Test Validation Strategies

2

Mark LoVerde and Ken Lahti

CONTENTS

M. LoVerde (✉)
CEB, 1701 Golf Road, 3–900,
Rolling Meadows, IL, USA
e-mail: mloverde@executiveboard.com

K. Lahti
CEB, 555 North Point Center East, Floor 6, Alpharetta,
GA 30022, USA
e-mail: mail@kenlahti.com

C. Hanvey, K. Sady (eds.), *Practitioner's Guide to Legal Issues in Organizations,*
DOI 10.1007/978-3-319-11143-8_2, © Springer International Publishing Switzerland 2015

Validity is one of the most important considerations for an employer to take into account when using a test or assessment—yet it is often misunderstood. Validity refers to whether a test score or other information about a job candidate is useful for making predictions about that candidate's potential job behaviors (for example, on-the-job performance, attendance, turnover, or theft). In other words, validity evidence answers the question of whether information provided by the test can help you make better employment decisions. In the USA in particular, validity evidence can also be crucial to demonstrating the legal defensibility of employment decisions when those decisions could result in a charge of discrimination.

Our goal in this chapter is to provide readers with a basic understanding of validity and why it is important. We will describe:

- the various forms of validation evidence (aka, validation strategies) and their application, including alternatives to the more standard or traditional approaches,
- when the application of each validation strategy might be more or less appropriate,
- what each strategy requires in terms of activities and employer resources, and
- issues that might interfere with demonstrating validity evidence.

This chapter provides guidance to practitioners on how to properly conduct a validation study (i.e., what you should do). In contrast, the chapter written by Tonowski provides a very informative discussion of mistakes that practitioners should avoid (i.e., what you should not do).

2.1 Key Concepts

To begin with, it is helpful to discuss some terms that are relevant to the concepts of validity and validation. We will try to make this as painless as possible.

2.1.1 Test

Practically speaking—and for the purpose of this chapter—a test is often thought of as a process for gathering and evaluating information about an individual in which stimuli (items, questions, or tasks), scoring (evaluation of responses or behaviors), and benchmarks (for example, sten scores, percentile scores, score ranges, job "fit" or job-match scores, or pass–fail thresholds) are to some degree formalized and standardized. However, any instrument, procedure, or source of information that is used to measure an individual's employment qualifications and make employment-affecting decisions can be considered a test from a legal perspective. Evaluations made of a candidate's suitability while reviewing a resume, interview responses, or their appearance or demeanor are as much a *test* as a multiple-choice instrument containing math problems. Likewise, training completion requirements, job tryouts, and certification standards are tests. Thus, employers cannot avoid the legal ramifications of their human resource decisions simply by using less structured and more subjective hiring standards.

Further, the Federal Government's Uniform Guidelines on Employee Selection Procedures (which we will discuss shortly) equate tests and selection procedures. These guidelines apply to all employment decisions, which include not only hiring but also:

> ...promotion, demotion, membership (for example, in a labor organization), referral, retention, and licensing and certification, to the extent that licensing and certification may be covered by Federal equal employment opportunity law. Other selection decisions, such as selection for training or transfer, may also be considered employment decisions if they lead to any of the decisions listed above. (Sect. 7.2.B)

Standardized tests used for employment decisions often include one or more of the following:

- Cognitive ability or aptitude tests, such as those measuring math skills or numerical reasoning, verbal reasoning, deductive reasoning, mechanical comprehension, visual inspection skills, or spatial ability.
- Personality tests and tests of related issues, such as values, interests, integrity, work styles, and preferences. Character reference checks, including standards applied when checking a candidate's social networking profiles, could also be considered as falling in this category.

Personality tests that are developed for diagnosing mental disorders should generally not be used for employee selection, and there are plenty of alternatives designed for use with "normal" populations.
- Knowledge tests, common examples of which include written driver tests, nursing boards, realtor certification, or technical job-knowledge tests (e.g., database administration; project management processes; marketing best practices).
- Experience evaluations, such as application forms, minimal qualification standard evaluations, experience reference checks, and biodata tests.
- Physical ability/performance tests, which may include demonstrations of visual acuity, manual dexterity, strength, agility, stamina, etc. (Police and fire departments frequently use such tests).
- Simulations, in which candidates demonstrate their abilities by completing activities that mimic those required on the job for which they are applying. Examples are diverse, depending on the type of job, and can include such things as diagnosing a predetermined technical problem on a piece of test equipment, operating a flight simulator, role playing a sales call, or completing a situational judgment test.
- Performance tests, which are similar to simulations but in which candidates are required to directly demonstrate their ability by performing an actual job-relevant activity, such as typing/keyboarding, shooting at a target, welding, or operating a vehicle.
- Medical tests and drug screens, although these are typically administered post job offer.

2.1.2 Assessment

Confusion arises over the distinction between testing and assessment. Assessment is really the broader term, with tests being a form of assessment, although this distinction is not critical to the discussion of validation. In the context of employee selection and evaluation, the terms individual assessment and assessment center are often used to describe processes where several types of tests are used—typically including "high touch" types such as interviews and role plays—and the information is integrated by one or more trained assessors, often to produce a narrative report.

2.1.3 Validity and Validation

A common misconception about validity is that it represents how good a test is. Actually, validity refers to the usefulness of a test as a basis for specific decisions or actions.

How test scores will be used to help make decisions about people—that is, the reason for administering the test—is a critical focus of validity and test validation research. A test that is useful for selecting high-potential candidates for a leadership development program may be of little use for selecting welders or long-haul truck drivers. Giving a less extreme example, a test that is useful for selecting entry-level supervisors may be insufficient for identifying C-suite leaders. Thus, although test professionals and users often talk about a "valid test," a critical question that needs to be answered is, "valid for what?"

The example above also highlights that validity is not an inherent characteristic of a test but rather a characteristic of interpretations that can be made based on test information. Decisions based on misinterpreted information drawn from an otherwise "good" test are no more useful than those based on a "bad" test.

The following definitions of validity come from professional standards regarding test use and validation:
- "Validity: The degree to which accumulated evidence and theory support specific interpretations of test scores entailed by proposed uses of a test." (*Standards for Educational and Psychological Testing, American Educational Research Association*, American Psychological Association, and National Council for Measurement in Education (1999).
- "The essential principle in the evaluation of any selection procedure is that evidence be accumulated to support an inference of job relatedness. Selection procedures are demonstrated

> to be job related when evidence supports the accuracy of inferences made from scores on, or evaluations derived from, those procedures with regard to some important aspect of work behavior" (*Principles for the Validation and Use of Personnel Selection Procedures*, Fourth Edition, Society for Industrial and Organizational Psychology, Inc. 2003).

Both definitions emphasize that validity pertains to evidence supporting the interpretation of test scores in light of their intended use. The second definition, which is more specifically focused on employment testing, points out that those inferences must accurately reflect one or more important aspects of the job for which they are being made.

As we noted earlier, validity evidence demonstrates that a test score or other information about a job candidate is useful for making employment-related predictions or decisions. Validation of an employment test, then, is the process of accumulating and evaluating validity evidence to determine what and whether job-related interpretations or conclusions are justified.

Much of this chapter will discuss the three most common approaches, or strategies, for establishing the validity of decisions based on test information. These are known as *criterion related, content,* and *construct* validation strategies. We will also discuss alternative strategies for demonstrating validity evidence. These validation strategies are not mutually exclusive; since validation involves the accumulation of evidence, ideally more than one of these approaches is used.

The concept of the **utility** of a test represents attempts to describe and quantify the usefulness of a test in terms that are more meaningful to a business audience. As we will discuss regarding criterion-related validation strategies, a correlation coefficient is a measure of the statistical relationship between two measures. In personnel selection, correlations are usually computed between scores from a test and some meaningful business outcome. A positive correlation means that higher scores on a test are associated with higher levels of a business outcome (for example, job performance), while lower scores are associated with lower levels of that outcome. A negative correlation means the opposite—that higher scores on a test are associated with lower levels of a business outcome, which, in the instance of some outcomes such as theft, absenteeism, and turnover, is a good thing. The closer the correlation coefficient is to 1 or −1, the stronger the relationship; the closer to 0, the weaker the relationship.

A correlation coefficient can be a very good indicator of the validity of decisions based on a test. However, it is a statistical measure with which many people are not familiar and it is not expressed in terms of business outcomes. Early writings about test utility converted correlation coefficients into predictions about "hit rates," demonstrating the usefulness of a test by showing how it could increase the proportion of correct hiring decisions. These types of utility estimates are still useful today. Other measures of utility estimate the differences in outcome levels associated with different test scores, for example, comparing the turnover rates for people with test scores above and below a certain score.

More advanced utility models demonstrate the utility of test usage in financial terms. These models are based on the assumption that higher levels of individual job performance represent greater levels of financial benefit for a company. Thus, use of a test that predicts job performance can be shown to have financial benefits. The most sophisticated of these models incorporate additional information about costs associated with selection processes, hiring volumes, selection rates, success rates of new hires, and average tenure of employees hired to more accurately represent the financial return associated with use of a given test or other selection procedure.

A fundamental challenge for any financially based utility estimate is translating available measures of job performance into accurate estimates of their financial impact on the company. Even using direct financial measures of performance such as sales can be tricky given that not all of the value associated with a sale can be attributed to the performance of the sales person.

A true estimate of the financial value strictly of a sales person's performance would need to factor out not only the costs associated with that person's sales (e.g., product development, manufacturing, logistics, marketing, and sales costs) but also the value added by tangible and intangible product characteristics (e.g., brand reputation, product quality, or product ease of use). That being said, it is usually possible to produce reasonable and likely conservative estimates of business impact of assessment and selection systems using either existing business metrics (e.g., % quota achieved), specialized quality of hire ratings (e.g., new hire time to proficiency, rehireability), or even simple estimates based on salary levels for a given role.

The concept of **adverse impact** is closely tied to, and often confused with, the concepts of discrimination, fairness, and bias. The chapters in this book by Richard Tonowski and by Scott Morris and Eric Dunleavy delve into these topics in greater detail, but we provide a brief overview below:

- Adverse impact refers to a differential, or disparate effect of an employment practice on members of a protected group, for example, a lower hiring rate for certain racial or ethnic minorities, when compared to other groups. (Adverse impact is most often a concern for employers within the USA, given the nature of US employment laws.) Adverse impact is not, in itself, illegal, though evidence of it is often used by government enforcement agencies as preliminary—or prima facie—evidence of potentially unfair practices.

  When adverse impact exists against a protected group, to avoid charges of discrimination and the penalties that go with such charges, an employer may be required to demonstrate the job relatedness of the employment practice. Validity evidence is often used to do so, which is why we mention adverse impact in this chapter. Employers may also be required to demonstrate that they have looked for and not been able to find equally or more valid alternative practices (e.g., tests or other standards) that have less adverse impact.

- **Fairness** and **bias** are closely related concepts. The main difference is that there are standards for bias that can be quantitatively defined, are clearly implied in the federal government's *Uniform Guidelines on Employee Selection Procedures,* and have been accepted by US courts. Fairness, on the other hand, is a more subjective concept. We will discuss bias below. For a description of four possible definitions of fairness, refer to the *Standards for Educational and Psychological Testing,* American Educational Research Association, American Psychological Association, and National Council for Measurement in Education (1999).

  Bias refers to job-irrelevant sources of variance that result in systematic differences in test scores (e.g., higher or lower test scores between two groups that do not reflect true differences in ability). Differences in measures of physical strength between men and women would not be biased if they reflected true differences in men's and women's ability to perform aspects of a job requiring strength. Differences between Whites and Hispanics on a test of critical reasoning skills that were due to differences in the ability to read English would be biased, however, if the job required critical reasoning skills but not the ability to read English.

2.2 Legal Context

We will briefly and broadly discuss applicable laws, professional standards, and administrative guidelines that apply to employment testing. By cataloguing and describing these laws and guidelines, we hope to provide a basic orientation and framework for practitioners to appraise their testing and selection programs, and to provide a foundation for further learning. We are not attorneys, and the legal arena can get complex quickly; fortunately, though, in our experience good science and practice is highly consistent with the law. In particular, a strong focus on validity combined with clear documented processes and fair

and consistent treatment of candidates will usually keep you on the right side of the law.

2.2.1 Federal and State Employment Laws

Many laws at the Federal and State levels are applicable to employers, employment decisions, and therefore testing as a part of employment decisions. The major Federal laws that apply to testing are listed below. State laws are highly variable and individual states may place additional requirements on employers, create additional protections for employees, or even create additional protected classes of candidates (e.g., based on sexual orientation). When in doubt, consult with an attorney who is licensed to practice in the state in question and who has expertise in employment law.

The **Civil Rights Act of 1964** created a framework for the Federal government to regulate businesses and public institutions to ensure equality and fair treatment of people across a number of areas with huge potential, social and economic impact, including voting rights, education, public accommodations, and employment. In particular, Title VII of the Civil Rights Act codified that employers shall not discriminate with respect to race, color, religion, sex, or national origin. This law applies to all employers with 15 or more employees.

The Civil Rights Act was amended in 1991 in response to several US Supreme Court decisions regarding employment discrimination. The effect of the Civil Rights Act on testing is the requirement that tests should not unfairly discriminate against people in the above protected groups. This seemingly simple requirement understates the complexity inherent in making a determination of unfair discrimination. However, a simple shorthand is that tests (and other selection procedures) should be administered consistently with all candidates treated similarly (that is, no disparate treatment), and tests should show no average score differences across these groups (that is, no disparate impact or **adverse impact**). If score differences create adverse impact, the tests must be shown to be job related and equally valid for each group.

The **Age Discrimination in Employment Act of 1967 (ADEA)** prohibits employment discrimination against anyone who is 40 years of age or older and applies to employers with 20 or more employees. Similar to the Civil Rights Act, this law places a requirement on employers that testing programs do not show adverse impact against older workers and/or are shown to be job related and valid.

The **Americans with Disabilities Act of 1990 (ADA)** extends prior civil rights laws by prohibiting discrimination based on disability, defined as "…a physical or mental impairment that substantially limits a major life activity." An individual's protected status and whether a particular condition is considered a disability under the law is determined on a case-by-case basis. Employers are prohibited from discriminating against a "qualified individual with a disability" meaning that the person is able to perform the job, albeit possibly requiring *reasonable accommodations* or changes to the work environment, processes, policies, or other elements as needed to provide an equal opportunity for effective job performance in the context of the individual's disability. For example, a wheelchair user may have the interpersonal and computer skills to be an excellent call center representative, but may require a larger workstation to accommodate their wheelchair, which the employer would be required to provide.

With regard to testing programs, most of the focus related to ADA has been on helping employers provide reasonable accommodations to candidates who declare a disability and request accommodations during the selection process. Practitioners continue to struggle, however, to provide confident recommendations for accommodations that retain the validity of testing programs, as scant research has been conducted on the effects on validity of making accommodations such as providing extra time to complete tests, providing alternative delivery/formats of tests (e.g., read aloud by a reader/assistant or by

screen-reader software), or skipping or substituting entire steps in the process when no reasonable adjustment can be made that would preserve the integrity of the test (e.g., how to accommodate a visually impaired candidate on a Numerical Reasoning test that is reliant on graphics and tables as stimuli). Please refer to Gutman's chapter in this book for additional discussion on ADA compliance.

The Uniform Guidelines on Employee Selection Procedures (1978). The "Uniform Guidelines" (http://www.gpo.gov/fdsys/pkg/CFR-2011-title29-vol4/xml/CFR-2011-title29-vol4-part1607.xml) were jointly adopted in 1978 by the US Equal Employment Opportunity Commission, the Civil Service Commission, and the Departments of Labor and Justice as a uniform set of principles and standards to guide employers and practitioners regarding the design, validation, and appropriate use of testing and selection procedures. The Uniform Guidelines are highly prescriptive, providing explicit guidance on appropriate nondiscriminatory test user behavior, technical definitions of terms like "discrimination" that appear in legislation and have been debated in case law, technical standards for a number of validation strategies, and clear guidance on what does *not* constitute validity in the eyes of the US Federal Government. Although the Uniform Guidelines are not law, when legal cases arise involving testing and selection (under State or Federal laws, including the Civil Rights Act or ADEA, for example), the Uniform Guidelines are given great deference by judges, attorneys, and testing professionals/expert witnesses.

2.2.2 Professional Standards

A number of professional standards exist that guide professional practice and inform professional judgment with regard to test development, validation, and usage. Professional standards of practice are usually created by a committee of subject matter experts (SMEs) based on considerable review, debate, and discussion. Members of the profession are expected to promote and adhere to the standards, but the public and the courts also reference these as official declarations from experts in the profession regarding what is considered good or acceptable practice.

The Society for Industrial and Organizational Psychology (SIOP) created **The Principles for the Use and Validation of Personnel Selection Procedures (Fourth Edition, 2003)** (http://www.siop.org/_principles/principles.pdf) as a roadmap for professional practice in personnel selection. The purpose of "The Principles" is to specify best practices in selection that are consistent with established scientific evidence.

The **Standards for Educational and Psychological Tests (1999)** was jointly developed by the American Psychological Association (www.apa.org), the American Educational Research Association (AERA), and the National Council on Measurement in Education (NCME). The standards are considered by many professionals to be the "gold standard" regarding technical issues of test development and use in education and employment.

Other notable standards that provide guidance to practitioners with regard to test development, validation, and usage in employment decisions, include:

- International Test Commission (www.intestcom.org): International Guidelines for Test Use (2013); International Guidelines on Computer-Based and Internet Delivered Testing (2005); and International Test Commission Guidelines for Translating and Adapting Tests (2010)
- ISO 9000 Standards for Assessment (2013): Procedures and Methods to Assess People in Work and Organizational Settings
- International Task Force on Assessment Center Guidelines (www.assessmentcenters.org): Guidelines and Ethical Considerations for Assessment Center Operations (2009)

Case law also frequently comes into play in legal cases involving selection and other employment decisions. A suitable review of relevant case law would be extensive and is beyond the scope of this chapter. In addition, a problem with discussing case law in a book is that the landscape can

shift quickly with a single court decision, even between the time writing ceases and the book is published. However, for a brief review of some of the most important selection-testing-related court cases over the past 45 years, see Chap. 2 of the Gatewood, Field, and Barrick book listed at the end of this chapter.

2.2.3 Validity and the Law

In the USA, it is illegal to make employment decisions based on a person's race, color, religion, sex, national origin, age, or disability status; additional criteria (e.g., sexual orientation) vary by state. It is also illegal to apply different standards or treat people differently based on these characteristics when administering selection procedures used to make employment decisions (other than when making reasonable accommodations to comply with the ADA). However, it is important to note that, despite all the legislation regulating employment decisions and the numerous professional standards describing how to use science to make better hiring decisions, employers are not actually required by law to make good decisions about who to hire or promote, nor are they required to use valid tests or anything scientifically based to make employment decisions. They are simply prohibited from making decisions that discriminate unfairly against people who fall into one or more of the protected classes previously discussed. It is only when adverse impact occurs that an employer is required to justify its use of a test. Of course, aside from legal concerns, employers may want to base hiring decisions on validated selection instruments because the validation evidence provides them with proof that their selection instruments allow them to make better hiring decisions.

So, for example, assume an employer decided to hire only people who showed up to the interview wearing blue shirts. Although this seemingly arbitrary "test" and hiring criteria would likely be a poor business decision (analogous to selecting people randomly) and almost certainly would not improve hiring decisions or job performance, there is no applicable law that would prevent an organization from applying this policy to make (bad) employment decisions. …that is, unless wearing a blue shirt happened to be related to one of the protected groups (e.g., if men are more likely to wear blue shirts than women). If this were the case, then the organization would be unintentionally discriminating against women through consistent application of their hiring policy. Unless "wearing a blue shirt to the interview" was shown to be valid for predicting job performance, the employer would be violating the Civil Rights Act.

This example, although hopefully unrealistic, is illustrative of the dance between validity and the law. Intentional discrimination (i.e., disparate treatment) is expressly illegal. Disparate impact, or adverse impact, may or may not be illegal, depending on the quality and appropriateness of the test and selection procedures used. When it is caused by an employment practice and an employer cannot satisfactorily demonstrate the job relatedness of that practice, it may represent *unintentional* discrimination, which would still be illegal.

The dance starts with a plaintiff (or a class of plaintiffs, or an enforcement agency) who alleges employment discrimination. If plaintiffs can show clear intent to discriminate on the part of the employer (e.g., overt mistreatment, discriminatory comments made), then they can pursue a disparate treatment claim. If, however, they cannot show clear intent to discriminate (as is often the case), but they can show that there is adverse impact against a protected group resulting from the organization's policies and procedures, plaintiffs can pursue a disparate impact claim.

Under the theory of disparate impact, if adverse impact is "proven" by plaintiffs, that is, if statistics show that there are indeed meaningful differences in passing rates across relevant demographic groups for one or more steps of the employer's selection process, then the burden shifts. The defendant (employer) must now prove that the procedure being used to make employment decisions is job related and consistent with business necessity—in other words, valid for predicting job performance. If the test or other selection procedure is found to be valid, the burden shifts back

to the plaintiffs to make the case that equally effective but less discriminatory "alternatives" exist that the employer should have considered and implemented instead. Many employers have thus been encouraged by case law (and their in-house legal counsel) to conduct some form of search for alternatives during the design of selection testing programs, and to document what types of tests were considered along with why they decided on the final test battery they are using. To be clear—the legal requirement is not that employers use only tests that have no adverse impact, but simply that they should search for alternative procedures that would be equally effective with less adverse impact and use such tools whenever feasible.

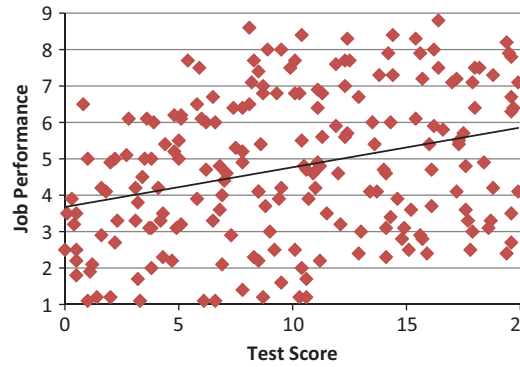
2.3 Validation Strategies

We now turn our attention to the three most common approaches, or strategies, for establishing the validity of decisions based on test information. As we noted earlier, these are known as *criterion-related, content,* and *construct* validation strategies. They do not represent different "types" of validity, but rather three complementary ways of building the case for the validity of decisions based on a test.

2.3.1 Criterion-related Validity

Criterion-related validity evidence is primarily statistical in nature, and is generally based on correlation coefficients, which we described earlier. The underlying concept is that levels of test performance are statistically related to levels of a job-relevant outcome. Most commonly criterion-related validity evidence is based on the demonstration of a positive relationship, or correlation, between test performance and job performance; although, as we noted earlier, the desired relationship can be negative when the outcome is something the employer wants to reduce (e.g., higher levels of test performance associated with lower levels of absenteeism or time to successfully complete training).

The graph below shows the relationship between scores on a test (that range from 0 to 20) and performance scores (that range from 1 to 9). The correlation is 0.30, which is in the moderate range for test validities.



The graph shows two things that are important to note:

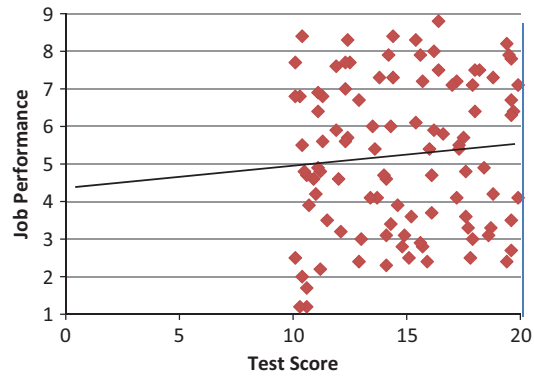
- First, higher test scores are associated with higher job performance, on average. The line on the graph reflects the linear relationship (termed a regression line) between test scores and job performance such that expected job performance can be estimated for any given test score. The regression line slopes up to the right, denoting that on average higher test scores are associated with higher job performance. Note for example that, even though there is a lot of variability at any given test score, nobody with a test score below 5 fell in one of the top three performance levels (i.e., 7, 8, or 9), nobody with a test score below 8 fell in one of the top two performance levels (i.e., 8 or 9), and nobody with a test score above 11 fell in one of the bottom two performance categories (i.e., 1 or 2). So, even though the data are messy (and they always are for anything less than a perfect correlation of +1.0), there is still a clear pattern or trend of higher test scores being related to higher performance.
- Second, the chart also clearly shows that use of a test (or any selection standard) will not result in perfect prediction. For a given hiring decision, a person who scored lower on

the test might perform better than another person who scored higher on the test, because the correlation is not perfect and there are other determinants of individual job performance besides just this test. However, use of the test will result in higher job performance on average (that is, better hiring decisions across a number of people) than not using the test. Put another way, use of an appropriate selection test can dramatically increase the odds of making a successful hiring decision.

One advantage of using a criterion-related validation approach is that the evidence used to demonstrate the statistical relationship between a test and a criterion can also be used to provide a rationale for setting qualification standards, that is, a "passing" or "cut" score. Look again at the regression line in the graph above. If a performance score of 5 (on the vertical axis) was deemed by the company as "acceptable," people scoring below 12 on the test (the horizontal axis) would on average have job performance that was less than acceptable, while those scoring above 12 on the test would, on average have performance that was more than acceptable. Thus, a passing score of 12 could be set on the test to hire people who would achieve acceptable or higher performance levels, on average. For a more detailed discussion of setting and using cut-scores, see the chapter in this book by Lorin Mueller and Liberty Munson.

Some problem conditions can affect (generally reduce) an employer's ability to detect the true relationship (correlation) between a test and a desired outcome. Among the most common of these are:

Range restriction: As the name implies, this refers to a limitation in the range of test or job performance scores from what normally would be expected to occur. For example, the range on test performance might be restricted if people with low scores on the test (or another, closely related test) were previously selected-out (not hired) before the validation study, with the result that information about their job performance is not available. Such a situation is shown in the graph below where (using the same data as the graph above) people with scores below 10 were excluded. It is certainly harder to see the general trend of higher test scores relating to higher performance scores and, in fact, the correlation you would find in this situation is 0.11, well below the 0.30 we found without range restriction in our sample. In real world situations, you are also likely to have range restriction on the criterion of job performance, as low-performing employees may have been terminated while top-performing employees may have been promoted. It is worth noting that, in rare situations, range restriction might result in finding an artificially higher correlation, such as when an employer decided to include only top and bottom performing employees in a validation study while leaving out employees performing in the middle.



Representativeness of the people in a validation study sample (the degree to which they are like or unlike job applicants for whom the test will ultimately be used) can also affect its findings. In some cases, lack of representativeness can result in range restriction. This might occur if, for example, an employer that hired from the general population conducted a validation study only on college students applying for summer jobs; if those students' range of cognitive ability was higher than that of the typical applicant, the validity estimates from this study would not match what would be expected among the real applicant population. Lack of representativeness can also occur if other demographic characteristics of validation study participants differ from those of typical applicants (e.g., the study was conducted in a plant in rural Idaho but most of the employer's facilities were in urban areas on

http://www.springer.com/978-3-319-11142-1

Practitioner's Guide to Legal Issues in Organizations

Hanvey, C.; Sady, K. (Eds.)

2015, XVI, 325 p. 11 illus., Hardcover
ISBN: 978-3-319-11142-1