Clinical and Research Implications of Revising Psychological Tests

Marshall L. Silverstein
Long Island University

Linda D. Nelson
University of California, Irvine

This article integrates those of other contributors to this special section, "Methods and Implications of Revising Assessment Instruments," to underscore important conceptual factors to consider when undertaking test revisions. These considerations include determination of when test measures have become sufficiently understood to be incorporated in a test revision, cohort effects, revision of administration formats and test instructions, and comparisons of performance levels across test versions. The discussion of these factors also takes into consideration clinical practice and educational implications of making a transition to revised test versions.

The revision of a major psychological test is an effort that is intended to have an influence extending approximately one generation. However, there exist few guidelines in the literature regarding clinical and research adaptations to incorporating new editions of tests, despite the fact that planned revisions of tests have been undertaken for well over 50 years. Although revisions are more frequent now, to our knowledge there has not been a systematic examination of considerations involved in revising tests. Our concluding article represents an attempt to synthesize the views of the other contributors to this special section, adding several of our own.

Because revisions emerge out of a need to enhance the psychometric characteristics of tests, including updating of norms, what a test assesses conceptually is not often reconsidered apart from its psychometric properties. Further, incorporating views about sociocultural factors and cohort effects such as progressive increments in intellectual capabilities (the so-called Flynn effect) has implications extending beyond achieving superior psychometric or measurement sophistication. In addition, facilitating comparisons between existing versions of tests and their revised editions (e.g., equating test scores) is rarely anticipated. Consequently, the implications of converting to a revised version require greater attention concerning clinical examinations and longitudinal research that study performance levels over time. There is also room for more carefully addressing how to go about revising test formats and instructions, particularly when there exists more than one format of a test in common use, for example, tests of verbal fluency or word list learning.

Our discussion emphasizes conceptual concerns that merit consideration when one creates new test editions. We begin with a discussion of the other contributions to this special section, organized by broad themes to emphasize the links among the articles. We thus stress integrating factors that are usually considered on their own merit but that may be seen to better advantage as congruent ideas. Accordingly, we discuss the articles grouped by the following ideas: (a) establishing when test measures and the psychological concepts underlying these tests have become sufficiently understood to be incorporated in a test revision, (b) considering cohort effects such as the Flynn effect and sociocultural factors, and (c) reconceptualizing the psychological meaning of tests and their principal measures.

In our second section, we address the problem of standardization of test instructions and administration formats to be considered when tests are revised. We include here a comment about the matter of updating tests such as the Wechsler intelligence scales that are often used in conjunction with tests that are not themselves revised very often. We also provide a discussion of implications of test revision on longitudinal studies and comparability of test measures during the transition from one version to another.

Discussion of the Special Section

When Is a New Measure or Concept Sufficiently Understood to Be Incorporated in a Test Revision?

Strauss and Spreen (2000) provided a discussion of the fundamental question underlying much that is crucial to this special section by asking, in effect, why tests should be revised. They suggested that a revision provides a better means to answer questions about individuals’ functioning than the preexisting version of the instrument. Thus, tests should be revised when their principal constructs are better understood and when levels of performance can be reliably measured or discriminated. Reise, Waller, and Conroy (2000) discussed a different but nevertheless related point about the psychometric purity of a test. They argued that some good reasons for revising tests result from matters such as the internal consistency reliability of test scores, generalizability of factor structures across groups, and the persistence of questions.