



# Standards for Evaluating the Quality of Assessment Tools

Charles A. Handler, Ph.D.  
President/Founder, Rocket-Hire

## Executive Summary

This whitepaper attempts to provide an understanding of the concept of quality as it relates to the use of assessment tools. The inherent complexity of assessment tools and recent increases in their use have created a need for this type of information. This paper provides an overview of what assessment quality means and provides information designed to help organizations better evaluate the quality of assessment tools. This goal is accomplished via the presentation of the following information.

## Section I: Introduction to the Problem

Recent advances in hiring technology and practice have continued to demonstrate the ability of correctly implemented tools and processes to help organizations meet their strategic objectives. This evolution of products and processes have created a growing demand for a wide range of products focused on providing solutions to a variety of different specific issues related to the hiring process. Assessment tools represent one such tool, and although they have been successfully used for decades, interest in their utilization has increased substantially over the past several years.

While there are many potentially positive outcomes associated with assessment tools, their use can be a relatively complex proposition. One of the most common issues faced by organizations interested in using assessment tools is the fact that there is a good deal of variation in the quality of these tools. This can be problematic because many organizations lack the experience and methodology necessary to make accurate evaluations of assessment quality, thus creating a situation in which assessments may end up "underachieving" when it comes to demonstrating results.



The purpose of this paper is to explore the concept of assessment quality in order to help those interested in using assessment tools to make more informed decisions that will result in a greater ability of these tools to help provide value. In order to set the stage for the discussions provided in this paper, our working definition of quality is as follows:

*Quality assessment tools provide organizations with the ability to systematically make accurate predictions regarding the ability of potential applicants to perform the requirements of a job within a specific environment. The greater the accuracy of the predictions that result from the use of an assessment tool, the higher its quality.*

Given this definition, the quality of any assessment tool is directly related to its ability to assist organizations in ensuring its hiring process results in effective decision making. Unfortunately, there are a variety of factors which can make it difficult for organizations to make sound decisions when attempting to select assessment tools. These include the following:

**Lack of governance.** While there are standards and guidelines which have been created to help guide the implementation of assessment tools, there remains almost no governance related to who can sell an assessment tool or even what an assessment tool actually is. There are presently no limits on who can sell tools that can be used to help make selection decisions.

**Bad press.** Negative experiences related to the use of poor quality tools and a lack of understanding of the fact that some tools are simply not designed to be used in specific situations has led to some negative publicity surrounding the use of assessment tools.

**Lack of understanding.** Assessment is a complex subject matter and even the most carefully constructed assessments will fail to deliver value if not used correctly.

**Increasing demand.** The evolution of hiring practices has placed a spotlight on the use of assessment tools. As more organizations begin to realize that the use of these tools is essential for truly delivering value through the hiring process, interest in this area is growing. Unfortunately, this has created a situation in which there has been a proliferation of new products and approaches. While many of these are highly effective, others represent solutions which have been hastily created in order to "cash in" on this trend.



Given these issues, the need has become apparent for information which can help consumers of assessment tools understand how to evaluate the quality of these tools. The following section provides a detailed discussion of the concept of quality as it relates to assessment tools.

## Section II: Definition of Assessment Quality

The need is clear for information that can help provide organizations with the ability to more fully understand the concept of assessment quality. This section provides an overview of the concept of assessment quality, first discussing existing standards put forth by the US Government and then proposing an additional model to help supplement these. This model draws on a combination of both methodology and practicality to define assessment quality as a matter of contingency requiring an awareness of the situation in which the assessment tool is to be used.

### Existing Standards for Assessment Quality

While there are a variety of unwritten rules for defining assessment quality, the only actual real standards that have been created are the Uniform Guidelines on Employee Selection Procedures published by the US Government (<http://www.uniformguidelines.com/>). The mission of these guidelines is to:

*"Incorporate a single set of principles which are designed to assist employers, labor organizations, employment agencies, and licensing and certification boards to comply with requirements of Federal laws prohibiting employment practices which discriminate on grounds of race, color, religion, sex, and national origin. They are designed to provide a framework for determining the proper use of tests and other selection procedures."*

These guidelines are several decades old and while they do provide an important set of standards, they may be difficult for untrained individuals to fully comprehend. For the most part, the Guidelines focus mostly on the idea of test validation (i.e., verifying the fact that the test measures what it is supposed to measure) and what forms of validation are acceptable. These Guidelines serve as "the standard" when it comes to the technical evaluation of acceptable personnel testing procedures. While these provide excellent standards for organizations creating assessments, they are not very "consumer oriented" and do not offer organizations a clear path to understanding how to choose a particular assessment.

In 1999, the US Department of Labor (DOL) released an accompanying document, *"Testing and Assessment - An Employer's Guide to Good Practices"* (available at no cost from the



Uniform Guidelines website). This 80 page document is basically a "field guide" that communicates the legal context, laws, and regulations that surround testing as well as provides information on the statistical aspects of test quality, how to select an assessment test, how to use and score assessments, and how to evaluate the effectiveness of an assessment test. This document offers a very solid foundation for determining what is legal and what is not when it comes to testing and assessment and is highly recommended reading for those interested in better understanding what it takes to "do it right" when it comes to using tests and assessments.

The information in the "Guide to Good Practices" is directly extracted from the Uniform Guidelines and provides a good set of clear ideas regarding the use of assessments. However, this document alone does not offer the information that may be required to help an organization evaluate the quality and relevance of a specific assessment tool which they may be interested in using.

This level of understanding requires organizations to take a more active role in evaluating assessment tools relative to their specific needs.

### **A Practical Definition of Assessment Quality**

Unfortunately, the idea of Assessment Quality is not a simple one. There are many factors involved in this determination and quality goes well beyond the technical aspects of the assessment itself such as those offered by the DOL.

This paper views "Assessment Quality" as a broader idea, one which goes beyond the mere technical construction of a test to include the concept that "doing it right" requires carefully selecting the right tool for the job.

Specifically, assessment quality is defined as having two facets. The first of these involves the methodology used to construct the tool itself while the second represents the contingencies surrounding the proposed use of the tool.

**Area One: The technical dimension.** This aspect of quality is defined by best practices in test construction and technical standards including those put forth by third party organizations such as the Association of Test Publishers and the American Psychological Association. The material such as that found in the Uniform Guidelines also provides a good foundation for the technical evaluation of assessment tools. While a detailed description of these sources is beyond the scope of this paper, there are some basic tenants on which the assumption of technical assessment quality should be based, as they represent

some of the major things that are required in order to support the assumption that a test has been properly constructed. From a technical perspective, a quality assessment should be accompanied by the following:

- An operational definition of what the test measures
- Theoretical links explaining why the test is a reasonable measure of these things
- Documentation of psychometric analyses conducted during the test construction process
- Evidence indicating that the test measures what it is supposed to measure (i.e., validation evidence)
- Evidence that the test does not predict differentially for members of certain protected groups (a concept known as adverse impact)
- Additional evidence summarizing results that can be expected when using the test to measure the constructs it is supposed to measure (summary of ROI and the direct relationship between various levels of test scores and job performance).

The above should be contained in a test manual or test validation document, the evaluation of which is an absolute necessity for anyone who is interested in using a specific assessment tool. This documentation should be carefully evaluated to help provide an understanding of the care and technical procedures that went into the creation of the test.

Despite the primary importance of technical accuracy, the changing nature of the world of employee selection has increasingly begun to show that meeting the above standards are not enough. Evaluating the technical suitability of products is definitely a critical first step in the selection of an assessment tool. Technical standards are a baseline which provide some piece of mind and offer a good filter in that any organization that cannot provide the above information for their test or assessment they are offering for sale is basically failing to offer quality.

While a technical evaluation may help to identify those organizations whose products are simply not legitimate, most assessment vendors are able to provide this information to interested consumers in some form or another. However, there is no set formula for the presentation of the above information and because these studies are often presented as part of a sales cycle, they cannot be trusted to be the sole source of quality. So, they are a necessary, but not sufficient part of the quality equation. This paper suggests that, while the technical dimensions are definitely a pre-requisite, quality is really a reflection of the



relevance of the assessment tool to the specific needs of the organization. This is referred to as the "contingency dimension" of quality.

**Area Two: The contingency dimension.** The recent explosion of interest and availability of assessment tools has led to the need to evaluate assessment quality in a manner that goes beyond a mere review of the technical documentation that accompanies an assessment tool. The fact is that the majority of tools that are commercially available in today's assessment marketplace will meet minimum technical standards, with most being relatively well constructed. This means that there must be another dimension to help define assessment tool quality.

The contingency dimension is all about the manner in which the assessment tool is deployed relative to the specific needs of the consumer. A description of a full methodology for determining the specific parameters of contingency is beyond the scope of this paper. However, evaluating contingencies associated with the use of an assessment tool generally involves an understanding of the following:

**The key elements of the job (or jobs) for which the tool will be used.** This involves the level of the job, the competencies, knowledges, skills, abilities, etc that are to be measured using the tool. Effective prediction for different types of jobs requires the use of different types of tools.

**The outcomes that are desired from the use of the tool.** This involves the actual performance criteria or objectives that relate to the use of the tool. These can vary quite a bit as different types of tools are designed to help the user achieve different outcomes.

**A consideration of the tradeoff between speed and accuracy.** This is perhaps one of the most important aspects of the contingency dimension. The decisions made in determining how important it is to customize an assessment tool for use in a specific situation vs. using a tool that is more generic will have a tremendous impact on the overall quality of the assessment tool?

**The manner in which the tool fits into the existing hiring process.** This relates to the philosophy with which the tool is implemented relative to the overall hiring process. That is, will the tool be used as a single test or will it be used as part of a more integrated process in which test data can be combined with other information in a manner that is supported by decision making tools.



**An understanding of how the effectiveness of the tool will be measured.**

The manner in which the effectiveness of the tool will be verified is an important consideration when choosing a tool. For instance, smaller organizations may not have the volume of applicants required to support empirical evaluations of the effectiveness of the tool. This can have a definite impact on an understanding of the quality of the overall process.

At the end of the day, a consideration of the above factors when selecting an assessment tool will have a major impact on the results that are returned. In other words, quality is about using the right tool for the right job but also about finding a tool that fits with the considerations driving the need for the tool itself.

Ultimately, it is the combination of the technical dimension and the contingency dimension that will end up defining the ability of the assessment tool to provide optimal outcomes that will offer value to the organization. The following section describes the manner in which an evaluation of both dimensions can help organizations optimize assessment quality as it relates to their specific needs.

### **Section III: Optimizing Quality**

At the end of the day, the concept of quality as it relates to the use of assessment tools is all about prediction. The extent to which the tool is able to accurately and systematically predict the outcomes for which it was implemented is what will define its quality. This concept requires an understanding that, while theoretically possible, from a practical point of view perfect prediction using assessment tools is simply not possible. After all, employee selection deals with humans whose individual differences make them unpredictable.

Despite the lack of ability to achieve perfect prediction, the accuracy with which predictions are made can be controlled. This is extremely useful because the more accurate these predictions are, the greater the returns on the investment made in the use of the assessment tool. This paper suggests that the combination of technical accuracy and the extent to which contingency parameters are considered creates a "quality continuum" with which the quality of the tool will be defined by its ability to offer a high degree of accuracy in prediction. Although there are an infinite number of possible points on this continuum, this paper will discuss quality in terms of the three key anchor points on the continuum: high quality, medium quality and low quality. Each of these is described in more detail below.

Figure 1 demonstrates the relationship between the content of an assessment tool and job performance that can be associated with high quality solutions. The idea behind this figure is a simple one; the higher the degree of overlap between the construct measured by the assessment and the constructs that define job performance, the more closely the circles will overlap.

High quality solutions will first and foremost meet all technical requirements in terms of test development and construction. This is a pre-requisite. However, even the most carefully created tool will offer no value if not used correctly. High quality solutions are generally those that are closely mapped to the job in question via thorough job analysis process, have evidence to support their ability to predict accurately for similar jobs, and are used as part of a process as opposed to one isolated piece of information.

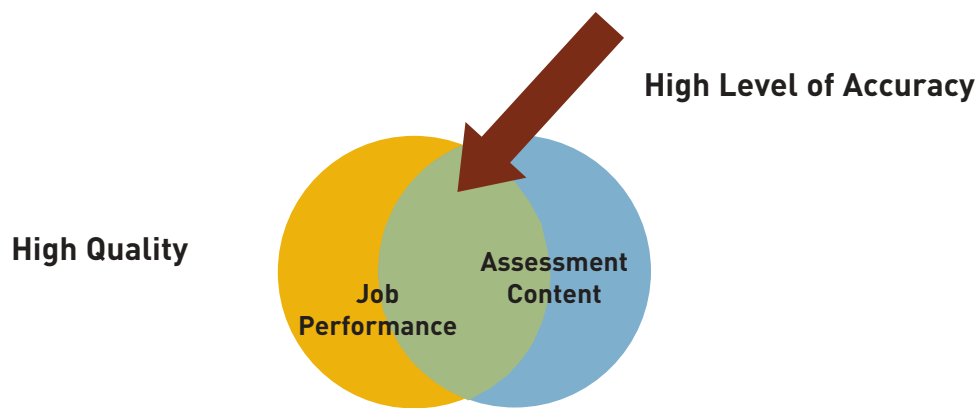


Figure 1. High Quality

Solutions meeting the above criteria will generally deliver the most accurate levels of prediction but at a cost. This cost being the fact that it takes time, money and a good deal of effort to ensure the assessment has been properly created and that the tool is "dialed in" to the unique aspects of the job or organization for which it will be used. The tradeoff here is that, up to a point, the greater the effort put forth to deploy a high quality solution, the greater the returns on the investment made in it will be.

Figure 2 represents the relationship between the content of the assessment tool and job performance that can be associated with medium quality solutions. Note that the circles in this figure have a good deal less overlap than those in Figure 1 but still offer the ability to provide some degree of accuracy. This level of accuracy can still provide significant returns on investment but not at the level associated with situations in which both technical and contingency dimensions are optimized.

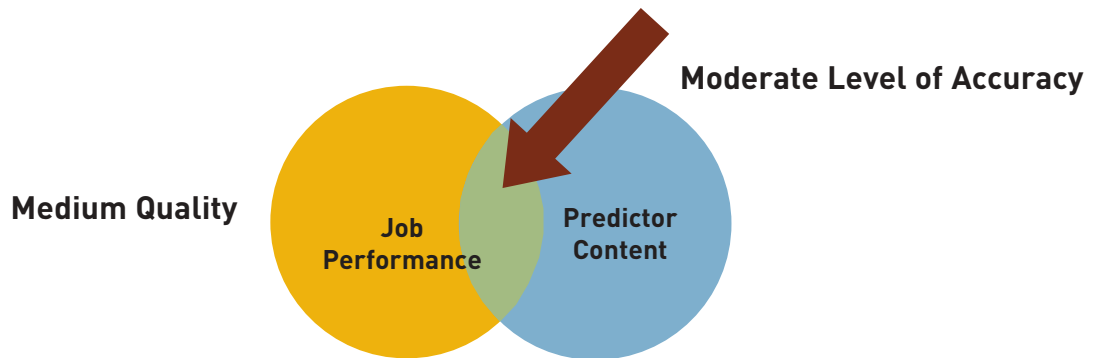


Figure 2. Medium Quality

Medium quality assessments fall in a range that includes the middle of the quality continuum. These assessments most often have been created in a manner that meet technical standards, although in some situations a failure to fully meet technical standards may be one reason for reduced levels of accuracy. The biggest determinant that creates a situation of medium quality is the contingency dimension. That is, the less the assessment tool is customized to meet specific needs of the job, the less accurate it will be. This is not necessarily a bad thing, as the value provided by quality depends on the needs of the organization.

The medium quality range typically includes "off the shelf" assessments or "one size fits all" assessments. That is, assessments that have been created for use in a wider range of situations with the goal being rapid deployment and lower cost to the consumer. Provided the organization using the assessment understands this, assessments of medium quality can offer value. These tools are especially useful to smaller organizations that do not have the resources required to invest in a higher quality solution.

Figure 3 represents the lowest point on the quality continuum. Note that circles in this figure show that there is virtually no relationship between the content of the assessment and the elements of the job for which it is being used. This creates a situation where accurate and systematic prediction is either not possible or far from optimal.

There are a variety of reasons why an assessment may be of low quality. However, these are generally related to one of two possible situations. Either the assessment itself has not been created using proper technical techniques, or the assessment that has been chosen has very little relevance for the situation in which it is being used. Regardless of the reason for the low level of accuracy, this situation is to be avoided because it will fail to provide users with the ability to systematically predict outcomes that have value to the organization.

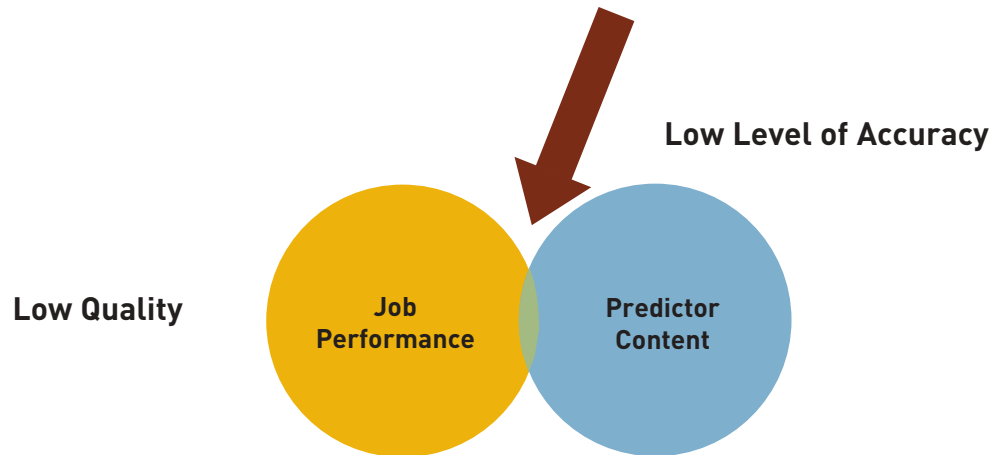


Figure 3. Low Quality

This situation is a bit different from that of high and medium quality because, while those require a choice related to the needs and resources of the organization, low quality assessments should be avoided at all costs.

The following section provides some direction to help organizations avoid selecting low quality solutions.

#### Section IV: Choosing a Quality Assessment Tool

This paper suggests that choosing an assessment tool that will deliver quality requires two important considerations. The first of these is determining what level of quality is desirable given the contingencies that surround the implementation. The second is conducting the due diligence required to avoid choosing a low quality solution.

##### A. Optimizing quality relative to needs

Ensuring the selection of a quality assessment tool requires a variety of careful considerations and an awareness both of the organization's specific needs and the resources available to meet those needs. This requires a careful review of the parameters that create the contingency dimension of quality and an understanding of the following key decision making factors:

**Your needs.** What is the primary driver for the use of the assessment tool? What do you hope to gain via its use?



**The substance.** What human dimensions define job performance and thus will help to ensure you are able to realize the outcomes you seek? Is the job a complex one or one that is relatively simple? How unique are these aspects and to what degree do unique aspects determine outcomes?

**Available resources.** How much time and money do you have to spend on this project? What is the comfort level in investment in this tool? Smaller organizations or those with fewer resources to invest may not be able to seek out the highest quality solutions, but still may be able to benefit from those of medium quality.

**Scope.** How broadly will the assessment tool be implemented? Will one tool be expected to carry the load for more than one job?

**Process.** How will the assessment tool be embedded into your process? This will be an important consideration as the ultimate ability to make accurate selection decisions is a result of the entire selection process rather than one specific assessment.

Answering the above questions will result in a decision regarding the strategy used when evaluating the suitability of an assessment. For instance, if you have a high volume job that is relatively unique, and have time to work through implementation, then choosing an assessment that is customized to your needs may make sense. On the other hand, it may make sense for smaller organizations with fewer resources to deploy an assessment that will be implemented off the shelf with no customization. While the quality of this solution may not be as high as that of a more customized solution, this type of tool can be deployed quickly and easily and can still offer value through systematic prediction of valued outcomes.

## **B. Avoiding low quality solutions: Due Diligence**

A second and critical step in optimizing the use of assessment tools involves ensuring the avoidance of low quality solutions. This requires two types of activities. The first is an audit of the technical aspects related to the assessment while the second involves a thorough understanding of the match between the content of the test and the parameters that define job performance.

### **1. Technical evaluation**

The technical evaluation of an assessment measure can be a relatively complex endeavor. At a minimum, any organization offering an assessment for use in pre-employment testing should be able to provide a technical manual to demonstrate the "pedigree" of the

assessment. Unfortunately, while the provision of a technical manual is a pre-requisite, these manuals should be scrutinized carefully because most technical manuals have been created to shine a positive light on the assessment in question. The goal of reviewing the technical manual should be to gain insight into the following:

- An understanding of the theory on which the assessment is based
- An understanding of the process used to construct the assessment
- Documentation of validation efforts demonstrating the effectiveness of the tool
- Evidence of positive outcomes related to the use of the assessment
- Evaluation of legal implications associated with using the assessment (e.g., adverse impact)

## **2. Relevance of the assessment to job performance**

The second aspect of due diligence involves gaining a thorough understanding of the match between the actual content of the assessment measure and the various human elements that are required for effective job performance.

This type of evaluation is best approached by ensuring that the idea of what defines job performance is documented before any evaluation is begun.

This will help to ensure a solid foundation with which the relevance of an assessment tool can be gauged. Once this information has been documented, the contents of various assessment tools can be evaluated relative to a set of desired performance dimensions.

This type of evaluation requires a review of manner in which job performance dimensions are identified by the assessment provider. The greater the investment made by the test provider in ensuring a close mapping of test content to desired outcomes, the higher the quality the assessment should be. For instance, highly customized assessments should be of higher quality because they have been developed or modified to match specific job performance requirements. On the other hand, assessments that are used off the shelf with less modification may have a good deal of overlap with desired performance outcomes but may fail to measure some critical dimensions and measure other dimensions that are not relevant. Thus the off the shelf approach is most likely to offer a medium level of quality. This is not necessarily a problem, but rather a consideration that must be taken into account when reviewing the details of any assessment tool.

The situation that must be avoided is one in which there is little or no relation between the content of the assessment and performance on the job in question. This situation will offer little or no quality and therefore no value through the use of the assessment. Therefore, when performing due diligence ensuring the strength of the relationship between test content and performance is a critical step in ensuring quality. Once some connection between the constructs measured by a test and those required for job performance has been established, determining an acceptable level for the strength of this connection will require an evaluation of the contingencies associated with the implementation of the assessment tool. Opting for a higher level of connection between test content and job performance has a direct relationship to tangible outcomes that can have value for the organization. These outcomes are discussed in more detail in the following section.

#### **IV. Conclusion**

There are two important reasons why organizations should take time to ensure the quality of any assessment measure they are considering: legal defensibility and return on investment. These things are actually directly related to one another via the fact that they both rely on the accuracy of the assessment in predicting job performance (i.e., assessment quality).

##### **Legal defensibility**

The greater the quality of an assessment the closer the link between that assessment and job performance. This link is the primary determinant of legality when it comes to the use of assessment tools. So, organizations who are concerned about legal issues related to the use of an assessment tool should work to ensure that the relationship between test and job performance is as strong as possible. It is important to understand that even assessments of medium quality should provide legal defensibility but that the level at which quality diminishes will have a direct impact on the legal risk associated with the use of the assessment tool. After a certain point, one in which there is little or no relationship between the test and job performance, legal defensibility will be compromised.

##### **Return on investment**

The same relationship which defines the legal defensibility of an assessment measure has a direct impact on its ability to provide return on the investment made in implementing it. The higher the quality of the assessment tool, the more direct the link between the constructs it measures and job performance requirements. This means that, in general, selecting a higher quality assessment tool should lead to larger returns because of the increased accuracy associated with the measure.



Just as with legal defensibility, there are judgment calls that must be made when determining the required level of quality relative to the ROI of an assessment measure. In the right circumstances an assessment of medium quality is capable of providing significant return on investment. This is why contingency factors become an important factor when choosing an assessment tool. For instance, using an off the shelf assessment that will provide only a medium level of return on investment but which can be deployed rapidly with little disruption and which can provide strong ROI may often be preferable to a more expensive and time consuming solution.

### **Verifying quality**

One important part of evaluating the return on investment provided by an assessment tool involves taking steps to verify its quality. There are a variety of ways to accomplish this but no matter what the method, the end goal involves an evaluation of the strength of the relationship between the content of an assessment and the outcomes it has been designed to predict. The stronger this relationship is, the higher the level of quality of the assessment tool (and the greater the return on the investment made in the tool). This is commonly done via a process known as test validation. There are a variety of ways to validate a test but the only way to obtain an actual numerical index of test quality is via a process in which is known as a criterion-related validation study. A full description of this process is beyond the scope of this paper, however the basic idea is that it provides a way to verify the accuracy of an assessment via statistical evaluation. The results of such studies are often an important part of the technical documentation provided by assessment vendors in order to demonstrate the effectiveness of their products.

### **Conclusion**

This is an important time in the evolution of hiring processes. Technology has changed many things about the way assessments are configured, delivered, and evaluated. Despite these changes, the concept of assessment quality remains grounded in the central idea that, to be effective, assessments must accurately predict key constructs required for effective job performance. While all tools must meet certain technical standards in order to be effective, understanding assessment quality is relative in that it requires consumers first understand their own needs and make decisions accordingly. These decisions will be unique to the organization interested in using the assessment tool and will ultimately define the degree of quality that is right for each specific situation. Thus, ensuring quality of assessment tools requires awareness on the part of the consumer and willingness to conduct the due diligence required to understand both their own needs and the degree to which an individual tool can meet them.



### About the Author

Charles Handler, Ph.D., is the president and founder of Rocket-Hire (<http://www.rocket-hire.com>), a consultancy dedicated to helping organizations use technology and best practices to build effective, legally sound employee selection systems. Dr. Handler is an internationally known thought leader in the development of online screening and assessment technology. Since starting Rocket-Hire, he has specialized in partnering with customers to develop strategies to help increase the efficiency and effectiveness of the employee selection process. In addition to Internet based hiring practices, Charles has also conducted extensive survey work and created training/development and performance management systems for a variety of companies. He has recently launched the online version of his one of a kind publication: The Buyer's Guide to Online Screening and Assessment (<http://buyersguide.rocket-hire.com>)

Dr. Handler's work in building new models for selection using technology has appeared in HR publications worldwide. He is a regular columnist for the Electronic Recruiting Exchange, Workforce Management Magazine, and SHRM's Employment Management Today Magazine. Dr. Handler has also played an active role in education, serving as a member of Tulane University's Business Studies faculty, where he currently teaches Recruitment and Selection, and other HR related topics. He is frequently asked to speak about best practices for employee selection at events such as the Society for Industrial/Organizational Psychology's annual conference, the Electronic Recruiting Exchange Annual Expo, and a many other user conferences and workshops.

Dr. Handler is a member of the American Psychology Association (APA), The Society for Human Resource Management (SHRM), and The Society for Industrial/Organizational Psychology (SIOP). He holds a Ph.D. in Industrial/Organizational Psychology from Louisiana State University and has been working as an employee selection specialist for over 10 years.