

Standards for Educational and Psychological Tests and Manuals, prepared by a joint committee of the American Psychological Association, the American Educational Research Association, and the National Council on Measurement in Education. Washington, D. C.: American Psychological Association, Inc., 1966. Pp. iv + 40. \$1.00.

The Professional Service Representative of a Test Publisher

It is the opinion of one test developer and publisher (this one) that *Standards* may well be one of the half dozen most important publications in the history of measurement. Written in clear, concise English that is a tribute to the skill of its several authors, it enunciates fully a set of principles which have been growing for fifty years among the professionals of testing.

This booklet's two predecessor publications—*Technical Recommendations for Psychological Tests and Diagnostic Techniques* published by the American Psychological Association (APA) in 1954 and *Technical Recommendations for Achievement Tests* published by the American Educational Research Association (AERA) and the National Council on Measurements Used in Education in 1955 (now called the National Council on Measurement in Education [NCME])—represented a first and very large step toward codification of professional expectations for published tests. *Standards* is a second and even more impressive stride in the same direction.

The book opens with an excellent description of its own history and purposes and methods, of the audiences who may utilize it, of the cautions to be observed by those who do use it—a modest and wholly objective set of directions for the reader. Then it goes on to present the “standards” in well-organized sections devoted to (a) dissemination of information, (b) interpretation, (c) validity, (d) reliability, (e) administration and scoring, and (f) scales and

[Editor's note: Shortly before the *Standards for Educational and Psychological Tests and Manuals* was released, the editor sought reviews from several individuals with different professional orientations. At the beginning of each review is a caption or heading that describes the editor's categorization of the professional identification of each person who on extremely short notice contributed a review of the *Standards*. Deep appreciation is expressed to each reviewer who participated in evaluating the *Standards*.]

norms. The clarity and comprehensiveness of its coverage of critical topics are sure to gain wide use for this publication in the *teaching* of measurement as well as in the technical assessment of standardized tests.

Not all publishers will agree that all the standards are appropriate, of course. Perhaps not even one publisher will think that every recommendation is both important and well-stated. But nearly all will agree that on every important point the committee has taken a position that is shared by substantial numbers of specialists in the field—a position that can be defended with vigor. This means that the committee actually is speaking for the profession by reflecting its divisions of opinion as well as its more nearly unanimous points of view, instead of legislating its own rules for the publishers. So a listing of the specific points upon which the reviewer holds an opinion differing from that of the committee would amount to little more than an interesting exercise; a different reviewer would produce a different list—and have an equally hard time proving that his position on a given point is better than the committee's.

As one would expect in a publication developed by a committee (even when the final writing is done by one or two very competent people) there is some unevenness in technical sophistication, as well as in expository and editorial styles, from section to section. Compared with the concepts and procedures specified in the section on reliability, for example, the recommendations pertaining to scales and norms are more often unspecific and occasionally over-simplified. And the section on scales and norms would be better with an introduction as useful and well-written as the introductions to the sections on validity and reliability. But this is strictly a *comparative* criticism, for all sections of the book are admirably well done, just as they are.

Continuing the trend which was accelerated by its predecessor publications, this book will make things even more difficult for the individual test-developer as a publisher of operational tests. Authors who act as their own publishers just cannot afford the costs of meeting these standards at the point of publication; by the time they can hope to complete most of the essential standards, the content of their tests will have become obsolete. In this particular effect upon publishing practice, *Standards* offers a fairly accurate reflection of economic and technical circumstances that influence publishing. Development of tests that satisfy current requirements in the field requires large financial investment and enormous technical resources.

Standards should serve its central purpose well. By the time this review appears, surely, the major developers of tests will know the recommendations by heart and will be planning to include in fu-

ture instruments those elements or standards which they have not already built into their publications. Further, measurement technicians who buy and use tests will very quickly build *Standards* into assessment procedures for test materials. Finally, it ought to become evident to authors and teachers in the field that this publication is an admirable statement of proper usage and nomenclature in the language of testing—amounting almost to a manual of style—which should be imitated widely.

In conclusion, any reviewer whose career is related to test publishing should be permitted one small, wistful hope. Partly as a consequence of the efforts of professional groups like APA, AERA, and NCME the technical standards of test publication have improved tremendously in the last two decades. It would be fair to say, however, that as a result of such general improvement a great many tests and manuals now are far better than the uses to which they often are put. A highly sophisticated instrument works no better than a cruder tool in the hands of an apprentice just learning the trade. To put high-standard instruments more often to high-standard uses, then, the technical qualifications of those who choose, administer, and interpret tests should be brought to the technical level of the best tests they employ. Because it is evident that nearly all the recommendations in *Standards* need only the slightest turning or re-phrasing to be wonderfully applicable as *standards for the users of tests*, and because no other group could possibly muster a comparable combination of professional influence and skill, it is to be hoped that the combined committees on test standards will undertake next the development of just such a publication; and make it as good as this one is.

JOHN E. DOBBIN
Educational Testing Service

A Measurement Specialist and Executive of a Large Test Publishing Firm

Publication of *Standards for Educational and Psychological Tests and Manuals*, the up-dated version of the *Technical Recommendations* for Psychological Tests and Diagnostic Techniques, and for Achievement Tests, prompts this test publisher to ask himself several questions: What effects have the *Technical Recommendations* had on test publishing in the decade since their appearance? What effects have they had on test selection, interpretation, and use? Are the new *Standards* likely to be better or worse in any of these respects than the original standards?

One approach to the first question would be to compare the manuals for tests published by the major publishers during, say, the past three years with the manuals of tests published in the

three years prior to issuance of the *Recommendations*. While the reviewer has not gone through this exercise, he does not hesitate to assert that the quality of test manuals has, indeed, improved over the past decade, nor to attribute much, perhaps most, of this improvement to publication of the *Standards*. Other influences for good have, to be sure, been operative: the *Mental Measurements Yearbooks*, the general increase in level of sophistication concerning tests, even, paradoxically, the wave of criticism of testing, which has forced the test makers to be on their mettle; but there is no gainsaying the impact of the *Recommendations*.

To the reviewer's knowledge, the *Recommendations* have been very much in the minds of the staffs of the major test publishers, in whose hands lies the responsibility for the preparation of most test manuals. They have perceived the *Recommendations* as a helpful codification of what the profession thinks it reasonable to expect test makers to provide in the way of information about their instruments. "Conscience," someone has written, "is a small voice telling us what we know." The *Standards* may tell the responsible test maker little he does not, or should not know, but it is not a bad thing that he be told, in some organized authoritative way, what his fellow professionals expect of him. Most publishers, the reviewer would judge, view the *Standards* (even those classed as "Essential") as definitions of ideal practice rather than rigid prescriptions. Few, if any, manuals, even among those published in recent years, conform to the letter of the *Standards* in every respect; but it is a fair presumption that where a manual does not accord in every particular with the *Standards*, it is not because the *Standards* were ignored, but rather because a decision was taken, for good reasons or bad, not to abide by the *Recommendations* in the particular instance. This reviewer has sensed no disposition on the part of any major publisher to dismiss the *Standards* as unreasonable, unattainable, or irrelevant; indeed, it is hard to imagine that any serious test author or publisher would lightly disregard a body of recommendations developed with such obvious thought and care, and enjoying the endorsement of the most directly involved professional groups.

It is, of course, not to be expected that every recommendation, or the classification of recommendations as "essential," "very desirable," or "desirable," will command universal endorsement. This reviewer, for example, does not agree totally with the formulations in the revised *Standards* on the use of correction-for-guessing formulas, on provision of tables of equivalence between new and revised forms as a universal practice, on handling the matter of local norms, on treatment of test scores over time, on provision of slope and intercept information in connection with all validity coefficients; he shares the committee's bias in favor of a standard-score

system of reporting and interpreting, but can sympathize with competing views, and so on. Every practitioner will have his own list of places where his judgment differs to some extent from that embodied in the *Standards*. Such lack of total agreement seems to the reviewer unimportant, and perhaps an inevitable consequence of the less-than-fully-developed state of our science. What is important is that the reader sense the goal the committee had in mind in calling for certain kinds of information, and be guided by the spirit of the recommendations.

Beyond the obvious use of the *Standards* as a guide in the preparation of test manuals, this publisher, at least, has found them helpful both with respect to the evaluation of test manuscripts being considered for publication, and in the training of editorial staff.

Less easy to answer are the questions on the effect of the *Recommendations* on test selection, use, and interpretation. There is little evidence to lead the reviewer to believe that many test users are now influenced in their selection of tests by the extent to which their manuals conform to the *Technical Recommendations*. Neither does there appear to be convincing evidence of correlation between improvement in test manuals and improvement in test use. Does this argue against the usefulness of the *Standards*, or call into question the wisdom of the sponsoring bodies in devoting so much time and effort to their production? The reviewer thinks not. He concludes that what is called for is much better dissemination of the *Standards*, and better training of test users in the application of the information which, thanks to the *Standards*, is increasingly available in test manuals. It is unrealistic to pretend that the modal user of educational or psychological tests today is behaving very differently from his counterpart of a decade ago as a result of publication of the *Standards*, and perhaps the *Standards*-writing groups never expected such an outcome. But there are grounds for some small optimism: most measurement textbooks that have appeared in the past five years have devoted attention to the *Standards*, which attention, it is to be hoped, will result in improved test interpretation and use.

How do the revised *Standards* compare with the earlier ones? In structure, organization, scope, and general tone, the revised *Standards* are very similar to the original *Recommendations*—indeed, surprisingly similar in view of the amount of time and study devoted to preparation of the revision. Insofar as the *Standards* constitute a kind of index of the state of the science and the art, one is impressed at how little things seem to have changed in the ten years between the two editions. The treatment of validity, while somewhat expanded and recast, adheres to the concepts of content, construct, predictive and concurrent validity elaborated in the original *Recommendations*, although predictive and con-

current validity have been combined into one category. Most notable changes occur in the section on reliability; here the committee has espoused wholeheartedly an analysis-of-error-variance approach, a development which this reviewer counts as a decided improvement. Recommendations in the areas of administration and scoring, and scales and norms, are much like the original.

Whether or not the revised *Standards* will be more efficacious in improving the quality of test manuals and level of test usage depends less on the changes between the new and the old editions than it does upon what steps are taken to bring the new *Standards* to the attention of test makers and test users. As far as the test makers are concerned, it is safe to assume that the new *Standards* will receive as much attention and observation as did the *Recommendations*. The real task continues to be that of consumer education aimed at increasing the user's ability to utilize the information that the test publishers are being stimulated to provide.

A final word seems in order. To conform fully to the spirit of the *Standards* imposes formidable obligations on the test maker. The development and dissemination of all the information called for is a costly process, and often implies a long period of pre-publication experimentation. If the conscientious publisher is willing to undertake these additional developmental costs, and seeks to recover them, as indeed he must, in the form of higher prices for materials, he would like to feel that he will not thereby be placed at a competitive disadvantage and that the test user, in turn, will recognize an obligation to support better standards of test development by not preferring materials simply because they are less expensive.

Test makers and test users, APA and AERA, are indebted to the committees that labored so hard both on the original *Recommendations* and on the new *Standards*. Their task has been a difficult and perhaps a frustrating one, since in the nature of things the outcomes of their labors are so hard to assess. This reviewer judges that they have discharged their duties with sophistication, technical competence, and good sense. And if the *Standards* now and again seem to set goals that are unrealistic in a world of time and money, well, what is wrong with a dash of idealism?

ROGER LENNON
*Harcourt, Brace, and World
Test Division*

The Professor of Psychology

After comparing the 1966 *Standards* with the 1954 *Recommendations*, the reviewer gained the impression that there are not many major changes. The 1966 Committee worked long and hard. Thus it

is a tribute to the perspicacity of the 1954 group that their work has lasted so well.

The new publication is, however, a little firmer in tone. One facet of this attitude is the change in title from *Recommendations* to *Standards*. Another example is the fact that in the 1954 *Recommendations* there appeared, in a section on Revision and Extension, the statement "The recommendations are intended to be used without reference to any enforcement machinery." No such statement appears in the 1966 *Standards*. Still another instance appears in Item A1.22, which introduces the idea that promotional material for a test should also not be misleading. Likewise, Item A2.31, concerning the necessity for new data when a short form of a test is brought out, has been raised from Very Desirable to Essential.

A new feature is a topical index, and this should be very useful. It appears, too, that it would have been very helpful for the Committee to have prepared one or more outlines for the organization of a test manual, with citations of the paragraphs of the Standards that are relevant to each item. The reviewer had the feeling that standards bearing on a particular portion of a test manual were somewhat scattered. A confirmation of this impression is found in the number of different sections listed under some of the headings in the index.

One prominent modification is in the discussion of validity, where the terms predictive and concurrent validity have been combined and labeled criterion-related validity. The appropriateness of content validity for achievement tests comes in for its proper emphasis, and the discussion of construct validity has been considerably clarified. Apparently twelve years of discussion and controversy about this latter type has borne some fruit.

The section on Reliability has become more sophisticated, with advocacy of analysis of error variance as the most informative approach. Reliability is defined as accuracy, and the reviewer must confess the Committee has not brought him along with them on this point. It seems that a test can be reliable but biased.

One notes with approbation that the audience to whom the *Standards* are directed has been upgraded from competence at the level of a single course in tests and measurements to one of two or three such courses, plus two semesters of statistics. This explicit recognition that not just anyone is capable of judging the merit of a test is a healthy sign.

The reviewer hopes that the shift in these *Standards* toward more rigor will stimulate the testing fraternity also to move toward a more stringent exercise of concern over what appears on the market. There has been reliance upon voluntary cooperation for a long time and, although there has been some progress, sometimes the most ethical of the test producers have been the ones to suffer.

The necessity for some kind of sanctions seems to be indicated both by the current importance of testing in our society and by the criticisms which have grown out of, and fed upon, legitimate public interest in what is going on. Perhaps a first step could be the establishment, by the three bodies issuing these *Standards*, of a "Seal of Approval" for test manuals!

JOHN E. MILHOLLAND
University of Michigan

The Professor of Educational Measurement

This brief set of guidelines for published assessment devices, like its predecessors, is certain to have a marked impact in the testing field. The document is a well integrated and substantially improved revision of two earlier independent, and partially overlapping, efforts by APA and AERA-NCME committees. It is clearly written in language that places minimal demand on the reader's measurement background. The tone of the *Standards* is suggestive, not dictatorial; care is taken neither to dictate practice nor to discourage change and innovation. The principal request throughout the volume is for an honest portrayal of known information about an instrument. The reader will observe that most of the issues raised in the review are of minor consequence, often relating to matters of personal preference.

The format employed closely parallels that of the earlier versions, categorizing the recommendations into six divisions. The initial section on "Dissemination of Information" presents excellent guidelines for promotional literature on published tests. The second section, "Interpretation," makes helpful recommendations regarding information that should be included in test manuals to assist users in making correct interpretations of the test's results, e.g., making the distinction between statistical and practical significance explicit.

The "Validity" section is substantially modified from previous versions. The three primary validity types, content, criterion-related, and construct, are presented in a unified framework in such a way that it is clear to the reader that a complete study of any test typically involves information about all three types. Topic-by-process matrices are recommended for standardized achievement tests. Criterion-related validity is an integration of the former predictive and concurrent categories. The generally excellent presentation of construct validity fails to make one helpful distinction; namely that between the validity of a construct (e.g. test anxiety) versus the adequacy (or validity) of a particular test as a measure of the construct. The recommendation for uniformity in the use of the term "item discrimination" rather than "item validity" for item-total score relationships is to be commended. The reviewer

would like to have seen some suggested guidance regarding the use of particular item discrimination indexes to assist in reducing the undesirable heterogeneity in current practice.

Although the need for cross-validation is clearly indicated, especially when multiple predictors are involved with small samples, the recommendation to apply a correction for shrinkage to multiple correlations when they are to be presented as evidence of criterion-related validity would seem to have been appropriate. The reviewer would have preferred to have seen greater encouragement given to the reporting of confidence intervals and/or standard errors of validity and reliability coefficients. In addition, a suggestion to explore possible non-linear relationships between novel tests or criteria, particularly when concerned with non-cognitive variables would seem to have been in order.

The excellent "Reliability" section emphasizes the determination of components of error variance in tests, a distinct improvement in approach. No longer are coefficients classified into types; the use of suitable descriptive phrases that convey the meaning of reported coefficients is encouraged.

The brief section on "Administration and Scoring" parallels the general high quality of the document. The reviewer would have preferred, instead of recommending that the "correction for guessing" formulas be applied on non-power tests, that experimental data supporting the efficacy of the particular method employed be provided. To the reviewer's knowledge, the relative value of a "correction for guessing" formula in all situations has not been definitively established. It does not seem inconceivable that in some situations the gambling set would be positively correlated with a criterion; hence corrected scores might have less validity than uncorrected scores.

The final section on "Scales and Norms" commendably recommends the use of standard scores in reporting test results more strongly than was done formerly. With respect to the definition and meaning of grade placement scales, the reviewer would have preferred some encouragement toward increased uniformity. Some such direction would probably help *reduce* the unnecessary lack of comparability of grade equivalence from test to test.

In summary, this document is an outstanding example of high quality of content expressed in succinct, non-technical language, both of which will greatly contribute to its usability. Every major objective of the report appears to have been abundantly achieved; in fact, the product might aptly be considered as typifying a standard for such *standards*.

KENNETH D. HOPKINS
Laboratory of Educational Research
University of Colorado

The Clinical Psychologist

Clinical psychology, to which the testing movement owes much of its early origins (and present plight), has an important stake in the continuing health and vitality of the movement. Any effort that will help the clinical psychologist to select from the many and varied test procedures now available and in development is bound to have a salutary effect on the currently chaotic scene. The criteria developed for reporting test information in the current *Standards*, if accepted by those who write test manuals, should ease the problem considerably.

Clinicians are singled out for mild rebuke on the problem of quantification of the results of clinical instruments, particularly the projective techniques. The claim that projective techniques cannot be submitted to the same standards employed in evaluating other psychological tests is examined briefly and, in the main, dismissed. Much more on this particular topic can be found in Zubin *et al.* (1965), who argue for the return of projective techniques to the psychometric tradition. An additional implied criticism of test validation of clinical instruments is the collection of data on the basis of availability. The number of studies conducted on the basis of searching through hospital or clinic files after reading an article on a new scoring method for an old technique is beyond calculation.

In addition to these comments, clinical psychologists will find many lively topics in this jam-packed pamphlet which will have particular relevance for testing in clinical situations. The committees, for example, suggest that test manuals, where appropriate, report on the results of fakability studies. This might be expanded to a suggestion that a "faking" key be developed, put through a try-out, and reported. Concern for acquiescence response set or "yes-no" type responses is also expressed, along with suggestions for its detection. The problem of "base rates" for diagnostic instruments appears to have been ignored, although the amount of misclassification or overlapping is listed as an essential item for inclusion in the test manual. How well will a test work in a practical situation if, in the validation study the cut-off score makes very few errors in distinguishing between fifty normals and fifty schizophrenics? While the base rate can be expected to vary from one practical situation to another, the reader of a test manual might be interested in the answer to this question in selected situations.

The problem of validity is simplified into three aspects: content, criterion-related, and construct. Most of the suggestions, as might be expected, are centered on criterion-related validity. Perhaps the one most ignored aspect of the problem is content validity: the nature of the universe being sampled and the adequacy of the sampling. The *Standards* suggest that "the manual should justify the

claim that the test content represents the assumed universe tasks, conditions, or processes (p. 12)." How well does a series of drawings, to which subjects make up stories, represent a universe of situations? What is the universe of items from which questions on a personality questionnaire should be drawn? The reader can see the impossibility of answering these questions at the present time. Psycholinguists are becoming concerned with this problem (Coleman, 1964), and a similar concern may take place in psychological testing.

When test manuals of assessment procedures now employed in clinical situations are evaluated by the criteria in the *Standards*, it will become apparent that clinical psychology will become the chief beneficiary of the new standards. If psychologists of all fields will insist on these standards in the test manuals of the future, the effect cannot help but be healthy and profound.

REFERENCES

- Coleman, E. B. Generalizing to a Language Population. *Psychological Reports*, 1964, 14, 219-226.
Zubin, J. Eron, L. D., and Schumer, Florence. *An Experimental Approach to Projective Techniques*. New York: Wiley, 1965.

PHILIP HIMELSTEIN
*Texas Western College of the
University of Texas*

The Measurement and Evaluation Specialist in a Large Public School System

The authors of *Standards for Educational and Psychological Tests and Manuals* are to be commended for their efficient and easy-to-read presentation of so many well conceived recommendations which in some cases utilized some rather involved concepts. Only occasionally are there sections that discuss concepts which exceed the technical competencies of the typical educator. In those areas where educators may encounter cognitive difficulties in measurement concepts the document has presented background information to the reader. In addition, examples and comments are provided to explain the standards. The profession should salute the cooperative efforts of such a distinguished group of mental measurement authorities for their extended dedication to complete the difficult assignment.

To the educator especially, criteria for determining certain aspects of the quality of tests and test manuals are strongly needed because of the variety of backgrounds and training in educational measurement of teachers, principals, and guidance workers and because of the many and sundry tests for school use that have man-

uals with diverse emphases. The document has developed such criteria. Although the publication may be used by the most astute student of mental measurement, it lends itself well to the school teacher, counselor, and administrator. The *Standards* include examples based on subjects ranging from reading readiness tests of the kindergarten child to college and vocational placement tests.

In an era where commercialism has been developed to a high degree in all aspects of business, it is quite appropriate that the publication remind test publishers that manuals, "must avoid using high-pressure advertising techniques." For many of the currently used tests to include all the recommendations, although highly desirable, would pose a rather monumental task on their part. However, it is hoped that test publishers will accept the document as a standard and construct their tests and manuals in keeping with the recommendations in accordance with the standard as conscientiously as they would adhere to a code of ethics.

The format of the document is such that it includes for purposes of clarification through example strengths (as well as deficiencies) of many tests in the field. These models are used frequently as bases for many of the standards.

Although this reviewer feels that the *Standards* could be adhered to without undue stress by test publishers, one statement, (C5.3), has elements of fantasy to expect test publishers to admit that any of their validity samples are, "made up of records accumulated haphazardly." One area untouched by the standard which would be of special assistance to elementary school educators is that of the use of practice tests for students. Primary teachers frequently seek inquiry regarding the use of practice tests prior to the administration of pupils' first tests or types of tests new to pupils.

A concern of the present reviewer is that in section D1.5 from the background of D1.4 a reader may become confused that a statistical significance of difference between two scores of a profile is associated with reliability *per se*. Students of measurement and school staffs who are consumers of tests, all too frequently, become confused from their own self-imposed misconceptions of this difference. Nothing in the *Standards* should enhance this misconception.

The *Standards* not only have presented recommended characteristics which should be included in tests and test manuals but also have, in some cases, indicated challenges or recommended activities that publishers should engage. Two examples which are exemplary of such activities are C7.21 which indicate that new ability tests, "must do more than simply duplicate the measurement of verbal and quantitative ability" and D6.30 which indicates that manuals for general mental ability tests should, "report correlations and

changes in means and standard deviations between tests administered one year apart, two years apart, and three years apart."

The usefulness of the *Standards* to the school workers can be tremendous. The document as a potential to assist educators to assess more effectively tests and manuals is indisputable. However, the previous *Standards* often were not utilized or publicized to the staffs of a large portion of school districts in the country. As these revisions have been made of the 1954 and 1955 *Recommendations*, so too, should ways for implementing the present standards be reviewed. It is the challenge of present educators to develop procedures to publicize and to implement the standards. Educators could publish a supplementary document which would provide additional recommendations for school use. Three examples of materials that could be included in such a document are recommendations for establishing elements of a good testing program, ways to provide in-service training of test interpretation to teachers, and characteristics of a good board-of-education report of a district's test data.

With regard to the myriad of schools and school districts, those agencies which have responsibilities to serve schools have a problem of how to disseminate and implement the standards. Whether the *Standards* become or are used as a tool to improve the use of tests so that schools receive the benefits of the intent of the *Standards* is a function of those staff members in schools and district office personnel who are responsible for testing programs. It is recommended that educators establish local and regional meetings to discuss implications of the *Standards* at their respective level.

Local, regional, and state educational organizations are urged to coordinate the assessment of tests and manuals according to the *Standards*.

JERRY G. GARLOCK
County Schools of Los Angeles

The Textbook Writer in Measurement and Evaluation

One of the truly helpful documents for test authors, test publishers, and test users was the supplement to the *Psychological Bulletin* entitled *Technical Recommendations for Psychological Tests and Diagnostic Techniques* published in 1954. Beyond a doubt, this publication represents a kind of benchmark in a series of steps leading to the establishment of widely accepted standards for psychological tests and inventories. Today, it is widely mentioned and quoted by the members of the audience to which it was addressed. Moreover, it tended to standardize terminology in testing as it appears in journals, books, and test manuals. Finally, it generated a number of useful manuscripts, including a number of journal ar-

ticles concerning construct validity and a companion publication entitled *Technical Recommendations for Achievement Tests* which appeared in 1955.

Over ten years have passed since the *Technical Recommendations* were published. Now, a revision is available. It, too, is the product of extensive committee work. The product of the efforts of the committee is truly another contribution of significance. The revision concerns both psychological tests and inventories as well as educational tests, and thereby eliminates the need for the two existing publications.

The audience for the revision is described as one which is fairly sophisticated in testing matters. The committee intended that the material included be meaningful to those who have a level of formal training between the master's degree and doctorate in education or psychology at a superior university. Obviously, the classroom teacher who may have had little or no formal education in testing is excluded. This is regrettable, since this very large group is deeply involved in educational testing today, and should have an adequate grasp of major points included in the publication. Nevertheless, the elimination of this audience by the committee is understandable. Now, it is the task of textbooks in educational and psychological testing to capture the ideas prepared by the committee and present them to untrained teachers in such a way that they will be competent users of tests.

The principal section of the revision is the section concerning the standards themselves. As in the case of the 1954 publication, six major subdivisions are included; namely, Dissemination of Information, Interpretation, Validity, Reliability, Administration and Scoring, and Scales and Norms. The most significant of the six are the two dealing with validity and reliability. The remaining are shorter and in some ways less important.

The subdivision concerning dissemination of information and the subdivision concerning interpretation resemble greatly those of the 1954 publication. Two observations are in order, however. First of all, the point is made in several instances that test manuals should indicate clearly that which is *not* measured by the scores yielded. Warning should be given with regard to that which might be thought of as being measured by a test score, but which actually is not. Realistic cautions of this type to the casual user of tests are invaluable. It would be refreshing, indeed, if test manuals would point out needed evidence concerning the usefulness of the test which is missing.

Secondly, one cannot help but note Section A-2 which states that "it would appear proper in most circumstances for the publisher to withdraw a test from the market if the manual is 15 or more years old and no revision can be obtained." Although one can

quarrel with the limitations of 15 years, one cannot quarrel with the intent of this standard. Revisions of tests and test manuals often appear too infrequently.

Even the casual reader of the *Standards* will quickly discover that the discussion of validity is changed in one major way. Whereas the 1954 publication cited four kinds of validity, the revised publication mentions three. Content and construct validity remain virtually unchanged. The original concurrent and predictive validity had been combined into one category known as criterion-related validity. This makes sense, since the only important difference between concurrent and predictive validity is the time element. Emphasis on criterion-related validity is consistent with certain recent publications concerning criterion-related research. Such a categorization of research and categorization of validity reduces the need for detailed description of both.

A point of major emphasis in the validity standards as well as the reliability standards is sampling. Certainly, there has been a tendency by test authors and test publishers to underestimate the importance of this feature of their tests. Samples of overt behavior, samples of subject-matter areas, and samples of subjects are sometimes mentioned briefly with little attempt to defend the strength of the sampling process used. Indeed, sometimes the population which the sample is thought to represent is barely described. The significance of this point is presented well by several standards concerning criterion-related validity. Important criteria to be applied in evaluating the worth of criterion-related validity determinations are listed.

As in the case of the validity subdivision, a noteworthy change has taken place in the reliability subdivision. In addition to modernizing parts of the discussion, the committee decided to abandon some of the terminology used in the 1954 report. Rather than speaking of coefficients of equivalence, coefficients of stability, and the like, it prefers a more complete statement about the reliability coefficient. This statement would briefly describe the nature of the reliability determination, and would not necessarily give it a title such as those previously used.

The section concerning administration and scoring is one of the shorter sections and is essentially complete insofar as the audience of the manuscript is concerned. This brevity, however, should not be interpreted to mean that the standards listed are necessarily of less importance than others. Some of the practices found in schools with regard to test administration and scoring are deplorable. Statements in test manuals are often clear and emphatic. Somehow, however, the message is partially or totally lost to some teachers. One of the frustrations of the testing profession is its inability to communicate with the test users—for example, teachers—with re-

gard to the importance of standardizing administration and scoring practices.

In short, it must be said that the 1966 revision is a fine improvement over the 1954 publication. One of the strengths of the revision yet unmentioned is the set of parenthetical comments which provide examples of violations of a particular standard or successful adherence to it. Although the 1954 publication used this technique as well, the parenthetical comments are more numerous and more pertinent in the 1966 revision.

The addition of the index is excellent. Although it is not an essential part of the publication, it certainly is a helpful part. Another useful addition might have been a selected bibliography concerning standards for tests and manuals. It is obvious that particularly good readings exist in the areas such as validity and reliability. Perhaps in another decade still another committee will examine the current edition and see fit to make this improvement as well.

J. STANLEY AHMANN
Colorado State University

Personnel Psychologist in a Large Industry

Possibly never in the history of psychology has so much cogent content been packed in so few pages as in *Standards for Educational and Psychological Tests and Manuals*. In this document, every sentence is highly meaningful, and every example or comment appears necessary for clarity. The authors are to be commended for their conciseness without loss of meaningfulness. The publication represents a considerable improvement over its predecessor, *Technical Recommendations for Psychological Tests and Diagnostic Techniques*, although this earlier document was of extreme value in clarifying many of the issues involving psychological tests.

The strongest features of the new publication lie in the discussions of and the standards for validity and reliability. The discussion on each of these topics should be read by every person who uses tests in any way. The material in these sections has relevance for any psychological research involving tests and certainly should not be, in applicability, limited to test publishers.

It is noteworthy that the interrelations of various types of validity are stressed and that it has been suggested that an earlier classificatory system for reliability coefficients be abandoned. If test publishers follow the new recommendations for reporting reliability data, some of the problems the test user has often previously faced in interpreting such data will become minimal. It is also notable that the section on construct validity is considerably

augmented in the new publication; the cautions mentioned relative to construct validity should be carefully considered by every test publisher and user.

Some persons actively engaged in employment psychology may not feel that *Standards for Educational and Psychological Tests and Manuals* is responsive enough to some of the social and technical issues facing the personnel psychologist today. It is to be noted, however, that it is specifically stated that, ". . . primary responsibility for improvement of testing rests on the shoulders of test users." A further consideration is that some of the presently most pressing problems in employment psychology are likely to be of a temporary nature and will probably not be so important when educational opportunity is more nearly equalized throughout all socioeconomic levels.

There are, however, in the document several points which every personnel psychologist should consider. Among these are strictures against implying that a test measures an "innate" ability and using the labels, "culture-free" and "culture-fair." The cautions expressed relative to the use of moderator variables should be exercised in employment research.

From an editorial viewpoint, the publication is excellent. A rather complete index will aid in making *Standards for Educational and Psychological Tests* a more useful reference than its predecessor.

MARY L. TENOPYR
North American Aviation, Inc.

Language Testing: The Construction and Use of Foreign Language Tests by Robert Lado. New York: McGraw-Hill Book Co., 1964. Pp. xxiii + 389.

In an age of such unprecedented and universal change as ours it may seem presumptuous and pointless to single out change in any one educational field. Still, the upsurge of interest in language teaching and the curricular changes resulting from it present one aspect that is probably unique. While few if any American educators ever seriously advocated dropping science, or mathematics, or social studies, from the secondary school curriculum, a sizable part of the educational community considered the teaching of foreign languages, classical or modern, largely a waste of time. The events of the last thirty years have taught us that the world outside our borders cannot safely be ignored and that, in order to coexist with other peoples, it is necessary to understand and speak their languages. When we took a good look at the type of language teaching going on in our schools, however, we were dismayed to discover that it was indeed a waste of time. The student who left college—