concept of validation, different validation strategies, and the legal implications of employment testing to the three managers. They informed the three that valid selection processes use only job related measures. The managers were told that there are four major types of validation — criterion related validity (which is statistical and includes predictive and concurrent), content validity, construct validity, and transportability of validity. (Appendix A provides more information regarding each of these approaches, as described by Higgins and Hill to the managers).

A Validation Study Proposal

By the end of the February 2006 meeting, Smith and his two managers had expressed their interest in the prospect of using tests for entry level selection and for bidding into higher classification jobs and promotions into the supervisory ranks. Higgins and Hill cautioned them that there were no guarantees that a validation study would support the validity (job-relatedness) of a test or test battery. The purpose of a validation study is just that – to “study” prospective tests to find whether the tests are valid in the employer situation for the job or job classes studied. Only then may tests that result in adverse impact be used as evidence of job relatedness. Despite these caveats, Smith requested that Higgins and Hill develop a validation study proposal.

By mid-March, the two consultants had completed a project proposal that included details pertaining to the scope of the project, deliverables, caveats associated with validation study outcomes, a project schedule indicating the start dates, duration, and completion dates of major project activities, resources needed from Northwestern (management support, employee time required to complete documents and to discuss work content with the consultants), and the project budget. The project was subsequently accepted in the first week of April by Northwestern Sugar.

Job Analysis Process

In late April, within a couple of weeks of the project proposal being accepted, Higgins and Hill began conducting job analyses as part of the validation study. They had decided to focus the study in the packaging department of the plant. This location was selected because they already had the support of the two top managers of that unit and the packaging department included a relatively large concentration of Operator Helper and Operator jobs. As previously noted, almost all higher grade workers have worked as Operator Helpers and many also have worked as Operators. The packaging department employed approximately 50 Operators and Helpers out of its approximately 75 workers. The 50 employees exceeded the minimum sample size that Higgins and Hill felt was statistically necessary for a meaningful concurrent validity study. Workers from the other job classifications (Laborers, Material Handlers, Mechanics, Quality Assurance Technicians, and Supervisors) also contributed useful information to the job analyses through interviews with Higgins and Hill and by completing job analysis questionnaires. However, workers in these other jobs were not included in the criterion related validity portion of the project (and associated test-taking). Job analyses were conducted for all of the jobs and new job descriptions were to be written by Higgins and Hill for all packaging jobs, including Supervisors, and not just the Operator and Operator Helper positions.

Higgins and Hill utilized a variety of job analysis methods as part of the validation effort in the packaging department. They felt that the different methods, some conducted concurrently while others were performed in stages, would provide the best understanding and documentation for the jobs studied. The two researchers ended up spending over one hundred hours observing and interviewing different workers

performing their packaging department responsibilities. In addition, Higgins and Hill developed a customized job analysis questionnaire (JAQ) that was based on the organizationally specific information they obtained from observation of and interviews with workers and supervisors. The JAQ included closed- and open-ended questions, checklists, and rating scales. This instrument was administered to all operating employees in the department, including Laborers, Operator Helpers, Operators, Material Handlers, Quality Assurance Technicians, and Mechanics.

After administering and analyzing the JAQ documents, the investigators developed a task inventory comprised of more than one hundred statements of the job tasks performed by the Operators and Helpers. Higgins and Hill administered the task inventory to all of Operators and Operator Helpers during the first two weeks of July. Using five point rating scales, the workers rated each task statement on both the importance and frequency of the tasks performed in their jobs. Higgins and Hill also reviewed written materials such as standard operating procedures, line operations manuals, and safety procedures used by the workers. In addition, training materials related to the soon-to-be-implemented ERP system were reviewed and the two consultants attended company-conducted employee training sessions involving the system. They then determined the reading grade levels for the materials (e.g., 6th grade reading level).

Review of Job Analysis Findings and Consideration of Possible Selection Devices

The researchers completed the job analysis procedures and written job descriptions for all of the packaging department jobs in September. Qualitative data from the interviews, observation, and the job analysis questionnaire provided a great deal of useful information detailing the job duties and tasks. In addition, the information obtained from the task inventory ratings provided quantitative data regarding the similarities and differences between the Operator and Operator Helper jobs, and the relative importance and frequency of tasks performed for each of these two jobs. Most of the tasks included on the task inventory were rated as having been performed frequently and/or were considered important to the job by the worker raters.

Collectively, the job analysis information provided the researchers with job-related information to assist in their decision about which selection devices to include in the testing battery. The testing devices were to assess the abilities and skills that Higgins and Hill had determined to be important competencies for workers to possess in order to effectively perform the tasks, duties and responsibilities of the Operator Helper and Operator jobs. The abilities and skills assessed by these tests were also deemed important to other packaging and processing department jobs. Specifically, Higgins and Hill decided to conduct a concurrent validation study in which they would test current workers using a custom-developed math test and three tests available from test development/publishing companies: a reading test, a mechanical ability test, and a manual dexterity test.

The math test measured an applicant's ability to count, add, subtract, multiply, divide and to compute percentages, fractions and decimals. The questions reflected the types of mathematical calculations the workers performed on the job. The reading test measured an individual's ability to read and understand written directions and materials such as those that were common on the processing and packaging lines. The mechanical aptitude test measured a person's ability to perceive and understand mechanical processes in everyday life and in manufacturing, processing, and mechanical oriented work environments. The types of problems contained in this test reflected the types of mechanical processes workers in the processing and packaging department dealt with on an on-going basis. Finally, the manual dexterity test measured an individual's ability to make skillfully coordinated movements with one's hands, including the grasping and manipulation of objects. These types of activities were regularly performed by packaging and processing workers. Each of these tests measured behaviors, skills, and capabilities regularly required of plant workers, especially Operators and Operator Helpers.

Testing of current packaging department employees began in September of 2006 and was completed in the first week of October. The results of the four tests would be analyzed statistically along with performance measurement data (which would serve as the criterion in a concurrent validation study).

Performance Measurement (The Criterion)

Early discussions during the pre-proposal stage between the two consultants, Smith and the two top managers of the packing department clearly indicated that there were no meaningful performance appraisal data that could be used as the criterion for a validity study. The managers had expressed concern that some the ratings were not taken as seriously as perhaps they should have been and probably contained some biases as well. It became readily apparent that job related performance data would need to be collected subsequent to the completion of the job analysis phase. Accordingly, Higgins and Hill developed a job-related performance assessment instrument in September and in mid-October they had the packaging department supervisors rate their workers using this custom-developed instrument. This clinical approach allowed supervisors to rate their workers more honestly and systematically without having to worry about worker or union ramifications. Supervisors would not be affected by the behavioral issues commonly associated with performance appraisal such as fear of having to justify the ratings to the worker or the union or claims of favoritism. In an attempt to minimize rating errors and biases, Higgins conducted performance rater training before the supervisors evaluated their workers using the appraisal instrument.

Completing the Validation Study

In early November, Higgins and Hill had collected the test and performance data necessary to complete the statistical portion of the project. The data had been entered into a statistical software program and some preliminary statistical analysis had been

performed. However, a variety of analyses were needed before the output could be meaningfully interpreted and a lot of work was necessary to document validity if in fact the results supported a conclusion that some or all of the tests were valid. If the results demonstrated sufficient validity for any of the tests, they would have to document the results in a technical validation report as required by the EEOC Guidelines. They would also have to examine how the tests worked together. Even if two tests are found to be valid, one may not contribute meaningfully to the overall validity when used in conjunction with another test. This would signify a lack of what is called “incremental validity” (the additional total validity when a test is added to a battery). In addition, the consultants would need to determine cut-off scores for the different tests and also determine the extent of adverse impact for each test. Adverse impact can change when different cut-offs are used, so alternative cut-offs and the resulting adverse impacts would need to be evaluated. Additionally, Higgins and Hill would have to develop a Test User’s Manual that addresses issues such as test administration, scoring, security, and retesting policies. Despite all of the work that remained, Higgins and Hill were excited that they had completed the data collection phase and some statistical analyses. They knew they were well on their way to finalizing the project.

The November Meeting: Let’s Start Testing!

Smith had called a meeting for later that week to discuss the “progress” of the validation project. Also in attendance were the two packaging department managers, the Human Resource Manager, Higgins, and Hill. The two researchers proudly communicated that all of the testing and performance ratings data had been collected and some preliminary statistical analysis had been completed. “At this point, all we need to do is finish the statistical analyses, interpret the results, and if they turn out the way we expect them to, we will need to do additional analyses. In addition, we will need to develop and then write the testing and retesting procedures, and write and complete a validation report which provides important documentation for our study.” It was at this point that Higgins and Hill were thrown the curve when Smith announced his eagerness to begin testing and to expand the testing program to other plants: “Great, we will begin testing applicants in 10 days. Also, we have been talking with corporate and they are very interested in initiating the testing battery at our other four plants soon after we get started here. We have really impressed them by getting this program online.” Higgins was in shock. “Where in the hell did that come from and how did they jump to this conclusion at this point in the project?” he thought. The validation study was performed for jobs at one specific plant. At no time had there been any mention of testing in the immediate future or at any of the other plants. But Higgins wondered if there was a way to legally use the tests at the other plant facilities if he and Hill could establish the validity at the current location. At that point, Higgins finally managed to get some words out:

Where should I begin? We need to back off here a bit. At this time, we haven’t even established the validity of the tests for this plant, let alone the others. There are a number of activities that we need to perform in order to complete our validation study. Even then, there are a number of other issues that will need to

be addressed. I will say that the preliminary evidence looks good and appears supportive of validity for the test battery. However, this being said, the idea of extrapolating this study's results to the other plants is a whole different ballgame. We certainly will need to talk a lot more about that.

Smith appeared to be quite uncomfortable with what Higgins had said. "Well then, would you please detail to us exactly what needs to be done to start testing at this plant, and I mean soon." It was clear to everyone in the room he was agitated. "Also, will it be possible to use the tests, even if not in the short term, at the other four plants? We need to make this happen. Corporate headquarters is anxious to get this program online at all of the plants. Can you make this happen? If so, what will it take and when can we do it?" Smith and everyone else in the room turned to Higgins waiting for his response.

Appendix A

Additional Details Provided By Consultants: Validation Processes and Adverse Impact

(The following section represents a reproduction of the "lecture" that Higgins and Hill gave to the managers during the pre-proposal meetings in an attempt to clarify validation and the legal implications of employment testing.)

In a nutshell, criterion-related validity involves computing and testing the correlation(s) between one or more predictors (selection techniques) and one or more relevant outcome measures such as job performance (criterion or criteria). To the extent that a meaningful and statistically significant correlation exists between the two measures, a test or tests may be considered to have validity. For example, if a reading test is found to have a statistically significant correlation with success in training, the test may be considered a valid predictor of training success.

Content validity, on the other hand, focuses more on how much the content of the test and the content of the job reflect one another. For example, having an applicant for a driver's license take an actual test that involves driving an auto and making different sorts of maneuvers similar to what drivers normally encounter would be an example of a test with perhaps a relatively high degree of content validity. Similarly, a paper and pencil test about driving may also have some content validity though it would not likely be as high as an actual driving test. The content of the paper test is not "the same or as similar" as actually driving a car.

Construct validity requires establishing the similarity of factors (constructs) in the test with previously validated tests that measure the same factors (constructs). For example, if we were to try to develop our own test of intelligence, for the test to have construct validity, the results of our new test would need to be highly correlated with other already established (validated) tests of intelligence. For our purposes here at Northwestern, we would not use construct validity to establish the validity of our test. We may want to

consider using a test developed by a test publishing company where they have previously constructed and established the construct validity of the test. We could then use their test and attempt to establish the validity of the test for our jobs by using one of the other methods of validity.

The fourth validation method involves what is called the transportability of validity evidence. This approach is not as well known as criterion, content, and construct validities but it is considered an acceptable option to validate tests as described in the EEOC's Uniform Guidelines on Employee Selection and The Society for Industrial and Organization Psychology's Principles for Validation and Use of Personnel Selection Procedures. Transportability of validity involves efforts to establish that a validation study conducted at another organization or another facility for a job or group of jobs can be extrapolated as validity evidence for those jobs at a different location. An important premise of transportability is that the workers at the original location where the validation study was performed and workers at the other facility performed substantially the same major work activities or required similar job abilities. According to one expert [Mahaffey, 1996], the most important element to successfully transport validity is that an effective job analysis was performed that demonstrated that the job tasks or job abilities are similar.

Perhaps it is also worth mentioning that although testing experts generally distinguish among the different types of validity, in practice the establishment of validity often involves aspects of more than one validation method. Validation involves the "accumulation of validity evidence" and multiple validation approaches may be employed simultaneously. Overall, it is the totality of validity evidence that establishes validity. Validity can be thought of as a continuum where we look at the "degree" of validity evidence as opposed to a dichotomy such as "valid" or "invalid." Higgins came to realize that his long dissertation on the different approaches to validity was much more detailed than the managers cared to hear. At that point he finished his discussion of validity approaches but he felt a need to share additional information regarding adverse impact.

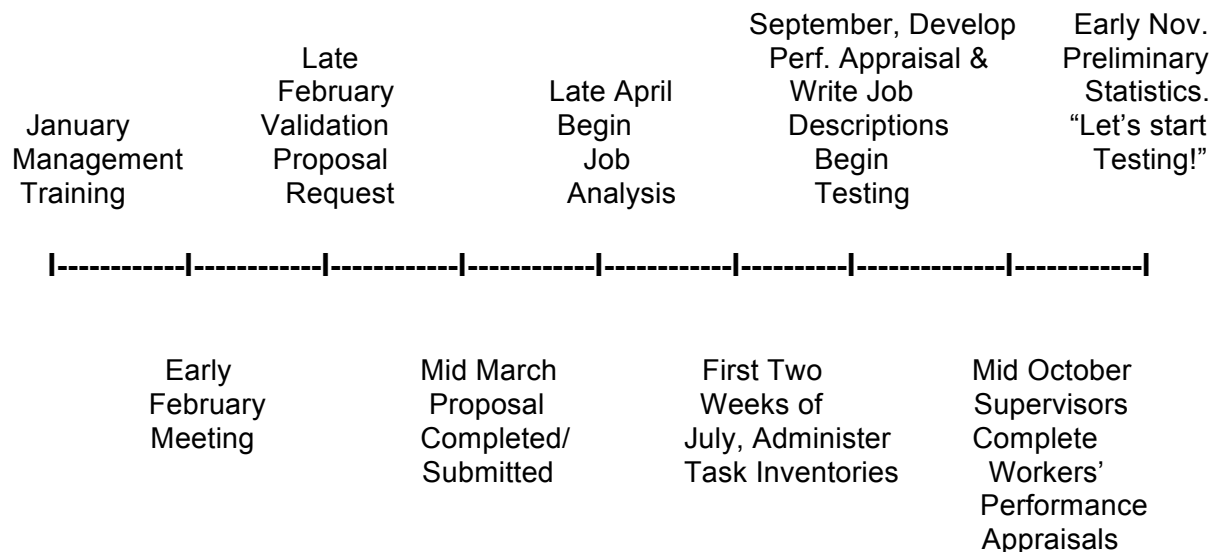
Higgins went on to discuss that most cognitive ability tests result in adverse impact. Adverse impact occurs in testing when a selection device selects proportionally more majority group members (e.g., Caucasians) than members of protected groups (e.g., African Americans or Hispanics). Adverse impact may also occur with differential selection rates between male (majority) and female (minority) applicants. The EEOC's Uniform Guidelines proposed the use of the 4/5ths rule (the 80 percent rule) to determine if adverse impact exists. The 4/5ths rule holds that if the selection rate for the minority group is less than 4/5ths the selection rate of the majority group, then adverse impact is presumed. For example, if an organization selects 20 blacks from among 50 black applicants (a selection rate of .40) and selects 60 white applicants from among 100 white applicants (a selection rate of .60), application of the 4/5ths rule indicates that adverse impact exists (.40/.60 is equal to .667 which is less than 4/5ths or .80). On the other hand, if 25 of the 50 blacks had been selected, adverse impact would not exist (.50/.60 is .83 which is greater than .80 or 4/5ths).

Adverse impact does not prove that a firm has illegally discriminated but merely establishes a prima facie (“on its face”) case of discrimination. Generally, employers can counter a prima facie case by demonstrating the validity of a selection instrument. However, the validation study must be professionally developed following the guidelines presented in the EEOC’s Uniform Guidelines and the APA’s Principles. Collectively, these references provide some guidance for using the validation approaches previously discussed.

Though it is generally considered desirable from a social and legal perspective to avoid adverse impact, valid selection methods may be used even when adverse impact occurs.

Appendix B

Case Timeline from January 2006 to November 2006



TEACHING NOTE

Conducting an Employment Selection Validation Study

Case Overview

Two consultants were hired to design and implement an employee selection system for a large sugar processing organization. They were faced with several misconceptions on the part of management, misconceptions that are fairly common among practicing managers. First, they had to convince these managers that employment testing is legal, when properly conducted. Then, they had to “educate” the managers about the requirements of testing and validation. The case also illustrates how a consulting team conducted a thorough job analysis and collected accurate performance appraisal data as part of their validation study. At the end of the case, the plant manager enthusiastically informed the consultants that testing would begin at the packing plant in ten days and that other company facilities would follow suit. The consultants were at a loss regarding how the managers could have arrived at the conclusions they had, given that they had proceeded as stated in the project proposal. The proposal described the “deliverables” (outcomes of the project) which included a number of concerns about the possibility of some or even all of the tests not proving to be valid. Moreover, the proposal clearly stated that the scope of the validation study was only for the one plant location. Now, the plant manager was eager to begin testing, even though all the validation work had not been completed. Additionally, he expected the program to be implemented at several facilities, although the necessary steps for transporting validity from one site to another had not even been discussed, let alone performed. The consultants now have to decide how to explain to the manager that his expectations for the project must be tempered.

Intended Courses and Levels

This case can be used in a variety of HRM courses including the Introductory Human Resource Management course (Junior/Senior level). It should be used after coverage of the Equal Employment Opportunity and Employee Selection/Testing chapters. The case may also be used effectively in higher-level Human Resource Management courses including Advanced Human Resource Management, Human Resource Staffing, or Human Resource Selection.

Case Objectives

1. To expose students to the complexities of designing and validating an employee selection/testing system.

2. To demonstrate the usage of job analysis techniques and performance appraisal techniques and their importance to the employee selection process.
3. To expose students to the various types of validity and the integration of multiple types of validity evidence.
4. To expose students to the requirements for transportability of validity evidence from one organizational setting to another.
5. To illustrate the importance of valid selection techniques to achieving strategic organizational goals. (By developing and implementing comprehensive job analyses and selection devices the company may attain a more productive workforce, one that is capable of more "big picture" understanding and actions.)

Key Learning Points

In addition to the Case Objectives and the Questions (that follow this section), the authors identify some Key Learning Points (specific HRM knowledge and applications) that the case presents to students. These include:

- The need to identify (internal and/or external) customer needs.
- The importance of effectively communicating project resource requirements to customers.
- The importance of keeping customers informed of program deliverables, scheduling milestones, and project completion dates.
- Recognition that employers have an increasing need for workers with higher aptitudes and abilities.
- The need to effectively explain to customers that validity is a matter of degree.
- Recognition that validity refers to the accumulation of validity evidence and involves inferences.
- Recognition that validation evidence includes quantitative and qualitative analyses.
- There are many methods for collecting validation evidence including content, criterion, construct, and transportability of validity.
- The EEOC's Uniform Guidelines on Employee Selection Procedures (1978) provide many procedures for conducting validation studies.
- The Society for Industrial and Organizational Psychology's Principles for the Validation and Use of Personnel Selection Procedures provide the testing profession's perspective regarding appropriate procedures for conducting validation studies.
- The importance of conducting job analysis to identify the job tasks and skill requirements.
- How job analysis information is useful for choosing which tests to validate.
- Tests that result in adverse impact need to be validated.
- A common method for determining adverse impact is the 4/5ths rule, also called the 80 percent rule.

- Selection cut-off scores can impact whether a test has adverse impact or can impact the degree of adverse impact a test has.
- A selection testing program requires consideration of test security and retesting policies.
- Invalid tests are not effective predictors of performance and are not useful for employee selection.
- Reliable and accurate performance appraisal ratings (or other important outcome measures) are necessary for criterion-related validity.
- Selection, career paths, and retention/employee turnover are related.
- Upgrading a workforce requires a strategic perspective and consideration of organizational and HRM systems (i.e., recruiting sources, relevant labor markets, and compensation).
- Labor unions impact an organization's ability to make changes and/or implement new systems.
- Test validation is a time consuming and complex process.
- Incremental validity is very important to assess and decide whether a test should be included in a test battery.
- Generally tests that have been validated at one organizational location or at another organization should not be used at a different location or organization without first demonstrating transportability by conducting a transportability of validity study.

Questions

1. What should Higgins tell Smith needs to be completed before Northwestern can start employment testing at his facility, and when they can begin?

Higgins' arguments should focus on two main areas. Some of this will be a reiteration and expansion of what Higgins stated in the meeting. First, he should detail to Smith that correlational and multiple regression statistical analyses (validity) need to be completed and evaluated to ensure that sufficient validity exists for the tests before using them at the current facility. There also are other types of analyses and documents that need to be completed. For one, Higgins and Hill need to examine the intercorrelations among the tests to determine how efficiently and effectively the tests work together. This would reveal the degree of incremental validity (the additional total validity when a test is added to a battery). In addition, Higgins and Hill should evaluate cut-off scores for the different tests and they also should determine the extent of adverse impact, if any, for different protected groups (e.g., African-Americans, Hispanics, Pacific Islanders and females). They should also examine the effects of different cut-off scores because different cut-offs often result in different levels of adverse impact. They would then have to develop a Test Users' Manual that addresses issues such as test administration, scoring, security, and retesting policies. Lastly, if the results demonstrate sufficient validity for any of the tests, they would have to document the results in a technical validation report as required by and described in the EEOC's Uniform Guidelines on Employee Selection Procedures.

Higgins should remind Smith that the project proposal specified that the validation study was being performed for the current plant location (scope). In addition, Higgins should reiterate that the project proposal specified the time frame for completion, the deliverables, and the resources needed from Northwestern to successfully complete the project. These were detailed in the case:

By mid-March, the two researchers had completed a project proposal that included details pertaining to scope of the project, deliverables, caveats associated with validation study outcomes, a project schedule indicating the start dates, duration, and completion dates of major project activities, resources needed from Northwestern (management support, employee time required to complete documents and to discuss work content with the consultants), and the project budget.

Higgins should also remind Smith of their previous discussion regarding project caveats. Higgins had informed Smith and others in the February, 2006 meeting that completing a validation study does not guarantee that the tests included in the validation study will be found to be valid. Rather, it is a study to determine if the tests are valid and can be used effectively in a specific situation. The following excerpts from the case describe that conversation.

Higgins and Hill cautioned them [Smith and others] that there were no guarantees that a validation study would support the validity (job-relatedness) of a test or test battery. The purpose of a validation study is just that – to “study” prospective tests to find whether the tests are valid in the employer situation for the job or job classes studied. Only then may tests that result in adverse impact be used as evidence of job relatedness. Despite these caveats, Smith requested that Higgins and Hill develop a validation study proposal and budget.

Moreover, Higgins should also point to other details in the proposal that are pertinent to Smith’s concerns and frustration. These provide support for Higgins’ argument that he is unable at this time to unequivocally declare that the tests are valid, given that all necessary statistical tests had not been completed. Between the meeting information and the proposal, Smith should have better understood this.

Higgins could take this a step further. Given that Higgins and Hill appeared pleased with their preliminary statistical analyses and presumably had a great deal of non-quantitative job analysis information that supported validation, it might be safe to assume that they believed validity would be demonstrated. During that final meeting with Smith, Higgins stated, “I will say that the preliminary evidence looks good and appears supportive of validity for the test battery.” If that is truly the case, Higgins might mitigate some of Smith’s frustration by again communicating to Smith that he and Hill believed the tests would prove valid, even though they could not state that with complete confidence without completing further analyses. Moreover, there is information in the case that indicates some content validity for the tests as well. Content validity evidence, if sufficient (i.e., “accumulation of validity evidence”) could be used to support

test validation. However, because a concurrent validity study has been conducted, its results should also be supportive of the validity of the tests. More details pertaining to the accumulation of validity evidence using more than one validation method is included in the discussion of question #4.

Finally, they could communicate the planned completion date for the project, as stated in the proposal. This would be the time that testing could begin at the current plant. A specific date was not provided in the case so we will not give one here.

2. How should Higgins respond to the question regarding the possibility of using the tests at the four other plants? Is this a possibility? If so, what would it take to successfully complete a transportability of validity evidence study?

Higgins could address this by indicating that this would be possible but only with additional resources, time, data collection, and analyses. Assuming that validity is established for the tests at the current location (which it appears it will be), transportability of validity studies would need to be conducted at each of the other four plants. In order to “transport” validity evidence to the other sites, thorough job analysis information must be obtained from each of those locations. This would be necessary to establish similarity of the jobs or job abilities at these sites to those at the current site (Smith’s plant). If the jobs are found to be essentially the same, then arguments of transportable validity may be made.

Appendix A of the case describes transportability of validity:

Transportability of validity involves efforts to establish that a validation study conducted at another organization or another facility for a job or group of jobs can be extrapolated as validity evidence for those jobs at a different location. An important premise of transportability is that the workers at the original location where the validation study was performed and workers at the other facility performed substantially the same major work activities or required similar job abilities. According to one expert [Mahaffey, 1996], the most important element to successfully transport validity is that an effective job analysis was performed that demonstrated that the jobs tasks or job abilities are similar.

Higgins should reiterate this information to the manager. He should discuss in general terms the types and extent of job analysis activities that would need to be performed to assess the similarities of jobs or job abilities at the other four plants compared to the current location. Higgins and Hill knew (as stated in the case) that the other four plants were also sugar processing plants. They could probably safely have assumed that the jobs were likely to be the same, or at least very similar to, those at the current location. Higgins might proceed by asking the managers present in the meeting if they could describe the types of jobs performed in the other plants and whether they appeared to be the same or similar to those at the present plant. Just as validation involves the “accumulation of validity evidence” the same notion would apply to transportability of validity evidence. Although there are no guarantees, assuming the jobs are in fact

similar, the more a firm invests in the study, the greater the likelihood of accumulating evidence that would withstand the scrutiny of the Equal Employment Opportunity Commission (EEOC) and the courts.

Higgins could suggest that if Northwestern would consider funding another project, that he and Hill would develop a second project proposal -- to complete a transportability of validity study at the other four plants. Only after that could they assess whether the tests were valid at the other locations. At that point, any estimate for completion of a transportability study would be very speculative.

3. Was the extensive nature of the job analysis conducted at the facility necessary or sufficient to collect information to help establish validity? Also, was it necessary for Higgins and Hill to obtain a new set of supervisory ratings of the workers for the validation study?

Meaningful job analysis data and performance appraisal data are essential to establishing the validity of selection instruments. A thorough job analysis uncovers the specific tasks, duties and responsibilities (TDRs) that make up the jobs as well as the necessary knowledge, skills, abilities and other personal characteristics (KSAOs) necessary to perform them. With documentation of these TDRs and KSAOs it is possible to develop a logical argument for why certain selection instruments were chosen: Because they measure the KSAOs necessary to perform the jobs in question. The same kinds of job analysis data also may be used, as was discussed in answer #2, to show the similarity of jobs across sites when one attempts to transport validity evidence from one location to another. Accurate performance appraisal data are crucial because these data serve as the measures of job performance that are compared to test scores in the criterion related (statistical) validation process. Statistical validation essentially demonstrates that test scores are correlated with, or predict, job performance measures. If these performance measures are not reliable, the observed correlations between them and the test results would be lower than the true correlations. This could result in obtaining statistically insignificant correlations which would not support the validity of the tests. An important psychometric concept is that reliability is a necessary but not sufficient condition for validity. Hence, reliability is important! As noted in the case, the consultants and managers agreed that new performance appraisal ratings data were necessary because the managers believed that existing performance measures available at the company were very likely biased, and therefore not accurate.

4. Are the different types of validity mutually exclusive or can they be used collectively to establish the validity of a test or test battery? Did the consultants rely solely on criterion-related validity or were any other validation approaches also employed to establish the potential validity of the test battery?

Though the EEOC's Uniform Guidelines identify and discuss a variety of validation approaches, including content validity, criterion-related validity, construct validity, and transportability of validity as if they are separate and distinct, this is not the case.

Evidence across these types can support and supplement one another. Validity is a unitary concept and the purpose of validity studies is to establish “accumulated” validity evidence to support the inferences drawn that tests predict performance (or some other meaningful outcome such as absenteeism). One might think of validation as “the degree of validation evidence” rather than some dichotomous classification (validity or no validity). The different types of validities are really different strategies for collecting validity evidence. Therefore, a comprehensive validation study might provide evidence using more than one strategy. Also, gaps in validity evidence resulting from one strategy might be supplemented with information and evidence obtained with a secondary strategy, or even a third type of validity evidence.

As noted in the discussion associated with question #1, content-related validity evidence was also collected by Higgins and Hill as they conducted extensive job analysis. They spent over one hundred hours observing workers perform their jobs and interviewing workers about their jobs. They developed a custom-made job analysis questionnaire that was based on the content of jobs performed in the plant. This questionnaire was completed by workers performing the different jobs in the packaging department. Based on this information, they wrote job descriptions for all packaging jobs. In addition, for the Operator Helper and Operator jobs, they developed a detailed task inventory rating form. They administered this task inventory to all of the workers who performed the Operator Helper and Operator jobs in the packaging department. The instrument asked the workers rate over a hundred task statements on both the frequency and importance of the tasks to the Operator Helper and Operator jobs. Higgins and Hill used the results to help them identify the important tasks, duties, and responsibilities (TDRs) of the two jobs and to determine the knowledge, skills, abilities and other factors (KSAOs) necessary to perform the jobs effectively.

Higgins and Hill used these data to develop a custom math test and they identified three commercially available tests that they determined were highly related to the work performed. These tests were to be used in their criterion-related validity study. Technically speaking, they did not have to perform the extensive job analysis to conduct the criterion-related study. However, by conducting the in-depth job analysis as they did, they were far more likely to select tests that would be job related and that would result in statistical significance and practical meaningfulness, thereby establishing validity.

In addition, the specific information they collected from all of the job analysis approaches was quantified in the task inventory ratings and it was documented in the job descriptions. As noted in the case, the content of all four tests reflected the content of the packaging jobs.

Collectively, the job analysis information provided the researchers with job-related information to assist in their decision about which selection devices to include in the testing battery. The testing devices were to assess the abilities and skills that Higgins and Hill had determined to be important competencies for workers to possess in order to effectively perform the tasks, duties and responsibilities of the Operator Helper and Operator jobs. The abilities and skills assessed by these tests were also deemed

important to other packaging and processing department jobs. Specifically, Higgins and Hill decided to conduct a concurrent validation study in which they would test current workers using a custom-developed math test and three tests available from test development/publishing companies: a reading test, a mechanical ability test, and a manual dexterity test. As stated in the case:

The math test measured an applicant's ability to count, add, subtract, multiply, divide and to compute percentages, fractions and decimals. The questions reflected the types of mathematical calculations the workers performed on the job. The reading test measured an individual's ability to read and understand written directions and materials such as those that were common on the processing and packaging lines. The mechanical aptitude test measured a person's ability to perceive and understand mechanical processes in everyday life and in manufacturing, processing, and mechanical oriented work environments. The types of problems contained in this test reflected the types of mechanical processes workers in the processing and packaging department dealt with on an on-going basis. Finally, the manual dexterity test measured an individual's ability to make skillfully coordinated movements with one's hands, including the grasping and manipulation of objects. These types of activities were regularly performed by packaging and processing workers. Each of these tests measured behaviors, skills, and capabilities regularly required of plant workers, especially Operators and Operator Helpers.

Collectively, Higgins and Hill performed many activities and obtained results that provided them with content validity evidence. However, for whatever reason, they did not articulate this to Smith. Although they wanted to fully complete their statistical analysis associated with the criterion-related data and they wanted to complete their documentation of the process, they may already have had sufficient information (based on content validity) to justify the validity of the tests. Presumably both types of validity evidence will support the use of the employment tests at Northwestern. However, if the statistical analysis is contrary to the content validity evidence the cumulative evidence about the job relatedness of these tests would not be as substantial. Again, the key phrase is "accumulation of validity evidence."

Epilogue

The case ended with the following:

Smith appeared to be quite uncomfortable with what Higgins had said. "Well then, would you please detail to us exactly what needs to be done to start testing at this plant, and I mean soon." It was clear to everyone in the room he was agitated. "Also, will it be possible to use the tests, even if not in the short term, at the other four plants? We need to make this happen. Corporate headquarters is anxious to get this program online at all of the plants. Can you make this happen? If so, what will it take and when can we do it?" Smith and everyone else in the room looked at Higgins waiting for his response.

In response to Smith's questions, Higgins, similar to part of the answer provided in response to question #1, stated that testing at the plant would have to wait until all of the analyses and required documentation were finalized. He also reminded Smith about the project schedule that was provided in the project proposal and informed him that he and Hill were on schedule and that, if all things continued going well, they would be able to begin testing in four weeks. He also stated that the project scope had always been for testing at the current plant though he had hints that management would likely want to investigate the possibility of using the tests at one or more of its other locations.

Higgins informed Smith that the plan to use the tests at the other plants was highly risky without the completion of transportability studies that would support their use at the other locations. He went on to describe in general terms what transportability of validity was and what would be required, including thorough job analyses at the other sites. Without such analyses, Northwestern Sugar would have difficulty successfully defending its testing at the other plants if it were challenged by the Equal Employment Opportunity Commission.

Smith was surprised that he could not legally begin his testing yet at his plant or at the other plants. Still, Higgins and Hill convinced him to delay any implementation until validity was adequately established at Smith's plant. Smith also agreed to talk to executives at corporate headquarters about obtaining funding for transportability studies at the other plants. These studies would only be considered if the tests were successfully validated at Smith's plant.

Five weeks later, Higgins and Smith determined that all of the tests in the battery were valid. In addition, each test had incremental validity in that each added to the overall validity of the battery. They also completed all of the other analyses needed such as determining cut-off scores and determining if adverse impact resulted from the use of one or more of the tests. They also constructed the Test User's Manual and the validation report. Employment testing of job applicants began a couple of weeks after the completion of the validation study.

Higgins and Hill were asked to submit a proposal to conduct a transportability of validity study at each of the other four plants. Their proposal was accepted. They analyzed jobs one plant at a time. They analyzed the jobs at the other plants and determined that the jobs at each plant were the same or very similar to Smith's plant. In addition, the abilities and skills required to perform the jobs in each of the other plants were the same as those required at Smith's plant. Therefore, they felt they had sufficient evidence of transportability of validity to the other plants and Northwestern Sugar began using the test battery. The transportability studies at each plant lasted approximately two months. Upon completion of each, testing began. All plants were testing within nine months of completion of the validation study at Smith's plant.

At the time this case was written Northwestern managers were quite happy with the selection system and with the increased quality and performance of the newly hired

workers. Implementation of the ERP system has gone relatively smoothly, due in no small part to the increased quality of workers being hired and promoted through use of the new selection system. Also, managers were hearing fewer complaints about biased and capricious promotion processes because of the job descriptions and selection devices now in place. Workers were commenting about the improved quality of co-workers. Managers were beginning to realize the added benefits, beyond testing, that could be obtained from the thorough job analysis work that had been done. For example, the job analyses provided objective KSAO information that could be used to identify training needs. Smith and his managers felt they had taken a major, positive first step toward building an HR system that would help them achieve overall organizational objectives.

References

- Bennett-Alexander, D., & Hartman, L. (2001). *Employment law for business* (3rd ed.). Boston: McGraw Hill-Irwin.
- Dessler, G. (2008). *Human Resource Management* (11th ed.). Upper Saddle River, NJ: Prentice Hall.
- Equal Employment Opportunity Commission. (1978). Uniform Guidelines on Employee Selection Procedures, *Federal Register* 43(166), Washington DC.
- Gatewood, R. D., Feild, H. S., & Barrick, M. (2008). *Human Resource Selection* (6th ed.). Mason, OH: Thomson South-Western.
- Griggs v. Duke Power Company, 401 U.S. 424 (1971).
- Heneman, H. G., & Judge, T. (2006). *Staffing Organizations* (2nd ed.). Middleton: Mendota House.
- Mahaffey, C. (1996). *Validity Generalization: Will it become the fourth major line of validity or a pipe dream?* Glendale: Psychological Services, Inc.
- Mathis, R., & Jackson, J. (2008). *Human Resource Management* (12th ed.). Mason, OH: Thomson South-Western.
- Society for Industrial and Organizational Psychology, Inc. (Division 14 of the American Psychological Association), (2003). *Principles for the Validation and Use of Personnel Selection Procedures* (4th ed.). Bowling Green, OH: Author.