



Overview of Validity and Business Outcome Studies at PreVisor Inc.

PreVisor Optimization Services

Table of Contents

| | |
|--|-------|
| Overview of Criterion Validation and Business Outcome Studies | p. 3 |
| Validation Process Overview | p. 3 |
| Background on Validation | p. 3 |
| Approaches to Criterion Validation and Business Outcome Studies | p. 4 |
| Validation vs. Business Outcome Studies | p. 4 |
| Purposes of the Studies | p. 4 |
| Pros and Cons of the Studies | p. 5 |
| Validation and Business Outcome Study Best Practices..... | p. 6 |
| Appropriateness of Metrics | p. 6 |
| Representativeness of the Sample..... | p. 7 |
| Statistical Feasibility | p. 7 |
| Optimizing the Selection Solution..... | p. 9 |
| Conclusion..... | p. 9 |
| References | p. 10 |
| Appendix A: Major Steps and Typical Timeline of PreVisor Validation and Business Outcome Studies | p. 11 |
| Appendix B: Client Worksheet for Understanding Objective Performance Data Issues..... | p. 12 |
| Appendix C: Sample Size and Statistical Power | p. 13 |

Overview of Criterion Validation and Business Outcome Studies at PreVisor

The field of personnel selection testing is founded on the premise that psychological assessments can be effectively used to predict future job performance. Selection tests have been developed to measure a variety of individual characteristics such as sales potential, conscientiousness, cognitive ability, customer service potential, interpersonal effectiveness, supervisory and leadership ability, and turnover, which are important for successful performance in a variety of jobs. Validation refers to a scientific process of demonstrating that a selection test effectively predicts job performance.

Validation is a core business practice of PreVisor for three important reasons:

1. It confirms the best combination of assessment tests are being used for a given position.
2. It serves to underscore for our clients the specific organizational and financial benefits of using PreVisor assessments.
3. It provides our clients with the highest level of protection from potential legal challenges of their hiring procedures.

Criterion validation refers to a validation process that directly links performance on a selection test to performance on the job. (While there are other scientifically sound validation processes such as content validation and validity generalization, these will not be addressed in this current paper.)

A parallel concept to validation is a business outcome study. A business outcome study involves systematic and scientific research aimed at demonstrating the operational and/or financial benefits resulting from the implementation and use of a selection test to make employment decisions.

This white paper is designed to help PreVisor clients (and prospective clients) understand the concepts and importance of assessment validation and business outcome studies. It gives an overview of the validation process, highlights the differences between validation and business outcome studies, and outlines PreVisor's approach to conducting these studies.

Validation Process Overview

Background on Validation

The use of testing for selection purposes has a long history dating back to the ancient Chinese practice of using written examinations to screen candidates for civil service positions.¹ Yet, the modern practice of scientifically evaluating the validity of selection assessments did not begin until the early 1900s.² As selection research proliferated throughout the 20th century, several key developments influenced test validation. First, the Civil Rights Act of 1964 outlined the need for fair selection practices (which the amended Act addresses in Title VII) and established the Equal Employment Opportunity Commission (EEOC).

Second, the EEOC and other enforcement agencies issued the Uniform Guidelines on Employee Selection Procedures in 1978, which remain the defining legal guidelines in personnel selection to the present day.³ This document served, in conjunction with case law (based on major Title VII court decisions), to solidify the fundamental role of test validation in personnel selection from a legal perspective.

Third, the Society for Industrial and Organizational Psychology (SIOP) published the first edition of the Principles for the Validation and Use of Personnel Selection Procedures in 1975 in order to outline best professional practices in this area.⁴ Recently, SIOP published the fourth edition of the Principles (2003), which now serves as the definitive set of professional guidelines in this area.⁵

From a practical business perspective, the most important consideration in assessing the usefulness of a selection test is its ability to predict job performance. From a legal standpoint, the ultimate goal of any validation study is to demonstrate that the selection test is job related. Fortunately, these objectives are not mutually exclusive. As noted in the Principles:

“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” (p. 4).⁵

Thus, demonstrating that an assessment is an effective predictor of key job performance outcomes serves both practical and legal purposes.

Approaches to Criterion Validation and Business Outcome Studies

The two major approaches in the area of validation research include **concurrent** and **predictive** study designs. Both techniques involve matching selection test scores with job incumbents' performance data.

In a **concurrent** study, a selection test battery is administered to existing employees. This information is then linked with supervisory ratings of job performance and/or objective performance metrics. A major advantage of concurrent studies is that using existing employees is faster and allows the selection process to be validated before it is ever actually implemented in the hiring process. A disadvantage of the concurrent approach is that current employees taking the assessment solution may not give their best effort, as they are not actually applying for a job. Also the employees used in the study may not be representative of the applicant pool from which the organization is currently drawing.

In a **predictive** validity study applicants take the selection tests prior to being hired. After a sufficient number of employees have been on the job long enough for appropriate performance measures to be collected (usually 6 to 12 months after the selection solution was initially implemented), the relationship between the performance measures and selection test scores are determined. Although the predictive study approach requires a longer timeframe, it has the advantage that the people taking the assessments tend to be more motivated to do well because they are actually applying for a job. Thus it can provide a somewhat more realistic picture of the effectiveness of the selection tests when applied to real applicants. Also the sample taking the solution is more reflective of the current applicant pool.

Another advantage of the predictive approach is that it provides a more solid foundation in saying that the solution predicts performance because the solution scores are assessed before any performance metrics have been collected. A disadvantage of a predictive study is that the statistical analyses are delayed until sufficient data have been collected.

The choice of a concurrent versus predictive validation approach should be based on a careful consideration of the advantages and disadvantages of each method for an organization's specific situation and needs. The pros and cons of concurrent and predictive studies are summarized in Table 1.

Table 1. Pros and Cons of Concurrent and Predictive Validity Studies

| | Concurrent Study | Predictive Study |
|---|------------------|------------------|
| Pros | | |
| Provides results relatively quickly | ✓ | |
| Applicants motivated to do their best on the selection tests | | ✓ |
| Allows validation before selection tests are used with applicants | ✓ | |
| Sample representative of the actual applicant pool | | ✓ |
| Provides a more solid foundation on the predictive nature of the selection test | | ✓ |
| Cons | | |
| Employees may not be motivated to perform their best on the selection tests | ✓ | |
| Results are delayed until a sufficient number of employees have been hired and have worked for a period of time | | ✓ |
| Employees may not be demographically representative of the current applicant pool | ✓ | |

Validation vs. Business Outcome Studies

Purposes of the Studies

PreVisor makes a clear distinction between validation and business outcome studies. At PreVisor, business outcome studies are specifically intended to highlight the operational and/or financial benefits of the assessments that are being used (the assessment solution) for a business audience. These studies typically use a predictive validity design, where follow-up performance measures are collected for employees who were initially hired based on PreVisor's assessments. Business outcome studies usually focus on existing objective performance metrics that are routinely collected by an organization's internal performance management system. These metrics can include a variety of performance outcomes, such as productivity, attendance, turnover, average call time, sales or collections per hour, and customer

service ratings. PreVisor recommends that judgmental supervisor ratings also be collected as part of the study when feasible. Results of business outcome studies are typically presented in the form of a PowerPoint presentation.

Validity studies are intended to provide solid legal documentation of assessment solution validity as well as to demonstrate the economic worth of the assessments. Although business outcome studies would be useful if a selection procedure were ever legally challenged, validity studies are specifically intended for this purpose. As a result, compared to business outcome studies, validity studies entail more detailed documentation of the statistical analyses that are conducted (including validity coefficients, adverse impact analysis, and examination of the psychometric properties of the performance criteria) and the findings are presented in a more lengthy technical report.

Another key distinction of validity studies is that they can include either a predictive or a concurrent approach. Thus, they afford more flexibility in the study design. In addition, validity studies typically include the use of a PreVisor Job Performance Rating (JPR) form, whereby each employee's leader is asked to complete a survey regarding their employees' job performance.

Both criterion-related and business outcome studies often lead to a decision to modify the assessment solution or passing scores in order to enhance utility or the prediction of job performance. This modification process is referred to as optimizing the solution. Optimization may be called for by factors such as changes in job duties, changes in the applicant pool or applicant flow or the development of new tests related to the job. Changes to the solution can include adjustments to the weights given to individual assessment components in the calculation of an overall score, scoring previously unscored experimental assessment components, or streamlining the solution to include the most predictive assessments. Appendix A reviews the key steps involved and typical timeframe of validation and business outcome studies.

Pros and Cons of the Studies

Both business outcome and validity studies have unique advantages and disadvantages. The general pros and cons of conducting business outcome and validation studies are outlined in Table 2. A key advantage of PreVisor's business outcome studies is their ability to show the financial or operational impact of using selection assessments with little or no disruption to the on-line production unit. In contrast, validity studies, especially those involving a concurrent design based on existing employees, do necessitate that time be

allocated for job incumbents to complete the PreVisor assessment solution. Finally, validation studies are typically more costly and labor-intensive than business outcome studies due to the additional analyses and time associated with the technical report write-up needed to provide solid legal defensibility.

Table 2. Pros and Cons of Validity and Business Outcome Studies

| | Validity Study | Business Outcome Study |
|---|----------------|------------------------|
| Pros | | |
| Requires little or no disruption to the on-line production unit | | ✓ |
| Can be completed in a relatively short timeframe (once an appropriate sample size has been reached) | | ✓ |
| Provides a good "spot check" of the benefits and outcomes of assessment testing | | ✓ |
| Provides the highest level of legal defensibility and the best picture of assessment testing effectiveness | ✓ | |
| Allows for more flexibility in the choice of study designs (i.e., concurrent vs. predictive approaches) | ✓ | |
| Cons | | |
| Is not specifically intended to demonstrate legal defensibility | | ✓ |
| The use of objective business metrics requires a larger sample size (typically at least 300) | | ✓ |
| Typical reliance on a predictive study design limits flexibility and takes a more lengthy period of time to reach an appropriate sample size | | ✓ |
| The concurrent design approach necessitates some disruption to the on-line production unit for test administration to existing job incumbents | ✓ | |
| Is more labor intensive and costly than business outcome studies | ✓ | |

In sum, the choice of conducting a business outcome versus validation study should be made based on a

careful consideration of the organization's needs – including the importance of legal defensibility – and weighing the pros and cons associated with each approach for the client's specific situation.

Validation and Business Outcome Study Best Practices

Although, there are differences between validity and business outcome studies, many of the same best practices apply to both. PreVisor has extensive experience in conducting validity and business outcome studies dating back over 50 years. Based on this experience, as well as recommendations outlined in the SIOP Principles,⁵ PreVisor researchers have developed a set of best practices and guidelines in this area.

Before deciding to conduct a business outcome or validity study, three key issues need to be considered in order to determine the feasibility of the research:

1. The appropriateness of performance measures
2. The representativeness of the sample
3. Statistical issues.⁵

After a study has been conducted, other key factors need to be considered in making decisions on optimizing the assessment solution.

Appropriateness of Metrics

Performance metrics can include a wide variety of criteria, such as the quality or quantity of job performance, promotions, job tenure, customer service ratings, training performance, sales output, 360-degree feedback ratings, or any other performance outcome directly related to job success. Before a criterion-related validity study or a business outcome study can be conducted appropriate performance measures that are relevant to the job must be identified.

Complexity of Job Performance

When considering performance metrics, it is also important to note that for most jobs performance is complex and multi-dimensional in nature. For example, a company might be interested in maximizing sales, maximizing adherence to company policy, minimizing turnover and maintaining a workforce that is reflective of the diversity found in the community. At times these

desired outcomes may actually be inversely related to each other. For instance, individuals who generate more sales may also be the same ones more likely to leave the company for another job. This complexity makes it difficult for any one measure to capture the full picture of performance.

Therefore, as much as possible, performance measures should provide a broad assessment of performance, or multiple performance metrics should be included in the study. The combination of the performance metrics that are used should be as representative of the whole job as possible.

Types of Performance Measures

Two basic types of performance measures can be collected: objective and judgmental. Objective metrics can be thought of as direct measures of countable behaviors or outcomes which have financial or other important implications for the organization. These include measures such as such as products scanned per hour, attendance, passing a certification exam, turnover, average call time, and sales or collections per hour. Objective measures have the advantage of being more easily linked to financial outcomes than judgmental measures. However, objective metrics tend to be fairly narrowly focused and it may be difficult to capture the full range of a job with them.

Judgmental measures are based on one or more individual's evaluation of a person's job performance. Typically these are ratings by a supervisor or another individual having the opportunity to observe an employee's performance. Judgmental measures have the advantage that they can be designed to measure multiple dimensions of a job which may not have clear objective outputs (such as teamwork). Their disadvantage is that they can be more difficult to tie to financial benefits for the company.

PreVisor recommends using a combination of objective and judgmental measures to get the full picture of an individual's job performance.

To assist in the collection of judgmental performance measures, PreVisor can provide clients with a job performance rating form (JPR). The use of the PreVisor JPR brings several advantages to a either a validity or a business outcome study. First, the JPR instrument is carefully designed to provide a reliable and robust measure of various aspects of job performance that link back to PreVisor assessments. Second, supervisors are specifically given assurances that their performance ratings will remain confidential and will only be used for research purposes. In general, previous research

has demonstrated that job performance ratings made for research purposes are less lenient and show more differentiation between employees (and thus less range restriction) than performance appraisal ratings used for administrative purpose (e.g., annual salary increases, promotions)⁶. Finally, the JPR form is more resistant to extreme outlier scores than most objective performance metrics.

Expected Relationship to the Selection Assessments

Not all selection assessments will be predictive of all aspects of job performance. Therefore a critical consideration in selecting performance measures is the logical relationship that they have with the assessments. For example, it would be expected that an assessment measuring sales potential would be related to sales performance but would not necessarily be related to attendance or turnover. If measures are included that have no theoretical connection with the assessments that have been implemented, it is unlikely that a relationship will be found.

Under the Employee's Control

Another important characteristic is that performance measures must be under the employee's control. Assessments can only predict variation in job performance that is related to individual differences in personal characteristics (e.g., knowledge, skills, abilities, personality characteristics, conscientiousness, energy level, self-confidence). However, job performance measures can be affected by factors other than employees' individual characteristics. For example, if productivity is determined by the speed of a machine rather than by the skill or effort of the employee, then an assessment solution will not be able to predict this aspect of performance.

Or consider a sales call center position in the mortgage and home equity loan industry. Loan Officers might have a lending advantage or disadvantage based on the socioeconomic status and housing market of their assigned territory. Further, some employees might be assigned to work a call shift that is more or less favorable within the loan industry. Other external factors include: job location, season, economic trends, supervisor, local job market, and minor differences in job duties. When external factors exert an influence on employee performance, these factors need to be understood and controlled for in order to examine the value and impact of assessment solutions.

To get an accurate picture of how an assessment solution is distinguishing between low and high performers, it may be necessary to adjust for these

outside influences. Fortunately, many of these factors (technically known as "moderators") can be corrected for statistically if they are identified early in the study and the information is properly documented and coded in the performance data that is made available for analyses.

Appropriateness of Metrics Summary

In sum, the selection of appropriate performance measures is a key success factor in validity and business outcome studies. There are a number of important considerations in this area. Performance measures should be job relevant, provide a comprehensive measure of important aspects of job performance, and be expected to relate to the selection assessments. Also, influences outside employees' control should be carefully considered. Appendix B provides a client worksheet for better understanding data issues commonly encountered when working with objective performance metrics.

Representativeness of the Sample

A second major consideration when assessing the practicality of conducting a validity or business outcome study is the representativeness of the research sample. As outlined in the Principles, it is important that the research sample be "reasonably representative" of the job and applicant pool (p. 14).⁵ Although business outcome and validity studies rarely include all possible employees in a target position, they should sample broadly from across the entire job incumbent pool, including location, job tenure, and previous performance ratings as well as demographic variables such as age, gender, and race/ethnicity.

Statistical Feasibility

A third key issue to consider before conducting a business outcome or validity study involves examining the statistical feasibility of the research. There are several important considerations in this area, including range restriction, performance metric reliability, and sample size.⁵

Range Restriction

Range restriction occurs when the variability on either the assessment solution or the performance measures (or both) in the sample used in the study is less than in the population in general. For example, if a company only hires applicants who score in the top 10 percent of the PreVisor assessment solution, or if all incumbents receive highly similar job performance ratings, then the statistical ability to show differences between employees is greatly reduced.

To avoid range restriction employees must vary in their scores on the performance measures. For example, if on 100-point call quality measure 95% of employees score between 95 and 100, then it will be very difficult to show that assessment scores relate to job performance. This type of restriction is regularly encountered in business outcome studies as employees scoring lower than 95 are often terminated. Another example of range restriction is if 80% of employees receive a rating of three on a five-point performance scale. In cases like this, there is not much variability in performance for an assessment solution to predict.

When assessment solution scores are available on a larger unrestricted applicant sample (such as a national sample of applicants or all applicants, hired or not hired, for an organization) statistical techniques are available to adjust for the restricted variance. These adjustments give a more accurate estimate of the relationship between assessment scores and the performance measures. They can also be used to project what the performance for low scorers on the solution would have been had they been hired. This helps the organization see the benefit of not hiring applicants who score low on the assessments.

Metric Reliability

Metric reliability (consistency) is also a key consideration. Generally speaking, high performers are expected to perform well from month to month. Thus, you would not expect someone to be the best performer one month, the worst in the next month, and an average performer in the third month. When the rank order of employees on a performance metric varies wildly from one month to the next it is probable that factors other than the personality, skills, and attitudes of the employee have a large influence on performance.

These inconsistencies (measurement error) in the performance metrics reduce the potential to statistically demonstrate that assessment scores are predictive of job performance. To be able to assess the reliability of performance metrics, it is best to have multiple months of performance data (at least three but ideally five or more) for each employee. The multiple points of measurement help ensure that the performance metrics are a consistent gauge of performance. If performance data are only collected at one point in time it is impossible to determine if the values represent an employees' typical performance.

For example, for any particular time period an employee may perform below his or her usual level due to illness, changing work conditions or personal issues. On the other hand, he or she may have performed better than

usual due to chance factors such as happening to be the one to receive a call which results in an exceptionally large order.

To help avoid measurement error, it is essential that performance metrics be collected and calculated consistently across all employees in the sample. For example, if sales revenue is calculated one way for half of the employees in the sample and another way for the other half, then statistical analysis based on this performance metric could be misleading and inaccurate.

Fortunately, statistical techniques are available to help correct for the effects of range restriction and measurement error, thus showing a more accurate picture of the true relationship between assessment scores and performance on the job. However, these techniques are only useful in mild to moderate cases and cannot overcome the impact of severe range restriction and measurement error.

Sample Size

Ensuring that there is an adequate sample size available for the study is another key statistical issue.

It is important to note that in order to be included in the research sample job incumbents must meet the following three criteria:

1. They have completed the PreVisor assessment solution
2. They have been on the job long enough to have 3 months of performance data after the initial training and ramp-up period
3. They have performance measures and/or supervisory job performance ratings that can be matched to their PreVisor assessment scores.

In general, results obtained from small samples are more unstable and inconsistent because the results of a few employees with extreme scores on the assessments or performance measures can have disproportionate effects on the statistical analyses. Extreme scores can serve to either artificially overestimate or underestimate statistical results. With larger sample sizes, the influence of these extreme scores is greatly reduced. A related concept is called statistical power, which is the ability to detect evidence of predictor-criterion relationships in a sample when a true relationship between the variables

exists. Sample size is an important component in statistical power, with larger sample sizes having a greater chance of obtaining significant results (see Appendix C for a more detailed technical explanation of statistical power).

Specifically, PreVisor recommends a minimum matched sample size of 200 incumbents for validity studies where JPR's are used and 300 for business outcome studies where objective performance metrics are used. Based on our extensive research experience, PreVisor has found these sample sizes overcome most of the previously noted statistical issues, and provide an acceptable level of confidence that validity or business outcome studies will demonstrate both reliable and accurate results. If an organization wants to compare the assessment solution across locations, shifts or other factors, data from more than 300 employees may be needed. In sum, larger samples provide more stable and consistent results because they are less impacted by extreme scores and provide more statistical power to detect true relationships.

Optimizing the Selection Solution

A key benefit of validity and business outcome studies is that they highlight opportunities to enhance prediction of job performance. Optimization opportunities include adjusting to the weights given to individual assessment components in the calculation of an overall score, adding new components, scoring previously unscored experimental assessment components, adjusting the passing score or streamlining the solution to include the most predictive assessments. When deciding which optimization options to pursue, several issues should be considered.

First, it is important to evaluate whether the conditions of the study warrant making changes based on the study's results. If conditions resulted in one or more of the above best practices being violated, then it may be best to delay making changes until the violation can be corrected (larger sample obtained, more reliable metrics collected, etc.).

If conditions merit considering changes to the selection solution, organizational leadership needs to consider the complex and multidimensional nature of performance discussed earlier. Enhancing the prediction of one aspect of performance may lead to lower prediction on another aspect. Particular caution must be exercised when the performance metrics used in the analyses do not completely capture all the important outcomes of the job. If the assessment solution has been based on a thorough job analyses, eliminating a solution

component that does not predict one of the observed metrics may result in selecting employees not as proficient on one of the important, but unmeasured, aspects of the job. Also broader factors such as organizational values, diversity, organizational culture and legal issues must be considered in decisions about changes to the assessment solution. During this process, PreVisor recommends optimization options to our clients and assists them in deciding on a solution that reinforces the balance the client wants to maintain on these important considerations.

Conclusion

In sum, PreVisor has tested over 35 million applicants, conducted hundreds of validation and business outcome studies on our assessment tools, and continues to seek opportunities to demonstrate the effectiveness of our solutions through validity and business outcome studies. As has been outlined here, many legal, practical, and technical issues must be carefully considered before proceeding with a validation or business outcome study. As noted by the authors of the SIOP Principles: "While validation efforts with a narrow focus may have value, poorly executed validation efforts may lead the employer to reject beneficial selection procedures or accept invalid ones" (p. 10). Careful coordination, planning, study design, and application of the best practices described here ensure that PreVisor business outcome and validation studies yield meaningful and accurate conclusions.

For more information about our validity and business outcome study process, please contact your account representative. To learn more about how PreVisor can help your organization, please contact sales@previsor.com or visit us at www.previsor.com.

References

- ¹Bowman, M.L. (1989). Testing individual differences in ancient China. *American Psychologist*, 44, 576-578.
- ²Schmidt, F.L., & Hunter, J.E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124, 262-274.
- ³Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor, and Department of Justice (1978). Uniform guidelines on employee selection procedures. *Federal Register*, 43 (166), 38295-38309.
- ⁴Division of Industrial-Organizational Psychology of American Psychological Association (1975). *Principles for the validation and use of personnel selection procedures*. Oxford, England: Author.
- ⁵Society for Industrial and Organizational Psychology (2003). *Principles for the validation and use of personnel selection procedures* (4th ed.). Bowling Green, OH: Author.
- ⁶Jawahar, I.M., & Williams, C.R. (1997). Where all the children are above average: The performance appraisal purpose effect. *Personnel Psychology*, 50, 905-926.

Appendix A

Major Steps and Typical Timeline of PreVisor Validation and Business Outcome Studies

| Step | Description | Party Responsible | Timeline |
|-----------------------|---|---------------------|---------------------|
| (1) Plan Study | <ul style="list-style-type: none"> • Kickoff Meeting • Discuss study type, steps, timelines, client requirements • Identify appropriate performance metrics, review sample metrics • Decision: Is the study generally feasible? • Present Analysis Plan and Timeline | PreVisor and Client | 2 - 3 weeks |
| (2) Collect Data | <ul style="list-style-type: none"> • Client collects internal performance metrics • Selection tests administered for concurrent studies • JPR's administered if used | Client | 2 - 3 weeks |
| (3) Analyze | <ul style="list-style-type: none"> • Performance data matched with selection test data • Decision: Is sufficient data available to make study feasible? • Conduct Analyses | PreVisor | 2 - 4 weeks |
| (4) Report and Review | <ul style="list-style-type: none"> • Present results to client • Conduct any needed follow up analyses • Agree on solution optimization options | PreVisor and Client | 1 - 2 weeks |
| (5) Optimize | <ul style="list-style-type: none"> • Optimize assessment solution • Deploy optimized assessment solution | PreVisor and Client | 1 - 3 weeks |
| | | Total | 8 - 15 weeks |

Appendix B

Client Worksheet for Understanding Objective Performance Data Issues

| | Metric/Data Issue | Comments | Suggestions for Metric/Data Improvement |
|-----|--|----------|--|
| (1) | Was the performance data gathered using a reliable and accurate process? What concerns do you have that the data is not of a high quality? | | <ul style="list-style-type: none"> • Seek ways to increase the reliability and accuracy of metrics already collected. • Work with internal job/data experts to identify ways to increase performance data quality moving forward. |
| (2) | Is there a system in place for ensuring the quality of the data? If yes, please describe this system. | | <ul style="list-style-type: none"> • Identify internal job/data experts who can review the existing data for quality issues. • Seek ways to improve data quality assurance moving forward. |
| (3) | Is any data missing? If yes, is the missing data represented as a blank value (preferred) or as zero (0) values (undesirable)? | | <ul style="list-style-type: none"> • Identify internal job/data experts who can review trends in missing data. • Recode zeros to blank values as appropriate. |
| (4) | Are there any extraneous factors (outside of the employees' control) that could serve to influence the performance metrics, such as: job location, season, supervisor, local job market, etc.? If yes, is it possible to code the data to identify these factors (such as providing employee job location)? | | <ul style="list-style-type: none"> • Talk to internal job experts who would best understand what extraneous factors influence employee performance. • Look for ways to code extraneous factors in the data set (e.g., including job location, month data was collected, etc.) |
| (5) | Has the data been transferred from one file format to another (e.g., from Microsoft Access to Excel)? If so, is there any possibility that the transfer process might have affected the quality of the data? | | <ul style="list-style-type: none"> • Identify an internal data expert who can review both the original and transferred files to insure 100% data integrity. |
| (6) | Are the minimum and maximum values for each performance variable accurate and plausible? In other words, do these numbers fall within a range of what would normally be expected for each variable? | | <ul style="list-style-type: none"> • Work with job/data experts to identify out-of-bound values (e.g., negative attendance, extreme values). • Work to understand where these values are coming from and how they should be addressed in the data file. |
| (7) | Is the average performance score "typical" for each variable? If no, what concerns you about the average performance scores? | | <ul style="list-style-type: none"> • Identify internal job/data experts to examine average performance scores and if the data makes sense. • If there are any surprises, work to understand what is going on with the data. |
| (8) | Are the values for each performance metric normally distributed throughout the employees included in this sample or are performance scores bunched-up at points? For example, on a 1-5 point performance scale, are there an equal number of 1's, 2's, 3's, 4's, and 5's, or do 90% of employees receive a 4 rating? | | <ul style="list-style-type: none"> • Work with job/data experts to understand the distribution of performance scores. • If scores are found to be bunched-up (e.g., everyone receives the same performance rating), seek ways to better differentiate employee performance moving forward. |

Appendix C

Sample Size and Statistical Power

The term statistical power in validation research refers to the probability of obtaining a statistically significant relationship in sample data when the true relationship in the population is of a certain magnitude. To evaluate the power of a given study design, four parameters need to be specified:

1. An estimate of the true effect size (e.g., a correlation between the predictor and criterion) in the population of job applicants,
2. A proposed sample size (or a range of possible sample sizes),
3. An alpha level (typically .05 by common convention),
4. A choice of a one-tailed or two-tailed significance test.

There are several methods available for estimating power, including published tables, statistical formulas, and computer software packages. The table below provides an overview of the most common situations faced in validation and business outcome study research based on a two-tailed test (which is the most common approach). As can be seen in this table, there are dramatic differences in the percent chance of obtaining significant correlations based on sample size, with larger samples having a much better likelihood of obtaining significance. For example, assuming an estimated population correlation of .20, the percent chance of obtaining significant correlations for a sample of 100, 200, and 300 job incumbents is 51%, 81%, and 94%, respectively. In sum, sample size is an important issue and has a direct relationship with the probability of demonstrating significant results in validation and business outcome studies.

**Percent Chance of Obtaining Significant Correlation
when Alpha = .05, Two-tailed Test**

| Sample Size | Estimated Population Correlation | | | | | | | | |
|-------------|----------------------------------|-----|-----|-----|-----|-----|-----|-----|-----|
| | .05 | .10 | .15 | .20 | .25 | .30 | .35 | .40 | .45 |
| 25 | 04 | 06 | 10 | 15 | 22 | 30 | 40 | 51 | 62 |
| 50 | 05 | 10 | 18 | 28 | 42 | 56 | 71 | 83 | 92 |
| 75 | 06 | 13 | 25 | 40 | 58 | 75 | 87 | 95 | 99 |
| 100 | 07 | 16 | 32 | 51 | 71 | 86 | 95 | 99 | |
| 125 | 08 | 20 | 39 | 61 | 81 | 93 | 98 | | |
| 150 | 09 | 23 | 45 | 69 | 87 | 96 | 99 | | |
| 175 | 10 | 26 | 51 | 76 | 92 | 98 | | | |
| 200 | 10 | 29 | 56 | 81 | 95 | 99 | | | |
| 300 | 14 | 41 | 74 | 94 | 99 | | | | |
| 500 | 20 | 61 | 92 | | | | | | |
| 1000 | 35 | 89 | | | | | | | |

Note: Validity correlations (or coefficients) typically range from .15 to .30.
Power is $\geq 99.5\%$ for all empty cells.