plumeus inc.

Compliance with APA's *Standards for Psychological and Educational Testing*

Table of Contents

Copyright © 2006 Plumeus Inc.

1-888-855-6975

Copyright © 2006 Plumeus Inc.

1-888-855-6975

INTRODUCTION

The intent of this document is to demonstrate how Plumeus Inc attempts to promote sound and ethical use of its tests and to provide descriptions on its methodology in maintaining the highest standards in the development and rationale of these tests.
To this extent, it was decided to use as a basis for guidelines, the document *"Standards for educational and psychological testing, 2004"*, and to provide in this document brief descriptions on how Plumeus attempts to adhere to the *Standards* when developing and using its questionnaires/tests.

Copyright © 2006 Plumeus Inc.

Test Construction, Evaluation, and Documentation

1. Validity

1.1
A rationale should be presented for each recommended interpretation and use of test scores, together with a comprehensive summary of the evidence and theory bearing on the intended use or interpretation.
yes
Each test is developed using up-to-date theories on the subject being assessed. Important references are presented in the test results, and are available in the technical manual as well. If test developer(s) include empirically selected scales in a test, the reasoning is clearly stated in the test interpretations and supporting materials, the scales must possess content validity, and be validated upon gathering of data.

1.2
The test developer should set forth clearly how test scores are intended to be interpreted and used. The population(s) for which a test is appropriate should be clearly delimited, and the construct that the test is intended to assess should be clearly described.
yes
Tests to be used in a high-stakes manner (pre-employment testing, for instance) are clearly labeled in ARCH Profile so employers can know whether they are appropriate for such purposes. Tests labeled with 'EAP' (for Employee Assistance Programs) should only be used for therapeutic purposes. The description of the test covers what its purpose is as well. The interpretation of all of Plumeus' tests is generated automatically by computer, eliminating issues related to misunderstanding of the test scores. However, if the test developer still feels that the potential for misunderstanding of scores exists, he or she clarifies this issue in the results. All tests are validated with a general population (visitors to the sites queendom.com and psychtests.com), but even so, there are tests that are inappropriate for certain users. If this is the case, it is stated clearly in the tests' supporting materials.

1.3
If validity for some common or likely interpretation has not been investigated, or if the interpretation is inconsistent with available evidence, that fact should be made clear and potential users should be cautioned about making unsupported interpretations.
Yes
When experience or intuition tells a test developer that the test is likely to be construed in a manner counter to how it was intended to be interpreted, this concern is stated in the supporting materials, such as the test access file, the pop-up containing information about the test, and in the test manual. However, it is incumbent on the test user to evaluate the materials available about the test and to decide if the test can be generalized for their own purposes.

1.4

Copyright © 2006 Plumeus Inc.

1-888-855-6975

If a test is used in a way that has not been validated, it is incumbent on the user to justify the new use, collecting new evidence if necessary.
No
This is the responsibility of the user; however, Plumeus staff is available to assist clients in validating different uses.

1.5
The composition of any sample of examinees from which validity evidence is obtained should be described in as much detail as is practical, including major relevant sociodemographic and developmental characteristics.
Yes
This information is included in the statistical manuals available for Plumeus' tests.

1.6
When the validation rests in part on the appropriateness of test content, the procedures followed in specifying and generating test content should be described and justified in reference to the construct of the test is intended to measure or the domain it is intended to represent. If the definition of the content sampled incorporates criteria such as importance, frequency, or criticality, these criteria should also be clearly explained and justified.
Yes
If a given assessment is deemed incomplete, and does not cover the full range of the given construct, this fact is clearly spelled out in the test results, both for the test-taker and the test user.

1.7
When a validation rests in part on the opinion or decisions of expert judges, observers, or raters, procedures for selecting such experts and for eliciting judgments or ratings should be fully described. The qualifications, and experience, of the judges should be presented. The description of procedures should include any training and instructions provided, should indicate whether participants reached their decisions independently, and should report the level of agreement reached. If participants interacted with one another or exchanged information, the procedures through which they may have influenced one another should be set forth.
No
We do not use expert judges to establish the validity of the tests, although experts often do offer feedback to us. This informal feedback is taken into account for revisions of tests.

1.8
If the rationale for a test use or score interpretation depends on premises about the psychological processes or cognitive operations used by examinees, then the theoretical or empirical evidence in support of those premises should be provided. When statements about the processes employed by observers or scorers are part of the argument for validity, similar information should be provided.
Yes
When assumptions such as these are made, the premise is tested during the validation process. In addition, the rationale for inclusion of such elements is included in the test.

Copyright © 2006 Plumeus Inc.

1-888-855-6975

We do not make statements about observers or scorers as part of the argument for a test's validity, therefore the last point is not relevant to Plumeus.

1.9
If a test is claimed to be essentially unaffected by practice and coaching, then the sensitivity of test performance to change with these forms of instruction should be documented.
No
Such claims are not made by Plumeus.

1.10
When interpretation of performance on specific items, or small subsets of items, is suggested, the rationale and relevant evidence in support of such interpretation should be provided. When interpretation of individual item responses is likely but is not recommended by the developer, the user should be warned against making such interpretations.
No
The results of Plumeus' tests are generated automatically by computer, therefore the issue concerning whether users make conclusions based on one item is not relevant. However, when subscales are included in the test, the rationale behind them is included in the results. At times, small subsets of items are used to generate specific pieces of information in the results. When this is the case, the conclusions made are only those that follow reasonably from the answers provided by the test taker, as judged by experts.

1.11
If the rationale for a test use or interpretation depends on premises about the relationships among parts of the test, evidence concerning the internal structure of the test should be provided.
Yes
Correlations, ANOVAs, factor analyses, and other statistical procedures are performed to support the structure of the test. Results are reported in statistical manuals.

1.12
When interpretation of subscores, score differences, or profiles is suggested, the rationale and relevant evidence in support of such interpretation should be provided. Where composite scores are developed, the basis and rationale for arriving at the composites should be given.
No
The automatic scoring of Plumeus' tests makes this point irrelevant.

1.13
When validity evidence includes statistical analyses of test results, either alone or together with data on other variables, the conditions under which the data were collected should be described in enough detail that users can judge the relevance of the statistical findings to local conditions. Attention should be drawn to any features of a validation data collection that are likely to differ from typical operational testing conditions and that could plausibly influence test performance.
Yes

Copyright © 2006 Plumeus Inc.

1-888-855-6975

The population is clearly defined in the statistical manuals available for tests. Analyses of specific groups within this general population is possible, and any major differences between groups is both investigated (for problems in test-construction and to ensure that these are real differences) and reported. Data for validation studies is collected in a way similar to suggested use.

1.14
When validity evidence includes empirical analyses of test responses together with data on other variables, the rationale for selecting the additional variables should be provided. Where appropriate and feasible, evidence concerning the constructs represented by other variables, as well as their technical properties, should be presented or cited. Attention should be drawn to any likely sources of dependence (or lack of independence) among variables other than dependencies among the construct(s) they represent.
Yes
Many, if not all, of Plumeus' validity studies involve pairing test responses with data on other variables. If the reasons behind the choice of pairing is not obvious, the logic for including the particular type of data will be included in the statistical report for that test. Possible reasons for pairing test responses with data that is not a clear assessment of construct validity include confirming relationship between the given construct and another element that has been established in the research on the topic.

1.15
When it is asserted that a certain level of test performance predicts adequate or inadequate criterion performance, information about the levels of criterion performance associated with given levels of test scores should be provided.
Yes
This is provided in the statistical manuals; specifically in the criterion validity section, containing ANOVAs and t-tests. Our major predictive validity study (in progress) will provide more detailed performance-related information and will also contain multiple regression results.

1.16
When validation relies on evidence that test scores are related to one or more criterion variables, information about the suitability and technical quality of the criteria should be reported.
Yes
This information is not provided, however, the validation questions selected are intended to assess both self-reported and objective measures of the given construct. Again, this is related to our predictive validity study, the criteria being the performance evaluation and other objective variables provided by the manager.

1.17
If test scores are used in conjunction with other quantifiable variables to predict some outcome or criterion, regression (or equivalent) analyses should include those additional relevant variables along with the test scores.
No

Copyright © 2006 Plumeus Inc.

1-888-855-6975

This is the responsibility of users, however, if advice on how to develop a system to predict an outcome (such as employee productivity) is needed, representatives are available to aid in validating the overall program.

1.18
When statistical adjustments, such as those for restriction of range or attenuation, are made, both adjusted and unadjusted coefficients, as well as the specific procedure used, and all statistics used in the adjustment, should be reported.
Yes
When we do adjust scores for one reason or another, we report it in the statistical manuals of that particular test. For instance, for the Classical IQ Test, we have adjusted the scores for age and gender.

1.19
If a test is recommended for use in assigning persons to alternative treatments or is likely to be used, and if outcomes from those treatments can reasonably be compared on a common criterion, then, whenever feasible, supporting evidence of differential outcomes should be provided.
No
Such recommendations are not made by Plumeus.

1.20
When a meta-analysis is used as evidence of the strength of a test-criterion relationship, the test and the criterion variables in the local situation should be comparable with those in the studies summarized. If the relevant research includes credible evidence that any other features of the testing application may influence the strength of the test-criterion relationship, the correspondence between those features in the local situation and in the meta-analysis should be reported. Any significant disparities that might limit the applicability of the meta-analytic findings to the local situation should be noted explicitly.
yes
Presently, this statement does not apply. In the future, if we do use meta-analyses to validate the relationship between a test and a criterion, we will use studies that utilize similar conditions as the local situation.

1.21
Any meta-analytic evidence used to support an intended test use should be clearly described, including methodological choices in identifying and coding studies, correcting for artifacts, and examining potential moderator variables. Assumptions made in correcting for artifacts such as criterion unreliability and range restriction should be presented, and the consequences of these assumptions made clear.
Yes
Presently, this statement does not apply. In the future, if we do use meta-analyses to support the use of a test, the evidence supporting methodological choices will be presented in the supporting materials.

1.22

Copyright © 2006 Plumeus Inc.

1-888-855-6975

When it is clearly stated or implied that a recommended test use will result in a specific outcome, the basis for expecting that outcome should be presented, together with relevant evidence.
Yes
We leave it up to the users of the test to determine the outcome of testing. For instance, if a user does pre-employment testing with an assessment, it is up to them to determine whether or not they hire the individual. However, when claiming that the tests predict performance, we will provide supporting evidence from the predictive validity study.

1.23
When a test use or score interpretation is recommended on the grounds that testing or the testing program per se will result in some indirect benefit in addition to the utility of information from the test scores themselves, the rationale for anticipating the indirect benefit should be made explicit. Logical or theoretical arguments and empirical evidence for the indirect benefit should be provided. Due weight should be given to any contradictory findings suggesting important indirect outcomes other than those predicted.
No
We do not generally claim that there are indirect benefits of using our tests. Instead the benefits are gained directly from the results of the test, and the advice and tips which are based on the literature and clinical experience.

1.24
When unintended consequences result from test use, an attempt should be made to investigate whether such consequences arise from the test's sensitivity to characteristics other than those it is intended to assess or to the test's failure to represent the intended construct.
Yes
If unintended consequences do result from test use, Plumeus makes every effort to explain the unintended consequences and rectify the situation by altering the tests or making recommendations to users about how to deal with these consequences.

Copyright © 2006 Plumeus Inc.

1-888-855-6975

2. Reliability and Errors of Measurement

2.1
For each total score, subscore, or combination of scores that is to be interpreted, estimates of relevant reliabilities and standard errors of measurement or test information functions should be reported.
Yes
We perform reliability analyses on the overall score, if applicable, and all subscores. The analyses performed are Cronbach's Coefficient Alpha, and the following split half tests: Correlation between forms, Spearman-Brown, and Guttman's formula. The Standard Error of Measurement scores will soon be added to all statistical reports.

2.2
The standard error of measurement, both overall and conditional (if relevant), should be reported both in raw score or original scale units and in units of each derived score recommended for use in test interpretation.
Yes
The SEM is (or will soon be) included in both the test manual and the statistical report.

2.3
When test interpretation emphasizes differences between two observed scores of an individual or two averages of a group, reliability data, including standard errors, should be provided for such differences.
Yes
Some of our tests do compare scores on subscales in order to make inferences about the test-taker. Regardless of whether or not this is true of a particular test, we always present reliability data in the statistical reports and technical manual.

2.4
Each method of quantifying the precision or consistency of scores should be described clearly and expressed in terms of statistics appropriate to the method. The sampling procedures used to select examinees for reliability analyses and descriptive statistics on these sample should be reported.
Yes
All of this information is provided in the statistical manuals available for each test. In addition, technical manuals will also soon be available with this information.

2.5
A reliability coefficient or standard error of measurement based on one approach should not be interpreted as interchangeable with another derived by a different technique unless their implicit definitions of measurement error are equivalent.
Yes
In general, this is not a procedure that is done at Plumeus.

2.6
If reliability coefficients are adjusted for restriction of range or variability, the adjustment procedure and both the adjusted and unadjusted coefficients should be reported. The

Copyright© 2006 Plumeus Inc.

1-888-855-6975

standard deviations of the group actually tested and of the target population, as well as the rationale for the adjustment, should be presented.
No
This doesn't apply as we don't adjust reliability coefficients for restriction of range or variability.

2.7
When subsets of items within a test are dictated by the test specifications and can be presumed to measure partially independent traits or abilities, reliability estimation procedures should recognize the multifactor character of the instrument.
Yes
Plumeus performs reliability studies on both the main score (if one exists) and all subscales of the tests. If any element on the tests lacks reliability, the scale or entire test is revised in order to improve the assessment.

2.8
Test users should be informed about the degree to which rate of work may affect examinee performance.
No
At this time, the tests available through Plumeus are not subject to time restrictions. Although time is reported in ARCH Profile tests, it is up to managers to decide what they wish to do with that information.

2.9
When a test is designed to reflect rate of work, reliability should be estimated by the alternate-form or test-retest approach, using separately timed administrations.
No
Tests are not designed to reflect rate of work.

2.10
When subjective judgment enters into test scoring, evidence should be provided on both inter-rater consistency in scoring and within-examinee consistency over repeated measurements. A clear distinction should be made among reliability data based on (a) independent panels of raters scoring the same performances or products, (b) a single panel scoring successive performance or new products, and (c) independent panels scoring successive performances or new products.
No
All tests are scored by computer, eliminating inter-rater consistency as an issue altogether.

2.11
If there are generally accepted theoretical or empirical reasons for expecting that reliability coefficients, standard errors of measurement, or test information functions will differ substantially for various subpopulations, publishers should provide reliability data as soon as feasible for each major population for which the test is recommended.
Yes/No
If such a finding led a test developer to expect such differences between subgroups, he or she would investigate and publish findings supporting or disproving this information in the statistical manual of the test involved.

Copyright© 2006 Plumeus Inc.

1-888-855-6975

2.12
If a test is proposed for use in several grades or over a range of chronological age groups and if separate norms are provided for each grade group, reliability data should be provided for each age or grade population, not solely for all grades or ages combined.
Yes
While this is rare, we do occasionally use different norms and scoring techniques for different age groups or genders. Is this is the case, we present the reliability information for each group.

2.13
If local scorers are employed to apply general scoring rules and principles specified by the test developer, local reliability data should be gathered and reported by local authorities when adequate size sample are available.
No
All scoring and data-gathering is done online, therefore it is not necessary to gather local norms.

2.14
Conditional standard errors of measurement should be reported at several score levels if constancy cannot be assumed. Where cut scores are specified for selection or classification, the standard errors of measurement should be reported in the vicinity of each cut score.
No/Yes
Presently not done, but will be done in the future

2.15
When a test or combination of measures is used to make categorical decisions, estimates should be provided of the percentage of examinees who would be classified in the same way on two applications of the procedure, using the same form or alternate forms of the instrument.
Yes
With rare exceptions, we don't generally recommend that users make categorical decisions based on the results of our tests. However, when we do, test-retest reliability or the reliability of alternate forms will be reported if this information is available.

2.16
In some testing situations, the items vary from examinee to examinee - through random selection from an extensive item pool or application of algorithms based on the examinee's level of performance on previous items or preferences with respect to item difficulty. In this type of testing, the preferred approach to reliability estimation is one based on successive administrations of the test under conditions similar to those prevailing in operational use.
Yes
See the documentation on dynamic test administration at the end of this document.

Copyright© 2006 Plumeus Inc.

1-888-855-6975

2.17
When a test is available in both long and short versions, reliability data should be reported for scores on each version, preferably based on an independent administration of each.
Yes
Both full and abridged versions pf Plumeus' tests face full statistical scrutiny, including reliability analyses. In most cases, this is performed on an independent administration of each, rather than simply calculating a score for the shorter version from items in the longer test.

2.18
When significant variations are permitted in test administration procedures, separate reliability analyses should be provided for scores produced under each major variation if adequate sample sizes are available.
No
We do not recommend allowing significant differences in the way tests are administrated, especially for high-stakes purposes. All tests are administered online, with the exception of for visually impaired clients. Refer to section XXXX on testing subjects with disabilities.
Make the doc

2.19
When average test scores for group are used in program evaluations, the groups tested should generally be regarded as a sample from a larger population, even if all examinees available at the time of measurement are tested. In such cases the standard error of the group mean should be reported, as it reflects variability due to sampling of examinees as well as variability due to measurement error.
Yes
We are planning to include this in the next version of ARCH Profile – the group mean, standard deviation, and standard error of measurement will be included in the results. However, the client is ultimately responsible for statistical analyses in this situation.

2.20
When the purpose of testing is to measure the performance of groups rather than individuals, a procedure frequently used is to assign a small subset of items to each of many sub samples of examinees. Data are aggregated across sub samples and item subsets to obtain a measure of group performance. When such procedures are used for program evaluation or population descriptions, reliability analyses must take the sampling scheme into account.
No
This section does not apply.

Copyright© 2006 Plumeus Inc.

1-888-855-6975

3. Test Development and Revision

3.1
Tests and testing programs should be developed on a sound scientific basis. Test developers and publishers should compile and document adequate evidence bearing on test development.
Yes
Plumeus is dedicated to producing psychometric tests of the highest quality and scientific standards. We use the most up-to-date research in the development of our tests, and document the process along the way. Periodically, the tests are revised so that current research can be incorporated. Interpretations include any highly relevant references and clearly explain the structure of the test. Documentation is available with references, statistical information, and general procedures for the development of a test.

3.2
The purpose(s) of the test, definition of the domain, and the test specifications should be stated clearly so that judgments can be made about the appropriateness of the defined domain for the stated purpose(s) of the test and about the relations of items to the dimensions of the domain they are intended to represent.
Yes
The intended audience, information about the appropriateness of the test, and definition of the construct being tested are available in supporting materials before the test is administered to users.

3.3
The test specifications should be documented, along with their rationale and the process by which they were developed. The test specifications should define the content of the test, the proposed number of items, the item formats, the desired psychometric properties of the items, and the item and section arrangement. They should also specify the amount of time for testing, directions to the test takers, procedures to be used for test administration and scoring, and other relevant information.
Yes
Test specifications are documented in the access file and in the info box of each test. Since the tests are scored automatically, the procedures for administration and scoring are omitted.

3.4
The procedures used to interpret test scores, and, when appropriate, the normative or standardization samples or the criterion used should be documented.
Yes
While the procedures are documented, this information is proprietary and is therefore not provided to our users. Normative samples are described in the statistical manuals.

Copyright © 2006 Plumeus Inc.

1-888-855-6975

3.5
When appropriate, relevant experts external to the testing program should review the test specifications. The purpose of the review, the process by which the review is conducted, and the results of the review should be documented. The qualifications, relevant experiences, and demographic characteristics of expert judges should also be documented.
Yes
The experts who review the tests currently do so on a volunteer basis. This procedure is at the moment an informal rather than formal process.

3.6
The type of items, the response formats, scoring procedures, and test administration procedures should be selected based on the purposes of the test, the domain to be measured, and the intended test takers. To the extent possible, test content should be chosen to ensure that intended inferences from test scores are equally valid for members of different groups of test takers. The test review process should include empirical analyses and, when appropriate, the use of expert judges to review items and response formats. The qualifications, relevant experiences ,and demographic characteristics of expert judges should also be documented.
Yes
The types of items, response formats, test administration and scoring procedures are intended for a general audience. The formats used are familiar to almost all groups of test-takers. The questions asked are intended to be general enough for all audiences to relate to. In addition, analyses are made during the statistical examination of the test to determine whether certain groups perform differently on any items or types of items included on the tests.

3.7
The procedures used to develop, review, and try out items, and to select items from the item pool should be documented. If the items were classified into different categories or subsets according to the test specifications, the procedures used for the classification and the appropriateness and accuracy of the classification should be documented.
Yes
This documentation is kept in the records for each test. If changes are made, items removed, or subscales created, they are recorded in different drafts of the questionnaire.

3.8
When item tryouts or field tests are conducted, the procedures used to select the sample(s) of test takers for item tryouts and the resulting characteristics of the sample(s) of test takers for item tryouts and the resulting characteristics of the sample(s) should be as representative as possible of the population(s) for which the test is intended
Yes
The field tests are generally performed on queendom.com and psychtests.com, where the audience is extremely diverse. Subsets of this population can be

Copyright© 2006 Plumeus Inc.

1-888-855-6975

examined in order to obtain a more representative sample if the test will be used for a more specific population

3.9
When a test developer evaluates the psychometric properties of items, the classical or item response theory (IRT) model used for evaluating the psychometric properties of items should be documented. The sample used for estimating item properties should be described and should be of adequate size and diversity for the procedure. The process by which items are selected and the data used for item selection, such as item difficulty, item discrimination, and/or item information, should also be documented. When IRT is used to estimate item parameters in test development, the item response model, estimation procedures, and evidence of model fit should be documented.
Yes
See the section on dynamic test administration at the end of this document.

3.10
Test developers should conduct cross-validation studies when items are selected primarily on the basis of empirical relationships rather than on the basis of content or theoretical considerations. The extent to which the different studies identify the same item set should be documented.
No
Plumeus does not select test items in this manner. We develop concepts and test items based on theoretical analyses. The validity of items is verified empirically.

3.11
Test developers should document the extent to which the content domain of a test represents the defined domain and test specifications.
Yes
Every attempt is made to cover the full extent of the domain assessed. If, for practical reasons, some aspects need to be left out, this is made clear in the results of the test. Any missing elements are usually reported in the introduction so that users and test takers can be made aware of this.

3.12
The rationale and supporting evidence for computerized adaptive tests should be documented. This documentation should include procedures used in selecting subsets of items for administration, in determining the starting point and termination conditions for the test, in scoring the test, and for controlling item exposure.
Yes
We select anchor items that cover the full spectrum of a construct. Additional items are administered according to consistency assessment. Consistency of answers is determined either based on the variability of responses using a measurement of the standard deviation, or fuzzy logic.

3.13

When a test score is derived from the differential weighting of items, the test developer should document the rationale and process used to develop, review, and assign item weights. When the item weights are obtained based on empirical data, the sample used for obtaining item weights should be sufficiently large and representative of the population for which the test is intended. When the item weights are obtained based on expert judgment, the qualifications of the judges should be documented.
Yes
Plumeus frequently employs differential weighting of items. First, the weighting is designed on theoretical basis. Adjustments are made according to empirical data. Although Plumeus does document the weighting of different items or scales, this information is proprietary and is not available to end users.

3.14
The criteria used for scoring test takers' performance on extended-response items should be documented. This documentation is especially important for performance assessments, such as scorable portfolios and essays, where the criteria for scoring may not be obvious to the user.
No
This item does not apply.

3.15
When using a standardized testing format to correct structured behavior samples, the domain, test design, test specifications, and material should be documented as for any other test. Such documentation should include a clear definition of the behavior expected of the test takers, the nature of the expected responses, and any materials or directions that are necessary to carry out the testing.
No
This item does not apply.

3.16
If a short form of a test is prepared, for example, by reducing the number of items on the original test to organizing portions of a test into a separate form, the specifications of the short form should be as similar as possible to those of the original test. The procedures used for the reduction of items should be documented
Yes
With rare exceptions, the administration, types of questions asked, answer options, and all other aspects of the tests remain the same when an abridged version of a test is created. Scoring methods are based on the same principles, but typically limit the amount of detail provided in the report. Short forms use a representative sample of the questions from the full version, the same weighting of answer options, and are derived using empirical data.

3.17
When previous research indicates that irrelevant variance could confound the domain definition underlying the test, then to the extent feasible, the test

Copyright© 2006 Plumeus Inc.

1-888-855-6975

developer should investigate such sources of irrelevant variance should be removed or reduced by the test developer.
Yes
If need be, the test scores are adjusted for age, gender, education, or other sources of variance unrelated to the domain definition, unless such differences are inherent to the domain being assessed.

3.18
For tests that have time limits, test development research should examine the degree to which scores include a speed component and evaluate the appropriateness of that component, given the domain the test is designed to measure.
Yes
With rare exceptions, our tests do not have a time limit for completion. However, in the cases that they do (only in those where speed is relevant to the construct, i.e., meticulousness practical test, time management practical test, etc.), the influence of speed in the score is clearly spelled out.

3.19
The directions for test administration should be presented with sufficient clarity and emphasis so that it is possible for others to replicate adequately the administration conditions under which the data on reliability and validity, and, where appropriate, norms were obtained.
Yes
Our tests are administered online and clear instructions are provided at the beginning of the assessment.

Copyright© 2006 Plumeus Inc.

1-888-855-6975

3.20
The instructions presented to test takers should contain sufficient detail so that test takers can respond to a task in the manner that the test developer intended. When appropriate, sample material, practice or sample questions, criteria for scoring, and a representative item identified with each major area in the test's classification or domain should be provided to the test takers prior to the administration of the test or included in the testing material as part of the standard administration instructions.
Yes
Detailed instructions are provided for all tests. When necessary, sample items and additional instructions are included in the questionnaires in order to clarify how test-takers should answer the questions on the test.

3.21
If the test developer indicates that the conditions of administration are permitted to vary from one test taker or group to another, permissible variation in conditions for administration should be identified, and a rationale for permitting the different conditions should be documented.
Yes
For ARCH Profile clients, the tests are computer-administered in all cases. The recommended method of test administration is to assign the tests to individuals and give them in-house. However, it is also acceptable for them to assign the tests and have test takers complete the test on home computers. In all cases, the test taker is informed that a quiet environment without distractions is needed in order for them to be able to complete the tests.

3.22
Procedures for scoring and, if relevant, scoring criteria should be presented by the test developer in sufficient detail and clarity to maximize the accuracy of scoring. Instructions for using rating scales or for deriving scores obtained by coding, scaling, or classifying constructed responses should be clear. This is especially critical if tests can be scored locally.
No
Tests that are scored automatically online do not contain instructions for scoring. For paper and pencil tests, clear instructions are provided.

3.23
The process for selecting, training, and qualifying scorers should be documented by the test developer. The training materials, such as the scoring rubrics and example of test takers' responses that illustrate the levels on the score scale, and the procedures for training scorers should result in a degree of agreement among scorers that allows for the scores to be interpreted as originally intended by the test developer. Scorer reliability and potential drift over time in raters' scoring standards should be evaluated and reported by the person(s) responsible for conducting the training session.
No
Most tests are scored by computer, eliminating inter-rater consistency as an issue altogether. Paper-pencil versions of certain tests (mostly the short

Copyright © 2006 Plumeus Inc.

1-888-855-6975

versions) are available. However, such tests are scored using a simple scoring formula. Complex scoring is done exclusively using a software scoring engine.

3.24
When scoring is done locally and requires scorer judgment, the test user is responsible for providing adequate training and instruction to the scorers and for examining scorer agreement and accuracy. The test developer should document the expected level of scorer agreement and accuracy.
No
With a few exceptions, tests are scored by computer, eliminating inter-rater consistency as an issue altogether. When paper and pencil administration of a test is allowed, scoring techniques are free from potential sources of unreliability, barring gross mistakes in calculations.

3.25
A test should be amended or revised when new research data, significant changes in the domain represented, or newly recommended conditions of test use may lower the validity of test score interpretations. Although a test that remains useful need not be withdrawn or revised simply because of the passage of time, test developers and test publishers are responsible for monitoring changing conditions and for amending, revising, or withdrawing the test as indicated.
Yes
If evidence surfaces that indicates the validity of a test has been compromised, necessary changes are made to compensate. Periodically, every two to three years, the tests are revalidated and renormed to ensure that the test is still valid and reliable, despite changes in the construct over time. If extensive changes are needed, a revised version of the test is created.

3.26
Tests should be labeled or advertised as "revised" only when they have been revised in significant ways. A phrase such as "with minor modification" should be used when the test has been modified in minor ways. The score scale should be adjusted to account for these modifications, and users should be informed of the adjustments made to the score scale.
Yes
Users are informed about the release of a revised version of the test; minor changes are reported in a newsletter.

3.27
If a test or part of a test is intended for research use only and is not distributed for operational use, statements to this effect should be displayed prominently on all relevant test administration and interpretation materials that are provided to the test user.
No
Our tests are for operational use as well as for research. Items intended purely for validation purposes are clearly separated from test items and participation is based on opt-in principles.

Copyright© 2006 Plumeus Inc.

1-888-855-6975

4. Scaling, Norming,and Score Comparability

4.1
Test documents should provide test users with clear explanations of the meaning intended interpretation of derived score scales, as well as their limitations.
Yes
In the introduction included in all the test reports, supporting materials, and the results themselves, the scales are described clearly, and limitations stated when necessary.

4.2
The construction of scales used for reporting scores should be described clearly in test documents.
Yes
When several scales make up an overall score, or top-level factors, the relevant scales are stated and defined in the documentation.

4.3
If there is sound reason to believe that specific misinterpretations of a score scale are likely, test users should be explicitly forewarned.
Yes
In both supporting materials, and in the test reports themselves, care is taken to inform test users about any potential for misunderstanding of the test results. In addition, we encourage user feedback and quickly work to change any unclear text or interpretations to try to remove sources of misunderstandings.

4.4
When raw scores are intended to be directly interpretable, their meanings, intended interpretations, and limitations should be described and justified in the same manner as is done for derived score scales.
No
Our tests scores are given in a standardized, rather than raw format.

4.5
Norms, if used, should refer to clearly described populations. These populations should include individuals or groups to whom test users will ordinarily wish to compare their own examinees
Yes
In the statistical reports, the population used to validate the test is specified. Typically, normative information provides break-down by gender, age, education, race/ethnicity, field of work and position.

4.6
Reports of norming studies should include precise specification of the populations that was sampled, sampling procedures and participation rates, any weighting of the sample, the dates of testing, and descriptive statistics. The information provided should be sufficient to enable users to judge the appropriateness of the norms for interpreting the scores of local examinees. Technical documentation should indicate the precision of the norms themselves.

Copyright© 2006 Plumeus Inc.

1-888-855-6975

Yes
The statistical report contains all such information. Upon request, statistics can be run on a sample of the population that is more appropriate for the user's intended use.

4.7
If local examinee groups differ materially from the populations to which norms refer, a user who reports derived scores based on the published norms has the responsibility to describe such differences if they bear upon the interpretation of the reported scores.
Yes
Renorming is performed for translated tests to be used in culturally distinct populations.

4.8
When norms are used to characterize examinee groups, the statistics used to summarize each group's performance and the norms to which those statistics are referred should be clearly defined and should support the intended use or interpretation.
No
In a future release of ARCH Profile, we are planning to provide mean, standard deviation and percentile ranks with reference to groups defined by the manager (for example, managers and sales people). However, comparisons will be made only between an individual and the reference group, not among groups themselves.

4.9
When raw score or derived score scales are designed for criterion-referenced interpretation, including the classification of examinees into separate categories, the rationale for recommended score interpretations should be clearly explained.
Yes
For all test batteries that are designed to assess the test takers' ability, suitability or aptitude for a certain criterion, we give the rationale for this in the supporting materials. The rationale is research-based, and we can provide info on demand. Score interpretation is computer generated, and the rationale is provided in the test intro and test manual.

4.10
A clear rationale and supporting evidence should be provided for any claim that scores earned on different forms of a test may be used interchangeably. In some cases, direct evidence of score equivalence may be provided. In other cases, evidence may come from a demonstration that the theoretical assumptions underlying procedures for establishing score comparability have been sufficiently satisfied. The specific rationale and the evidence required will depend in part on the intended uses for which score equivalence is claimed.
Yes
Correlations are run to determine whether the test forms can be used interchangeably. However, at this point, we have alternative versions only for the Verbal IQ Test.

4.11
When claims of form-to-form score equivalence are based on equating procedures, detailed technical information should be provided on the method by which equating functions or other linkages were established and on the accuracy of equating functions.
Yes

Copyright© 2006 Plumeus Inc.

1-888-855-6975

Correlations are reported in the statistical manuals of the relevant tests. However, at this point, we have alternative versions only for The Verbal IQ Test.

4.12
In equating studies that rely on the statistical equivalence of examining of examinee groups receiving different forms, methods of assuring such equivalence should be described in detail.
Yes
The only test for which this applies is The Verbal IQ Test. For this test, the population used to validate the tests is the same (users of queendom.com and psychtests.com of all different ages).

4.13
In equating studies that employ an anchor test design, the characteristics of the anchor test and its similarity to the forms being equated should be presented, including both content specifications and empirically determined relationships among test scores. If anchor items are used, as in some IRT-based and classical equating studies, the representativeness and psychometric characteristics of anchor items should be presented.
Yes
See section 2.16

4.14
When score conversions or comparison procedures are used to relate scores on tests or test forms that are not closely parallel, the construction, intended interpretation and limitations of those conversions or comparisons should be clearly described.
No
This standard does not apply.

4.15
When additional test forms are created by taking a subset of the items in an existing test form or by rearranging its items and there is sound reason to believe that scores on these forms may be influenced by item context effects, evidence should be provided that there is no undue distortion of norms for the different versions or of score linkages between them.
Yes
Correlation studies are run to ensure that alternate or abridged forms taken from an existing test are comparable.

4.16
If test specifications are changed from one version of a test to a subsequent version, such changes should be identified in the test manual, and an indication should be given that converted scores for the two versions may not be strictly equivalent. When substantial changes in test specifications occur, either scores should be reported on a new scale or a clear statement should be provided to alert users that the scores are not directly comparable with those in earlier versions of the test.
Yes

Copyright© 2006 Plumeus Inc.

1-888-855-6975

See section 3.26. In addition, we advise users not to compare results from different versions unless we have done studies about the equivalency of the results. New test manuals are provided for revised tests.

4.17
Testing programs that attempt to maintain a common scale over time should conduct periodic checks of the stability of the scale on which the scores are reported.
Yes
This standard does not apply.

4.18
If a publisher provides norms for use in test score interpretation, then so long as the test remains in print, it is the publisher's responsibility to assure that the test is renormed with sufficient frequency to permit continued accurate and appropriate score interpretations.
Yes
Plumeus performs norming studies periodically for all tests.

4.19
When proposed score interpretations involve one or more cut scores the rationale and procedures used for establishing cut scores should be clearly documented.
No
As a rule, it is the responsibility of the user to establish cut scores for their intended purposes. If need be, representatives are available to help in this process. For some job fit batteries (for example, The It Job Fit Test, The Customer Service Job Fit Test) we do make recommendations based on cut scores. In this case, these scores are based on theory and empirical data and confirmed in validation studies.

4.20
When feasible, cut scores defining categories with distinct substantive interpretations should be established on the basis of sound empirical data concerning the relation of test performance to relevant criteria.
No
See 4.19

4.21
When cut scores defining pass-fail or proficiency categories are based on direct judgments about the adequacy of item or test performances or performance levels, the judgmental process should be designed so that judges can bring their knowledge and experience to bear in a reasonable way.
No
See 4.19

Copyright© 2006 Plumeus Inc.

1-888-855-6975

5. Test Administration, Scoring, and Reporting

5.1
Test administration should follow carefully the standardized procedures for administration and scoring specified by the test developer, unless the situation or a test taker's disability dictates that an exception should be made.
Yes
The majority of Plumeus' tests are scored online using a standardized scoring system. As a result, all tests are scored in the same way. In the case of paper and pencil tests, detailed instructions for scoring are included in the documentation for the test. The instructions are very simple as the only tests that we allow to be scored by hand are our most basic tests. In the event of a test-taker needing accommodation, instructions and questions can be read out loud, set into Braille, or translated in order to comply with state and federal laws.

5.2
Modifications or disruptions of standardized test administration procedures or scoring should be documented.
Yes
Our tests should be completed in one sitting, without interruption. Test-takers are clearly instructed to follow these instructions. Disruptions should be reported to the manager.

5.3
When formal procedures have been established for requesting and receiving accommodations, test takers should be informed of these procedures in advance of testing.
No
Although Plumeus does provide recommendations about how to accommodate users with disabilities, it is up to the prospective employer to implement such accommodations and to inform test-takers of them.

5.4
The testing environment should furnish reasonable comfort with minimal distractions.
No
It is the user's responsibility to ensure that this is the case. We do however instruct the test takers to make sure they are in a quiet area with minimal distractions for the duration of the testing period.

5.5
Instructions to test takers should clearly indicate how to make responses. Instructions should also be given in the use of any equipment likely to be unfamiliar to test takers. Opportunity to practice responding should be given equipment is involved, unless use of the equipment is being assessed.
Yes
Instructions are provided, and, if necessary, an example included to illustrate how to answer questions on a particular test. However, there are no special skills required to know how to answer the tests. No test developed by Plumeus requires use of

Copyright© 2006 Plumeus Inc.

1-888-855-6975

equipment with which users might be unfamiliar, provided that they have basic computer skills.

5.6
Reasonable efforts should be made to assure the integrity of test scores by eliminating opportunities for test takers to attain scores by fraudulent means.
Yes
We encourage users to have test-takers whose test scores will be compared take the tests in similar conditions (i.e., in the office or potential workplace). In addition, some of our tests have elements put into place to identify cheaters. Finally, our instructions clarify the importance of taking the tests seriously and being honest.

5.7
Test users have the responsibility of protecting the security of test materials at all times.
no
Test materials are only available online; scoring materials are proprietary and are therefore not available to the public. Finally, questionnaires online are encrypted to further prevent security breaches.

5.8
Test scoring services should document the procedures that were followed to assure accuracy of scoring. The frequency of scoring errors should be monitored and reported to users of the service on reasonable request. Any systematic source of scoring errors should be corrected.
Yes
Test are scored automatically online, therefore minimizing the potential for human error.

5.9
When test scoring involves human judgment, scoring rubrics should specify criteria for scoring. Adherence to established scoring criteria should be monitored and checked regularly. Monitoring procedures should be documented.
No
Except in rare cases, tests available from Plumeus are scored automatically online; no human judgment is required.

5.10
When test score information is released to students, parents, legal representatives, teachers, clients, or the media, those responsible for testing programs should provide appropriate interpretations. The interpretations should describe in simple language what the test covers, common misinterpretations of test scores, and how scores will be used.
Yes
This information is all covered in the reports of the tests.

5.11
When computer-prepared interpretations of test response protocols are reported the sources rationale, and empirical basis for these interpretations should be available, and their limitations should be described.
Yes
This information is all covered in the reports of the tests and in test manuals.

Copyright© 2006 Plumeus Inc.

1-888-855-6975

5.12
When group-level information is obtained by aggregating the results of partial tests taken by individuals, validity and reliability should be reported for the level of aggregation at which results are reported. Scores should not be reported for individuals unless the validity, comparability and reliability of such scores have been established.
No
This standard does not apply.

5.13
Transmission of individually identified test scores to unauthorized individuals or institutions should be done in a manner that protects the confidential nature of the scores.
Yes
Individually identified test scores are provided only to authorized individuals. Test reports are password-protected.

5.14
When a material error is found in test scores or other important information released by a testing organization or other institution, a corrected score report should be distributed as soon as practicable to all known recipients who might otherwise use the erroneous scores as a basis for decision making. The corrected report should be labeled as such.
Yes
See section 3.16. In addition, we advise users not to compare results from different versions unless we have done studies about the equivalency of the results.

5.15
When test data about a person are retained, both the test protocol and any written report should also be preserved in some form. Test users should adhere to the policies and record-keeping practice of their professional organizations.
yes
We keep all the answers to test questions and scores in a database. Reports can be generated at any time based on the answers. Data are stored in ARCH Profile accounts until the organization cancels the account, at which point they are instructed to print reports for their records. Back-up files are stored by Plumeus even after the account is cancelled, unless a representative of the organization requests in writing that the back-up be destroyed.
We need to implement this in ARCH

5.16
Organizations that maintain test scores on individuals in data files or in an individual's records should develop a clear set of policy guidelines on the duration of retention of an individual's records, and on the availability, and use over time, or such data.
Yes
All ARCH records are deleted when the company cancels the account and we advise the client that we are keeping a back-up file containing their records of testing unless they sign a waiver requesting that the records will be destroyed. The data is only available to authorized members of the organization at all points.

Copyright© 2006 Plumeus Inc.

1-888-855-6975

6. Supporting Documentation for Tests

6.1
Test documents (e.g., test manuals, technical manuals, user's guides, and supplemental material) should be available to prospective users and other qualified persons at the time a test is published or released for use.
Yes
Plumeus is presently compiling test manuals for all available tests. Meanwhile, statistical reports are available, and pop-up windows containing information about the test are free to view online. The information is also available in the test catalogue.

6.2
Test documents should be complete, accurate and clearly written so the intended reader can readily understand the content.
Yes
Every effort is made to create tests, interpretations, and supporting materials that are accessible to the average user. If technical terms are used, they are explained using terms with which the average layperson will be comfortable.

6.3
The rationale for the test, recommended uses of the test, support for such uses, and information that assists in score interpretations should be documented. Where particular misuses of a test can be reasonably anticipated, cautions against such misuses should be specified.
Yes
An outline of the test rationale and information about the relevant scales is provided in the test interpretation and in the statistical and technical manuals of tests. Whenever appropriate, Plumeus makes an effort to educate users about any potential unintended consequences, misuses, and other problems related to misuse of information. If research indicates that test scores have different meaning across groups, we explore the possible sources of these differences but refrain from making overarching judgments about a group based on a particular test.

6.4
The population for whom the test is intended and the test specifications should be documented. If applicable, the item pool and scale development procedures should be described in the relevant test manuals.
Yes
We develop our tests with a general audience in mind, provided that they are adolescent or adults. If a test is designed with a more specific population in mind, this is specified in the supporting materials.

6.5
When statistical descriptions and analyses that provide evidence of the reliability of scores and the validity of their recommended interpretations are available, the information should be included in the test's documentation. When relevant for test interpretation, test documents ordinarily should include item level information, cut scores, normative data, the standard error of measurement, and a description of the procedure to equate multiple forms.

Copyright© 2006 Plumeus Inc.

1-888-855-6975

Yes
This information is provided in the statistical manuals of the tests.

6.6
When a test related to a course of training or study, a curriculum, a textbook, or packaged instruction, the documentation should include an identification and description of the course or instructional materials and should indicate the year in which these materials were prepared.
No
This standard does not apply to our organization.

6.7
Test documents should specify qualifications that are required to administer a test and to interpret scores accurately.
Yes
This information will be provided in technical manuals for each test. However, this standard is not a major concern, since all of Plumeus' tests are scored online and test results are interpreted automatically. As a result, they can be interpreted by anyone with a command of the English language. However, if a company desires to develop benchmarking for a particular purpose or needs help interpreting the results, advice is available through Plumeus.

6.8
If a test is designed to be scored or interpreted by test takers, the publisher and test developer should provide evidence that the test can be accurately scored or interpreted by the test takers. Tests that are designed to be scored and interpreted by the test taker should be accompanied by interpretive materials that assist the individual in understanding the test scores and that are written in language that the test taker can understand.
Yes
For abridged versions of tests, simple scoring and interpretive reports come with the test.

6.9
Test documents should cite a representative set of the available studies pertaining to general and specific uses of the test.
Yes
In house validity studies are reported in stats reports and test manuals.

6.10
Interpretative materials for tests, that include case studies, should provide examples illustrating the diversity of perspective test takers.
No
This standard does not apply, as Plumeus does not include case studies in our results.

6.11
If a test is designed so that more than one method can be used for administration or for recording responses - such as marking responses in a test booklet, or on a separate answer sheet, or on a computer keyboard - then the manual should clearly document

the extent to which scores arising from these methods are interchangeable. If the results are not interchangeable, this fact should be reported, and guidance should be given for the interpretation of scores obtained under the various conditions or methods of scoring.
No
Online methods of recording responses and the method for recording answers in paper-and-pencil versions of tests may be different, but the outcome is the same unless a test taker did not indicate his or her response clearly in the paper-and-pencil version. Therefore, any difference in scores is incidental, and the results can be seen as interchangeable.

6.12
Publishers and scoring services that offer computer-generated interpretations of test scores should provide a summary of the evidence supporting the interpretations given.
Yes
This information will be part of the test manuals, and is available in the introductions of the test reports. Further information is available upon request.

6.13
When substantial changes are made to a test, the test's documentation should be amended, supplemented, or revised to keep information for users current and to provide useful additional information or cautions.
Yes
If a test is revised, a new statistical report is created.

6.14
Every test form and supporting document should carry a copyright date or publication date.
Yes
The test reference online includes the publication date, as does the supporting materials such as statistical reports and manuals.

6.15
Test developers, publishers, and distributors should provide general information for test users and researchers who may be required to determine the appropriateness of an intended test use in a specific context. When a particular test use cannot be justified, the response to an inquiry from a prospective test user should indicate this fact clearly. General information also should be provided for test takers and legal guardians who must provide consent prior to a test's administration.
Yes
General information about the use of the test is provided in the test catalog, informational pop-ups and test manuals. Responses made to inquiries about the appropriateness of a specific test are made promptly and with the needs of the test taker in mind, rather than in the interest of Plumeus. If a proposed use of a test is deemed inappropriate, this is stated clearly, and alternatives offered if any exist.

Copyright© 2006 Plumeus Inc.
1-888-855-6975

Fairness in Testing

7. Fairness and Bias

7.1
When credible research reports that test scores differ in meaning across examinee subgroups for the type of test in question, then to the extent feasible, the same forms of validity evidence collected for the examinee population as a whole should also be collected for each relevant subgroup. Subgroups may be found to differ with respect to appropriateness of test content, internal structure of test responses, the relation of test scores to other variables, or the response processes employed by individual examinees. Any such findings should receive due consideration in the interpretation and use of scores as well as in subsequent test revisions.
Yes
All subjects taking tests online on queendom.com and psychtests.com are asked to answer the same demographic questions regarding gender, age, ethnicity, along with relevant questions to determine construct validity of the particular test. If credible research does indicate that certain groups' test scores differ from others, we are able to address this and perform research studies to confirm or disprove this research.

7.2
When credible research reports differences in the effects of construct-irrelevant variance across subgroups of test takers on performance on some part of the test, the test should be used if at all only for those subgroups for which evidence indicates that valid inferences can be drawn from test scores.
Yes
If evidence suggests that irrelevant sources of variance may affect test scores, the scores of different subgroups are reported in statistical reports and the test-user has the responsibility of using the test only in populations for which it has been validated.

7.3
When credible research reports that differential item functioning exists across age gender, racial/ethnic, cultural disability, and/or linguistic groups in the population of test takers in the content domain measured by the test, test developers should conduct appropriate studies when feasible. Such research should seek to detect and eliminate aspects of test design, content, and format that might bias test scores for particular groups.
Yes
Plumeus has been including questions about gender, age, ethnicity, and education level. If differences exist, they are reported in the statistical report.

7.4
Test developers should strive to identify and eliminate language, symbols, words, phrases, and content that are generally regarded as offensive by members of racial, ethnic, gender, or other groups, except when judged to be necessary for adequate representation of the domain.
Yes

Copyright© 2006 Plumeus Inc.

1-888-855-6975

All of Plumeus' tests are developed with sensitivity in mind. Several test-developers review the questionnaire and test results, and any user feedback from the validity study is taken into account in revising test-items deemed inappropriate or offensive.

7.5
In testing applications involving individualized interpretations of test scores other than selection, a test taker's score should not be accepted as a reflection of standing on the characteristic being assessed without consideration of alternate explanations for the test taker's performance on that test at that time.
Yes
Particularly with skills assessments, the results clearly state that the person's score may have been affected by factors such as lack of sleep, hunger, or other short-term issues. We take care to provide potential alternative interpretations in the test reports.

7.6
When empirical studies of differential prediction of a criterion for members of different subgroups are conducted, they should include regression equations (or an appropriate equivalent) computed separately for each group or treatment under consideration or an analysis in which the group or treatment variables are entered as moderator variables.
Yes
We perform studies such as these for some tests, and when we do so, we include the regression equations in calculations of the scores when appropriate.

7.7
In testing applications where the level of linguistic or reading ability is not part of the construct of interest, the linguistic or reading demands of the test should be kept to the minimum necessary for the valid assessment of the intended construct.
Yes
Attempts are made, through consensus, to limit the reading level of the testing materials, within reason. This is especially true for tests where comprehension of English is not part of the intended construct.

7.8
When scores are disaggregated and publicly reported for groups identified by characteristics such as gender, ethnicity, age, language proficiency, or disability, cautionary statements should be included whenever credible research reports that test scores may not have comparable meaning across groups.
Yes
Whenever possible, Plumeus makes an effort to educate users about any potential unintended consequences, misuses, and other problems related to misuse of information. If research indicates that test scores have different meaning across groups, we explore the possible sources of these differences but refrain from making overarching judgments about a particular test. Normative information about subgroup differences are investigated and reported if appropriate. We are cautious when developing item pool to phrase questions in a clear way to avoid misinterpretation of test results.

7.9

Copyright© 2006 Plumeus Inc.

1-888-855-6975

When tests or assessments are proposed for use as instruments of social, educational, or public policy, the test developers or users proposing the test should fully and accurately inform policymakers of the characteristics of the tests as well as any relevant and credible information that may be available concerning the likely consequences of test use.
No
This section does not apply.

7.10
When the use of a test results in outcomes that affect the life chances or educational opportunities of examinees, evidence of mean test score difference between relevant subgroups of examinees should, where feasible, be examined for subgroups for which credible research reports mean difference for similar tests. Where mean differences are found, an investigation should be undertaken to determine that such differences are not attributable to a source of construct under representation or construct-irrelevant variances. While initially the responsibility of the test developer, the test user bears responsibility for uses with groups other than those specified by the developer.
Yes
If evidence arises that one or a few groups perform differently than another group, Plumeus looks first at sample size, sampling bias, differential item responses, and other issues that may be causing one group to perform differently. If no such issues are found, research is performed to see if other measurements of this construct have found similar differences.

7.11
When a construct can be measured in different ways that are approximately equal in their degree of construct representation and freedom from construct-irrelevant variance, evidence of mean score differences across relevant subgroups of examinees should be considered in deciding which test to use.
No
This section is the user's responsibility - we typically have only one assessment for a trait.

7.12
The testing or assessment process should be carried out so that test takers receive comparable and equitable treatment during all phases of the testing or assessment process.
Yes
While the bulk of this responsibility falls upon the users of our tests, Plumeus does provide equitable treatment in that everyone receives standardized instructions on the tests and has the same user interface.

Copyright© 2006 Plumeus Inc.

1-888-855-6975

9. Testing Individuals with Diverse Linguistic Backgrounds

9.1
Testing practice should be designed to reduce threats to the reliability and validity of test score inferences that may arise from language differences.
Yes
Tests are screened for difficulty level and questions deemed too complex are altered or removed. Some tests have been designed to be culture-fair. These tests contain only images and therefore can be said to be free of linguistic bias. Other tests are designed to assess language ability or other closely related concepts. In this case, no concessions are made to linguistic minorities, and their scores should be compared with the group norms as a whole. People who use these tests for employment generally do so because the ability to converse in English is an essential job function.

9.2
When credible research evidence reports that test scores differ in meaning across subgroups of linguistically diverse test takers, then to the extent feasible, test developers should collect for each linguistic subgroup studied the same form of validity evidence collected for the examinee population as a whole.
Yes/no
Since such differences are inherent when using tests written in English, we do not generally gather data about this subject. However, for those tests which purport to be culture fair, we are currently gathering validity information to ensure that tests are valid across cultural and language groups.

9.3
When testing an examinee proficient in two or more languages for which the test is available, the examinee's relative language proficiencies should be determined. The test generally should be administered in the test taker's most proficient language, unless proficiency in the less proficient language is part of the assessment.
Yes
The majority of our tests are offered only in English; however some tests have been translated into other languages. For those tests available in different languages, the clients are instructed to offer the test-taker a choice.

9.4
Linguistic modifications recommended by test publishers, as well as the rationale for the modifications, should be described in detail in the test manual.
No
We do not currently make recommendations about linguistic modifications.

9.5
When there is credible evidence of score comparability across regular and modified tests or administrations, no flag should be attached to a score. When such evidence is lacking, specific information about the nature of the modification should be provided, if permitted by law, to assist test users properly to interpret and act on test scores.
No
Different forms of tests (abridged versus full versions, and revised versus original) are screened in order to ensure that the tests can be interpreted in the same manner.

Copyright© 2006 Plumeus Inc.

1-888-855-6975

9.6
When a test is recommended for use with linguistically diverse test-takers, test developers and publishers should provide the information necessary for appropriate test use and interpretation.
Yes
Statistical reports are available for the culture fair-IQ test along with the logic IQ tests. The scores of linguistic minorities should not be affected by lack of comprehension.

9.7
When a test is translated from one language to another, the methods used in establishing the adequacy of the translation should be described, and empirical and logical evidence should be provided for score reliability and the validity of the translated test's score inferences for the uses intended in the linguistic groups to be tested.
Yes
We use a forward translation technique in translations of tests. The process by which the tests was translated will be described in all supporting materials, and documentation on our translation policy available upon request. Supporting studies of validity and reliability will also be included in the statistical manual.

9.8
In employment and credentialing testing, the proficiency level required in the language of the test should not exceed that appropriate to the relevant occupation or profession.
Yes
The English required to take our tests is usually of a moderate level. This is appropriate for most positions. If a question as it is originally written is deemed to use vocabulary that is inappropriately difficult, it is altered or removed. The reading level required for a specific test is provided in supporting materials for tests.

9.9
When multiple language versions of a test are intended to be comparable, test developers should report evidence of test comparability.
Yes
Any test in our battery that has been translated and is intended to be equivalent to the original version will be subject to the same validation effort as the original. If the norms are significantly different, the new test will be renormed to make the test equivalent.

9.10
Inferences about test takers' general language proficiency should be based on tests that measure a range of language features, and not on a single linguistic skill.
Yes
The Verbal IQ Test (Forms A and B) is our only test that fits this category; this test measures a wide range of language features.

Copyright© 2006 Plumeus Inc.

1-888-855-6975

9.11
When an interpreter is used in testing, the interpreter should be fluent in both the language of the test and the examinee's native language, should have expertise in translating, and should have a basic understanding of the assessment process.
No
Users of ARCH Profile are strongly discouraged from using interpreters during the testing process; to do so raises both psychometric and legal issues.

Copyright© 2006 Plumeus Inc.

1-888-855-6975

10. Testing Individuals with Disabilities

10.1
In testing individuals with disabilities, test developers, test administrators, and test users, should take steps to ensure that the test score inferences accurately reflect the intended construct rather than any disabilities and their associated characteristics extraneous to the intent of the measurement.
Yes
While to the extent possible we do try to ensure that the questionnaires reflect the intended construct, rather than any disabilities and their associated characteristics, at times this is unavoidable. However, test users can minimize this effect by offering accommodations or alternative types of testing for those individuals who have disabilities that may affect their scores.

10.2
People who make decisions about accommodations and test modifications for individuals with disabilities should be knowledgeable of existing research on the effects of the disabilities in question on test performance. Those who modify tests should also have access to psychometric expertise for so doing.
No
This issue is the responsibility of the client.

10.3
Where feasible, tests that have been modified for use with individuals with disabilities should be pilot tested on individuals who have similar disabilities to investigate the appropriateness and feasibility of the modifications.
No
No such modifications are made, however, a statement (pending) containing recommendations for how to accommodate individuals with disabilities will be posted in ARCH Profile.

10.4
If modifications are made or recommended by test developers for test takers with specific disabilities, the modifications as well as the rationale for the modifications should be described in detail in the test manual and evidence of validity should be provided whenever available. Unless evidence of validity for a given inference has been established for individuals with the specific disabilities, test developers should issue cautionary statements in manuals or supplementary materials regarding confidence in interpretation based on such test scores.
No
No such recommendations are made, however, a statement (pending) containing recommendations for how to accommodate individuals with disabilities will be posted in ARCH Profile.

10.5
Technical materials and manuals that accompany modified tests should include a careful statement of the steps taken to modify the tests to alert users to changes that are likely to alter the validity of inferences drawn from the test score.
Yes

Copyright© 2006 Plumeus Inc.

1-888-855-6975

A statement (pending) containing recommendations for how to accommodate individuals with disabilities will be posted in ARCH Profile.

10.6
If a test developer recommends specific time limits for people with disabilities, empirical procedures should be used, whenever possible, to establish time limits for modified forms of timed tests rather than simply allowing test takers a multiple of the standard time. When possible, fatigue should be investigated as a potentially important factor when time limits are exceeded.
No
This standard does not apply.

10.7
When sample size permits, the validity of inferences made from test scores and the reliability of scores on test administered to individuals with various disabilities should be instigated and reported by the agency or publisher that makes the modification. Such investigations should examine the effects of modifications made for people with various disabilities on resulting scores, as well as the effects of administering standard unmodified tests to them.
No
This is the responsibility of the user.

10.8
Those responsible for decisions about test use with potential test takers who may need or may request specific accommodations should (a) possess the information necessary to make an appropriate selection of measures, (b) have current information regarding the availability of modified forms of the test in question, (c) inform individuals, when appropriate, about the existence of modified forms and (d) make these forms available to test takers when appropriate and feasible.
No
This standard does not apply.

10.9
When relying on norms as a basis for score interpretation in assessing individuals with disabilities, the norm group used depends upon the purpose of testing. Regular norms are appropriate when the purpose involves the test taker's functioning relative to the general population. If available, normative data from the population of individuals with the same level or degree of disability should be used when the test taker's functioning relative to individuals with similar disabilities is at issue.
yes
First of all, the tests administered to job candidates should be job-related, and by extension, all candidates be judged on the same terms.
If it is the case that the test-taker's performance on a test could be affected by a disability, the employer receives a warning when assigning the test.
This decision is up to the employer; s/he should be aware of the relevant legal issues when making such a decision.

Copyright© 2006 Plumeus Inc.
1-888-855-6975

10.10
Any test modifications adopted should be appropriate for the individual test taker, while maintaining all feasible standardized features. A test professional needs to consider reasonably available information about each test taker's experiences, characteristics and capabilities that might impact test performance, and document the grounds for the modification.
No
A statement (pending) containing recommendations for how to accommodate individuals with disabilities will be posted in ARCH Profile.

10.11
When there is credible evidence of score comparability across regular and modified administrations, no flag should be attached to a score. When such evidence is lacking, specific information about the nature of the modification should be provided, if permitted by law, to assist test users properly to interpret and act on test scores.
No
This standard does not apply.

Copyright© 2006 Plumeus Inc.

1-888-855-6975