



## INTERNATIONAL TEST COMMISSION

### Guidelines for Practitioner Use of Test Revisions, Obsolete Tests, and Test Disposal

**2<sup>nd</sup> March, 2015, Version 1.0**

**Final Version**

**Document reference: ITC-G-TD-02032015**

The contents of this document are copyrighted by the [International Test Commission \(ITC\)](#) © 2013. All rights reserved. Requests relating to the use, adaptation or translation of this document or any of its contents should be addressed to the Secretary-General: [Secretary@InTestCom.org](mailto:Secretary@InTestCom.org).

### **Formally adopted**

This document was formally adopted by the ITC Council on March 2<sup>nd</sup>, 2015.

### **Published online**

This document was officially published online on June 24<sup>th</sup>, 2015, and can since be found online on the ITC website at <http://www.intestcom.org>.

### **Published in print**

This document was not yet published in print, but will be published in the International Journal of Testing, the official publication of the ITC.

### **Please reference this document as:**

International Test Commission (2015). International Guidelines for Practitioner Use of Test Revisions, Obsolete Tests, and Test Disposal.

## ACKNOWLEDGEMENTS

The ITC thanks Thomas Oakland (United States) who drafted the guidelines and served as committee chair for this project.

The ITC also thanks members of the project committee for their contributions to the production of the draft. These included Dave Bartram (UK), Osman Tolga Aricak (Turkey), Bruce Bracken, (United States), Ann Edwards (South Africa), Dragos Iliescu (Romania), Muhammad Rizwan (Pakistan), Eva Tideman (Sweden), Hazel Wheldon (Canada), and Solange Wechsler (Brazil).

The ITC thanks members of the ITC Council who provided comments on the first draft produced by the Committee.

## SUMMARY

These guidelines address issues associated with when to adopt revisions to or new versions of tests. Revisions can range from a norm update to a complete new version of the contents and the scales. These guidelines also address issues associated with the disposal of obsolete test materials.

These guidelines emphasize the importance of using sound professional judgment when deciding, on a case by case basis, the appropriateness of using either a new or revised version of a test or continuing to use the older version. This judgment involves balancing considerations relating to improvements offered by the newer version against the need for continuity of assessment afforded by use of the older one. Psychometric considerations, although important, are not the only considerations. Guidelines are provided that address practical, applied, professional, and psychometric issues to help inform professional judgment.

**CONTENTS**

**ACKNOWLEDGEMENTS**.....3

**SUMMARY**.....4

**CONTENTS**.....5

**INTRODUCTION** .....6

    Background .....6

    Ethical and Other Professional Guidelines and Standards for Test Use .....6

**THE GUIDELINES** .....8

    1. Scope.....8

    2. Basic Principles for Test Selection.....8

    3. Some Practical Issues .....11

    4. Psychometric Issues.....11

    5. Applied Issues.....13

    6. Research Issues .....15

    7. When a Test Is Obsolete .....15

    8. Methods to Dispose of Tests .....15

**REFERENCES**.....17

## INTRODUCTION

These guidelines are intended to clarify issues pertaining to whether and when to adopt revised tests as well as the disposal of obsolete tests. The guidelines are intended to inform test developers and test users as to questions and issues that may be considered following the release of a revised test.

The guidelines focus on test use in high stakes assessments that guide diagnosis and other decision-making. Standards for the selection of revised tests can be expected to be highest when making high stakes decisions. These crucial decisions include such services as forming a diagnosis, identifying pathology, deciding on intervention methods, assessing a person referred for academic difficulties, reaching important individual work-related personnel decisions, and in some forensic settings providing information to a court that may influence incarceration as well as life or death decisions. High stakes decisions typically are made on an individual following the use of individually administered measures.

These guidelines encourage test developers to be proactive by addressing key issues during the test revision process. The guidelines also will enable practitioners to identify salient issues when making a professional decision as to whether and when to adopt a revised test as well as when and how to dispose of obsolete or unwanted tests.

### Background

Psychology displays an abiding commitment to the fundamental principle of understanding individual differences. Its century-long efforts examine and help explain the impact of biological, cultural, demographic, economic, medical, and other qualities on personal development and behaviors. Professionals engaged in providing assessment services draw on this information, in part, to assist in their efforts to accurately describe behaviors and to form professional judgments as to their degree of normality, their trajectory and prognosis, their diagnosis, desired intervention and other treatment methods, together with other issues. The belief that each person is unique runs deep in these practices.

Tests and other assessment methods rightfully serve an important role in all areas of applied services, including health, forensic, education, work, as well as research. Quality tests provide reliable, valid, and cost-efficient methods to acquire needed information. In addition, some tests provide a statistically based standard against which to decide the frequency of the behavior. The belief that thoughtful care is needed in the selection and disposal of tests that help address referral and other applied issues also runs deep.

### Ethical and Other Professional Guidelines and Standards for Test Use

Professional associations can be expected to assume leadership for the development and promotion of guidelines and standards that govern professional services, including test development and use. These guidelines and standards are intended to educate members of the

profession and the public; promote public trust; reflect laws, public policy, and practice; and advocate for high professional standards.

Professionals can be expected to be knowledgeable of and utilize widely regarded guidelines and standards that address test development and use. *Standards for Educational and Psychological Testing* (American Educational Research Association, 2014) and the *EFPA Review Model for the Description and Evaluation of Psychological and Educational Tests, Version 4.2.6* (European Federation of Psychologists' Associations, 2013) underscore the importance of considering a test's reliability and validity together with other important psychometric features, such as the availability and appropriateness of norms.

The International Test Commission (ITC) has a long-standing commitment to promote suitable practices governing test development and use (Oakland, Poortinga, Schlegel, & Hambleton, 2001). The ITC's guidelines emphasize that practitioners should use tests ethically and display good testing practices. Professionals engaged in good testing practices evaluate the potential utility of test use, select technically sound tests in light of intended standards, recognize the importance of fair testing practices, prepare for the test session, administer tests properly, score and analyze test results accurately, interpret tests results properly, communicate the results clearly and accurately, and review the appropriateness of the test and its uses (<http://www.intestcom.org/guidelines/index.php>). These guidelines on whether and when to adopt revised tests as well as the disposal of outmoded tests reflect the ITC's long-standing commitment to promote suitable testing practices.

## THE GUIDELINES

### 1. Scope

These guidelines are intended to inform test developers and test users as to questions and issues that may be considered following the release of a revised test. They address issues relating to whether and when test revisions should be adopted and what should be done with obsolete test materials.

These guidelines primarily address issues related to tests used with individuals or small groups in applied as well as research settings where the test materials are under the direct control and management of the test user. These guidelines are not intended to apply to Internet-delivered testing where the management of the test content is in the hands of the distributor rather than the end user.

A wide range of professionals use tests. They include but are not limited to audiologists, counselors, educators, human resource personnel, nurses, psychologists, psychiatrists, rehabilitation specialists, and speech pathologists. The term *practitioner* is used to describe them all. These guidelines are intended to apply to all practitioners who use tests.

### 2. Basic Principles for Test Selection

#### 2.1. Test Developers and Users Have a Reciprocal Relationship

The relationship between those who develop and those who use tests is reciprocal. Test developers respond to market needs made known by practitioners and other test consumers through their engagement in developing new tests and revising them. Practitioners respond by purchasing or obtaining licenses to use tests, their protocols, and any supporting materials. Various issues (e.g. new research, new diagnostic criteria, and demographic changes) drive revisions and market needs.

#### 2.2. Practitioners Should Make an Informed Professional Decision When Adopting Revised Versions of Tests

Practitioners need to consider various issues when deciding whether and when to purchase and use revised tests. Their ultimate concern often centers on whether the revised test promotes a better understanding of a client and his or her welfare. Whether the revision reflects changes in our understanding and conceptualization of a construct, condition, or disorder as well as whether the psychometrics properties (e.g., reliability and validity) are upheld or improved also are important. This requires professional judgment about the test's eventual use based on a professional's background, training, and experience in light of providing a specific professional service.



### **2.3. Practitioners May Use Tests Other Than Those That Are Considered the ‘Gold Standard’**

Some tests have acquired an outstanding reputation and are considered the ‘gold standard’ for the assessment of certain construct(s). Nevertheless, such esteemed tests may not be suitable for use with all persons. Test selection should be based on an individual’s personal characteristics, clinical or applied needs, legal and ethical principles, overall psychometric properties, best practice standards, the quality of local translations or adaptations, and the extent to which any given test meets those needs.

### **2.4. Test Publishers Shall Describe and Justify the Need for the Revised Test**

Test publishers shall describe their rationale for the publication of the newer version of the instrument and highlight the research that drove the need for revision and/or updates. Newsletters, journals, and other published test reviews also may provide this information.

Test revisions may be driven by knowledge that the assessed behaviors are subject to substantial change over time, by significant demographic changes, from research that leads to improvements in theories and concepts that should impact test use, from changes in diagnostic criteria, or in response to test consumers demands for improved versions. The presence of these qualities warrants consideration of whether a revised test may be more suitable than an older version.

### **2.5. Test Selection Decisions Should Be Based on Evidence Regarding the Scientific Merits of the Revised Version**

Once a revised test is released, practitioners should review evidence as to whether it or the prior version of the test is more appropriate for the purpose of a client’s assessment. Decisions should be based on one’s professional judgments of the scientific merits, clinical and applied utility, and in light of an individual client, particularly when re-testing to measure change. For example, a practitioner may decide to use an older test with one client to maintain continuity in the assessment results and the revised test with another new client. Practitioners also may decide to delay the purchase of a revised test until more information is known about the possible impact of its adoption and use.

### **2.6. Test Selection Shall Be Based, In Part, On a Review of Changes Made In the Revision**

Practitioner decisions should be based, in part, on a review of information provided by the test publisher about the nature of the changes made to the test, and about specific changes made. Practitioners should consider the soundness of evidence provided by the test publisher and through other reliable sources as to the need for the revised measure and how the revision provides added value. Information from test reviews of the original measure may enable practitioners to determine whether the revision addresses concerns or issues that may have existed within the prior version.

### **2.7. Practitioners Should Obtain Training in the Use of the Revised Test**

Practitioners should take workshops, attend online webinars, and engage in other practical training exercises that review these changes, and learn new methods to administer and score tests and to interpret test data. Publishers may inform professionals who have purchased or are interested in purchasing their revised assessments regarding the availability of training resources through their websites. Independent scholarly information on a recently revised test is less likely to be available immediately upon its release. Practitioners may decide to delay adoption of major revisions until independent research has validated the revision.

## **2.8. Practitioners Shall Review External Policies That May Govern Test Use**

Some policies may mandate the use of a revised measure. Thus, practitioners may not be allowed to select the tests to be used and instead may be required to use a specific test consistent with laws, public or private policy, or other external requirements. On occasion, practitioners may judge a mandated test to be unsuitable for use with a particular client. When this mismatch occurs, the practitioner should recommend to the prevailing authority a measure judged to be more suitable for the referral and the reasons for requesting permission for its use. The practitioner should record this effort in a report.

## **2.9. Practitioners Shall Not Justify the Use of an Older Version Due to Their Personal Attachment**

Various other issues influence practitioners' choice to obtain and use revised tests, including economic costs, availability of newer versions in the practitioner's country (including translations/adaptations), and the possible integration of the previous version with existing decision-making schemas that would need to be revised if the newer version rather than the previous version were used. Additionally, practitioners may form a strong and abiding attachment to an older version. Emotional feelings (e.g., attachment to the previous version) and reluctance to update one's knowledge of the new version shall not be considered valid reasons for failing to adopt a new version.

## **2.10. Professional Judgment Whether to Adopt and Use a Revised Test Shall Consider Multiple Issues**

No one issue, single standard, or criterion for the purchase and use of tests and their revisions takes precedence over all others. When selecting and using tests, practitioners shall consider multiple standards in light of the client's individual characteristics and the specific reasons for using the tests. Specific rules as to when an earlier version of a test should be discontinued and the newer one used exclusively do not exist.

Decisions as to whether and when to transition to a new version of a test should be based on a professional's informed judgment of the scientific merits and practical utility of the test, not an arbitrary time frame, and following careful consideration of the specific practice setting and client needs together with new research and diagnostic criteria.

### 3. Some Practical Issues

#### 3.1. Financial Considerations Influence Adoption of Revised Tests

The cost of test materials and test record forms impacts test purchases and use. Although the purchase and use of a revised test may be preferred for many reasons, its cost may be too burdensome to lead to its immediate adoption. For example, individual practitioners, researchers, schools, and other institutions, especially in emerging economies, may have limited funds to devote to test purchases. These entities often have fewer tests at their disposal and, once acquired, may use them for years, often bypassing the purchase of revised editions and forms.

#### 3.2. Practitioners Should Not Copy Test Materials

Practitioners may attempt to justify the photocopy of manuals and test forms due to financial considerations. However, they should not do so. Such practices violate legal and professional standards.

#### 3.3. Practitioners Should Consider Non-financial Considerations When Adoption a Revised Test

The decision whether to purchase a revised test is likely to be influenced by other practical issues: the ease and amount of time required to administer and score the revised test, whether the revised test requires interpretations consistent with current needs, the quality of its test materials, the need for local norms, and the availability of the test in one's dominant national language.

### 4. Psychometric Issues

Decisions as to whether to purchase and use a revised test often hinge on whether improvements are found in the newer edition's norms, reliability, and validity. Practitioners should decide whether the revised test is more useful for a particular purpose or whether the previous version's norms have become outdated.

#### 4.1. Test Selection Should Consider the Relevance of the Revised Test Norms

Revisions of tests that provide norm-referenced interpretations typically focus on updating norms. Thus, decisions regarding the purchase and use of a revised test may be more favorable when the norms for the revised test were obtained recently, are sufficient, and better represent the population. Practitioners should examine test norms closely to determine the extent to which they include evidence that a client's characteristics (e.g., race, SES, gender, age, language, impairments, diagnosis) are represented in the norms. Issues pertaining to cultural relevance and fairness also may need to be considered when considering the value of the revised instrument's norm sample.

#### **4.2. Test Selection Should Consider the Correspondence between Prior and New Norms and Their Possible Impact**

Mean score differences between the original and revised scales are common. These differences may have implications regarding a client's existing diagnosis, program eligibility, qualifications to obtain public financial support, and other possible life-altering decisions and thus deserve professional review and judgment together with client consultation. Advice from the publisher should be sought on the issue of equating new and old norm-based or raw score cutoffs.

#### **4.3. Test Selection Should Consider the Effects of New Norms on Inter- and Intra-individual Diagnostic Practices**

Assessment practices typically consider two forms of test data analysis: inter-individual differences and intra-individual differences. An examination of inter-individual differences allows practitioners to identify differences displayed by a client on one or more traits, behaviors, or other characteristics compared to the performance of a population of persons with whom the client is being compared. Intra-individual differences allow practitioners to compare variations between two or more traits, behaviors, or other characteristics based on data from one client. Although norm data may impact both forms of interpretation, the outcomes related to inter-individual differences typically are more apparent.

#### **4.4. Test Selection Should Consider Whether New Norms Are Needed For Tests That Are Not Norm-referenced**

Some tests do not employ norm-referenced interpretations and instead use criterion referencing to guide interpretations. Thus, although norms are not used in their interpretation, normative data may be obtained to examine the test's psychometric features. For tests without norms, the age of norms does not rise to the same level of importance as it is for norm-referenced instruments. However, the nature of the test items and the relevance of the test content are important and should be considered when deciding whether to use a revised measure. Evidence of equating is needed for revisions of diagnostic tests that define categories of characteristics in terms of raw score cutoffs.

#### **4.5. Test Selection Shall Consider the Revised Test's Reliability**

Evidence of the extent to which a revised test displays sufficiently high levels of measurement precision (or low standard errors of measurement), stability over time or forms, and inter-rater reliability may influence whether it is purchased and used. Reliability standards for individually administered measures generally should exceed those for group-administered measures. In addition, reliability standards for higher stakes decisions should be more stringent than those for lower stakes decisions.

#### **4.6. Test Developers Should Provide Evidence of the Revised Test's Validity**

The extent to which a revised test displays suitable validity is more complex than the assessment of its reliability. When viewed in light of psychometric qualities, test developers

should provide evidence of two forms of validity: construct validity (i.e. whether there is evidence that the test measures what it purports to measure) and criterion-related validity (i.e. whether the test scores relate to external criteria).

For construct validity, evidence can be internal (relating to the factorial structure of the test); convergent (i.e. evidence of relationships between the test scores and other measures of the same or similar constructs); discriminant (i.e. evidence of lack of relationships between the test scores and measures of different constructs); or experimental (i.e. evidence of change in the test score in response to manipulation of the construct that the tests purports to measure).

For criterion-related validity, distinctions often are drawn between concurrent (when both the test and the criterion are administered at the same time) and predictive (when the test scores are obtained in advance and used to make predictions about the criterion measures – as in job selection and subsequent measures of job performance) designs. Criterion-related validity can include the ability of an assessment to differentiate between different groups, the outcome of training interventions, or the relationship between psychological attributes and important attainments (e.g. academic performance).

Validation evidence supporting the original test may be generalized to the revised version when there is good evidence of measurement equivalence. Ideally, the revised tests should provide evidence of improvement in either construct validity, criterion-related validity or both, or demonstrate similar validity that can be obtained more efficiently.

#### **4.7. Test Selection Shall Consider Various Forms of the Revised Test’s Validity**

Practitioners shall examine evidence that the revised test displays adequate validity as well as avoiding construct irrelevance (i.e., that the test scores are not affected by processes extraneous to its intended construct). For example, efforts are needed to insure test data are not attenuated by personal qualities not intended to be assessed. Additional issues pertaining to test validity are discussed below under applied issues.

### **5. Applied Issues**

#### **5.1. Practitioners Shall Select Tests That Help Them Individualize Their Work**

Psychology’s abiding commitment to its fundamental principle of understanding individual differences may be seen most prominently in the services provided by practitioners to individual clients. Practitioners shall tailor their assessment methods in light of relevant client qualities. These may include age, family constellation, disabilities, gender, education, language, work histories, race, social class, and other personal qualities that may impact behavior. Thus, these qualities as well as the client’s needs, legal and ethical issues, and best practice standards inform test selection and use.

#### **5.2. Test Selection Shall Consider Referral Issues**

Test use often occurs in response to a referral that requests desired services. Referral issues may range from identifying early onset disorders in infants to later onset disorders in the elderly, and include issues pertaining to selection and promotion. Looked at more broadly, referral issues often warrant the assessment and diagnosis of persons with possible cognitive, motor, neuropsychological, personality, or sensory disorders. Tests that best address referral issues shall be selected.

Some presenting issues are simpler to diagnose (e.g., the identification of students who are intellectually gifted). Other issues are more difficult and require a complex process that warrants differential diagnostic methods—namely those that distinguish between two or more similar disorders by comparing their symptoms. Evidence of mean score differences across relevant subgroups is needed for this process (Thorne, 1955). Again, referral issues shall inform test selection, given the goal to identify and use the test that best enhances decision-making in light of a client’s personal qualities.

### **5.3. Practitioners Shall Select Tests That Best Improve Decisions**

Decisions to retain and use an existing test or to use a revised test shall consider which of the tests (i.e., either the prior test or the revised test) is most likely to yield data that lead to improved decision-making (e.g., diagnostic and treatment/intervention decisions, and behavioral and life-event changes). For example, a practitioner who has used an older version of the test for years may believe there is added value in continuing to use it. Another practitioner may need to determine the stability of a person’s data—information best achieved through the use of the previous test version (e.g., assessing premorbid conditions). Additionally, a practitioner may have years of experience using an older version of a test and find it more useful when examining intra-individual differences.

Some practitioners may rely on specific subtests of a test that no longer are included in the revised test. Practitioners may elect to continue to use subtests from an earlier version of the test and justify their use in reports provided the subtests are deemed to have adequate reliability and validity (in particular, construct/discriminant validity).

### **5.4. Practitioners Should Select Tests In Light Of Diagnostic Standards When Diagnosing Disorders**

Test selection should consider diagnostic standards. For example, the *Diagnostic and Statistical Manual of Mental Disorders: Fifth Edition* (DSM-5; American Psychiatric Association, 2013) and the World Health Organization’s *International Classification of Diseases-10* code (American Psychiatric Association, 2005) are used internationally to diagnose various psychiatric and psychological disorders. Their editions are subject to change, thus leading to possible changes in standards used to diagnose disorders. Tests used to diagnose disorders addressed in these manuals are likely to be revised, if needed, in light of diagnostic changes. Thus, practitioners should decide whether to purchase and use the revised test that may be more sensitive and specific to qualities important in the revised diagnostic standards.

Professional practices guided by the revised diagnostic standards are likely to be improved through the use of the revised measure—one designed to provide data that are more sensitive and specific to the behaviors that characterize the newly revised disorder. However, on occasion, assessment practice requires a retrospective evaluation, perhaps for forensic purposes, and thus may rely on a prior diagnostic standards and the use of an older version of a test to make this diagnosis. The decision not to use the latest version of a test and the reasons for this decision should be described in the report.

## **6. Research Issues**

Tests commonly are used as independent or dependent variables in research. Standards for deciding whether to use an older or a revised test in research are more flexible than standards that impact practice. Professionals engaged in research may find the above discussion to be relevant to their decision-making. For example, those engaged in longitudinal research may continue using the existing measures despite the availability of a revised form. Those planning new research have greater flexibility and may be attracted to the features of the revised instrument. Note, too, that some agencies that provide external support for research as well as review boards that examine research in light of legal and ethical issues promulgate standards for test selection and use. Research support may be contingent on the use of measures preferred by an agency.

## **7. When a Test Is Obsolete**

### **7.1. Professionals Should Be Able To Justify When a Test Is Obsolete**

A decision as to when a test is obsolete lies in the reflective and informed judgment of a professional who is committed to the selection and use of the best version of a test to address referral issues and client qualities. The professional can expect to be able to articulate support for this judgment. Generally speaking, a test is obsolete when its underlying theory, item content, norms, or technical adequacy no longer meet the needs for its intended purpose, professional standards, or when its continued use would lead to inappropriate or inaccurate decisions or diagnoses.

## **8. Methods to Dispose of Tests**

### **8.1. Professionals Shall Not Dispose of Tests Prematurely**

By itself, the launch of a revised test does not make the prior test version obsolete. Practitioners and agencies engaged in using a test shall retain the test, its manual and protocols, and other supporting materials at least until a subsequent version of the test and its supporting materials are obtained and fully adopted in practice. Practitioners recognize future needs and requirements may require the availability of the older version for some years. Examples include the need to refer to the test materials when engaged in scholarship

as well as addressing challenges to the accuracy of test data, the need to verify its administration and scoring, together with its use for forensic purposes.

## **8.2. Professionals Shall Dispose of Tests Properly To Maintain Test Security**

Upon making a decision to dispose of a test, professionals shall maintain test security by not giving or selling the test to unqualified persons. They should first determine if the test could be returned to the publisher for disposal (e.g. tests that are obtained under a license remain the property of the publisher and should be returned when the license expires). If they are not returned to the publisher, the contents shall be shredded or incinerated to ensure the test's continued security.



## REFERENCES

- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington DC: American Psychological Association.
- American Psychiatric Association (2005). *Diagnostic and statistical manual of mental disorders, fourth edition (DSM-IV: International Version with ICD-10 Codes)*. Washington DC: American Psychiatric Association.
- American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders, fifth edition*. Washington DC: American Psychiatric Association.
- European Federation of Psychologists' Associations (2013). *The EFPA review model for the description and evaluation of psychological and educational tests: Test review form and notes for reviewers. Version 4.2.6*. Brussels: EFPA.
- Oakland, T., Poortinga, Y., Schlegel, J., & Hambleton, R. (2001). International Test Commission: Its History, current status, and future directions. *International Journal of Testing*, 1 (1), 3-32.
- Thorne, F. (1955). *Principles of psychological examining*. Brandon, VT: Journal of Clinical Psychology.