

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Social and Technical Issues in Testing:
Implications for Test Construction and Usage

Buros-Nebraska Series on Measurement and
Testing

1984

1. Filling the Gaps Between Test Outcomes and Usage: An Introduction

Barbara S. Plake

University of Nebraska- Lincoln, bplake@unl.edu

Follow this and additional works at: <https://digitalcommons.unl.edu/burostestingissues>



Part of the [Educational Assessment, Evaluation, and Research Commons](#)

Plake, Barbara S., "1. Filling the Gaps Between Test Outcomes and Usage: An Introduction" (1984). *Social and Technical Issues in Testing: Implications for Test Construction and Usage*. 3. <https://digitalcommons.unl.edu/burostestingissues/3>

This Article is brought to you for free and open access by the Buros-Nebraska Series on Measurement and Testing at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Social and Technical Issues in Testing: Implications for Test Construction and Usage by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

1 Filling the Gaps Between Test Outcomes and Usage: An Introduction

Barbara S. Plake
University of Nebraska–Lincoln

Why do we have tests? What useful purposes do they serve? How can test results be used to make decisions? How can a test be proved to provide accurate and usable information? Questions such as these have been posed recently by a concerned public who have become more aware of and concerned about testing, test quality, and appropriate test usage. Their questions are challenging, legitimate queries that can and should be addressed by members of the measurement community.

Some of the questions being asked by the public are value laden, providing topics for many thoughtful but heated debates. For example: Would we be better off as a society if we did not have tests? Should testing be banned? Other questions are technical in nature and require accurate answers from the measurement community, which communicates to the public the present state of the art in measurement, assessment, and interpretation. Finally, questions such as “How can tests be used to eliminate the errors made in the selection process?” can provide an impetus within the measurement field for both theoretical and empirical development and yet are not ones that can, at least so far, be definitely answered.

The measurement field should take serious stock of itself and assess, as well as possible, the boundaries of its capabilities. From this assessment, it would be possible to communicate with the public about what testing *can* do, *may* be able to do, and is *incapable* of ever doing. At the present time, however, there appears to be an informational and expectational gap concerning what can be possible with the use of test results. Unless measurement experts and test users obtain a direct line to the angels, for example, error-free measurement will never be a reality!

Part of the communication and expectation gap can be assigned to a lack of measurement sophistication on the part of the public. Measurement course work and classes are not readily accessible to the public as a whole and may not be truly meaningful and usable to the public even if they were. A well-meaning but confused public provides fertile ground for test misunderstanding and misrepresentation by both knowledgeable and unknowledgeable test representatives. Tests enter into the lives of the public in so many ways; questionnaires, market surveys, school achievement batteries, classroom exams, and admission screenings are only a few possibilities. Yet the knowledge level of the public is minimal at best with regard to test information and interpretation.

Another part of the blame for the communication and expectation gap belongs to the measurement and testing professionals. Careful theoreticians are the first to caution on too rapid application of test advances into test usage and decisions. The state of the art is not as advanced in criterion-referenced testing or latent-trait modeling as some practitioners would want the public to believe. In addition, we are only now beginning to come to grips with decision-making models for test usage. Thus, a clear and purposeful statement (for public consumption) of what tests can and cannot do needs to be addressed by measurement and testing professionals. This would be an important *first* step in narrowing the gap.

Until such a statement is made, societal confusion and concern will abound. Confusion is fostered by the fact that decisions about test quality, application, and utility are made regularly by persons who are not trained as psychometricians. Legislative and legal decisions by politicians and judges who mandate and dictate test usage and disclosure only serve to widen the communication and expectation gap further.

PURPOSE OF THE VOLUME

The purpose of this volume is to investigate social and technical influences on test development and usage. As such, the volume can be viewed as making initial progress toward identifying what testing can and cannot do. This is accomplished first by establishing what some of the social influences are that impact tests and second by documenting some current technical aspects of testing. The volume provides essential preliminary information on how tests can be used and may be interpreted.

The intent of the volume is to present state-of-the-art content on: (1) characteristics that tests should have to be valid for use in decision making; (2) public awareness and social–legal issues that influence the credibility of tests that are used in decision making; (3) applications of tests in the decision-making process; (4) cognitive psychology’s impact on test development and vice versa; (5) quality issues of test development, packaging, sales, and usage; and (6) technical advances in test validation. These components are found in the five chapters of

Section I: *Social and Technical Influences*. Section II: *Influences on Aptitude and Achievement Test Development and Usage* is composed of three chapters that provide an integrated example of how social and technical issues have affected the development and usage of aptitude and achievement tests.

OVERVIEW OF THE CHAPTERS

Section I

Section I begins with the keynote presentation, “Struggles and Possibilities: The Use of Tests in Decision Making” from the first Buros–Nebraska Symposium on Measurement and Testing, and is authored by Dr. Ellis B. Page. Breaking from the style found in the other chapters in the volume, Dr. Page’s chapter is essentially a transcription of his symposium presentation because it was the keynote presentation for the symposium and therefore set the stage for the subsequent chapters within the section.

Dr. Page brings to focus a series of concerns that are relevant to the topic of uses of tests in a decision-making process. He chooses this forum to emphasize the social as well as technical issues in using tests for decision making. Dr. Page reviews factors that often influence perceptions of test quality, such as attacks on testing by the media, decisions with regard to test usage made by the courts, and concerns for test fairness and bias. Perception of test quality is identified as a fundamental factor in the use of tests for decision making. Unless tests are considered to provide valid, reliable, and reasonable pieces of information, he surmises, their role in making decisions will be subject to controversy and question. The chapter proceeds from a discussion of ways of establishing test quality and the reasons attitudes about the quality of tests may be threatened to a presentation of theoretical foundations for applying test results in the decision-making process. Page’s chapter therefore approaches the use of tests in decision making on two levels: initially, it must be demonstrated that the tests in question are in fact appropriate for use in a decision; second, a decision-making process should be employed to determine how the information provided by the test can be applied rationally to aid in making decisions.

Dr. Robert Sternberg presents an account of contributions of cognitive psychology to test development and usage in the following chapter titled “What Cognitive Psychology Can (and Cannot) Do for Test Development.” He contends that cognitive psychology stands to make substantial contributions to test development, although most of the contributions will be in the future. Sternberg discusses four topics: (1) what cognitive psychology is; (2) how cognitive psychologists study intelligence; (3) implications of cognitive psychological research for test validation; and (4) score interpretation and modification. Testing is presented in a reciprocal fashion whereby tests are used as assessment tools in

cognitive psychological research, the results of which can suggest modifications to test development and usage.

The next chapter presents fundamental and valuable information on the role and status of test validation research. In this chapter, Dr. Lyle Schoenfeldt reviews the history of test validation strategies, identifying methods of establishing test content, criterion-related, and construct validity. New advances in criterion-related validity, such as multivariate validation approaches, are presented and evaluated. In addition, validity generalization and Bayesian statistical approaches are discussed. The chapter presents recent advances and applications of test validation theory and research to the field of business (e.g., applicant selection and job satisfaction). Test validation is presented as an essential and legally necessary step in test usage. Some important and timely ramifications of not using tests with demonstrated validity are also discussed. Because the use of test results is only reasonable if the test is valid, this chapter presents the foundations upon which test usage relies.

“Social and Legal Influences on Test Development and Usage” is the title of the following chapter. After Schoenfeldt’s presentation of legal ramifications of inadequate test validation, Dr. Donald N. Bersoff posits three social influences that he regards as underlying all legal decisions pertaining to tests. These social influences are: (1) attempts to undo past injustices due to discrimination; (2) recognition of the public of their rights to privacy; and (3) negligence and lack of care by persons in positions to make decisions. Application of these social influences are illustrated in the fields of education, employment, and forensics. Bersoff continues his chapter with some examples of how social science research has and could be used to aid in court decisions on testing. He relates the impact of social influences and social science research to decisions in the cases of *Larry P v. Riles*, *PASE v. Hannon*, *Griggs v. Duke Power Co.*, and *Merriken v. Cressman*. The chapter concludes with a section on psychologists and public policy. Bersoff places the ultimate decision of test usage in the court’s hands, recognizing that the court’s decision will be influenced by the social and legal climate, which should be influenced further by test quality (validation) and expert psychometric testimony. He points out the final decision, however, is made by the judges, who are not generally psychometrically oriented.

Section I is concluded with a chapter from the Director of the Buros Institute of Mental Measurements, James V. Mitchell, Jr., which is titled, “Testing and the Oscar Buros Lament: From Knowledge to Implementation to Use.” Dr. Mitchell reviews the progress made in test development, using information accumulated from research and theoretical developments in testing knowledge. He reports that evidence of the status of test quality, as found in administration or technical manuals for tests, is often inadequate, and he contends test publishers are rewarded financially for test development by consumers who are, on the whole, psychometrically naive. If test sales are used as the guide, it appears that test users are, as a group, influenced by Madison-Avenue-type advertising and

tests' promises and titles, and they are not functioning as informed consumers. The responsibility for naive behavior of test consumers is traced to education and communication failures of professionals in the fields of testing and measurement. Dr. Mitchell concludes his chapter with specific recommendations that he believes will upgrade the education level of the consumers of tests that in turn will result in requiring test developers to upgrade the quality of their test documentation and development.

The perceptions of tests, especially as they are influenced by attacks on tests by persons in the media or courts who are in positions to make recommendations or decisions without adequate psychometric training, is one central theme that recurs in the chapters of Section I. The "call to action," issued by Mitchell, is reverberated in all the chapters of the first section. Improvement of test construction and test usage, viewed from utilization, theoretical support for and from cognitive psychology, test validation, legal and social influences, or quality control, require communication channels to the ultimate users of test results—the public.

Section II

Section II contains three chapters that originally were presented in the 1982 American Psychological Association's State of the Art Symposium. The symposium was organized by Dr. Carol Dwyer and focused on testing issues. The first chapter in Section II is authored by Dr. Anne Anastasi and is titled "Aptitude and Achievement Tests: The Curious Case of the Indestructible Strawperson." Dr. Anastasi initially reviews the traditional distinctions between aptitude and achievement testing, specifying that aptitude testing has been conceived as measuring "innate capacity" independent of learning, whereas achievement testing presumably assesses the effects of learning. The historical antecedents of this view are traced from Franzen's (1920, 1922) description of AQ (achievement quotient), the components of which were identified as EQ (educational quotient) and IQ (intelligence quotient). Dr. Anastasi then recounts efforts of psychometricians to disband the AQ terminology, beginning with Kelley (1927), noting that investigators repeatedly have reported extensive overlap of information obtained from these two types of tests. Yet despite the attempts by psychometricians to establish similarity between aptitude and achievement tests, the distinction reappears continually in presentations and writings of psychologists and psychometricians. Progress is being made though, as test companies recognize and communicate to the consumers that the distinction between aptitude and achievement tests is essentially one of breadth versus specificity of test content and antecedent learning experience. The conclusion of her chapter contains a more detailed analysis of the continuum of developed abilities, a continuum on which she places both aptitude and achievement tests. Thus, the major thrust of Dr. Anastasi's chapter is that psychologists and measurement experts have been

making steady progress in clarifying what aptitude and achievement tests measure; yet communication of this knowledge to test users, test takers, and the general public remains a problem.

Current issues in achievement testing is the topic of Dr. Robert Ebel's chapter, "Achievement Test Items: Current Issues." Dr. Ebel focuses his attention on the measurement of human characteristics and initially addresses the fundamental topic of the measurability of human characteristics. Within the domain of measuring human characteristics, Dr. Ebel considers the relative merits of various types of test items, such as: (1) essay and objective items; (2) realistic problem-solving items; and (3) alternate-choice items. Ebel concludes his chapter with a discussion of a technology of item writing. The major theme appears to be that any important human characteristic is necessarily measurable, and test items that focus on the basic components of knowledge are examples of an item-writing technology that has promise to yield highly reliable and valid assessments of human characteristics.

The final chapter in Section II, "Abilities and Knowledge in Educational Achievement Testing: The Assessment of Dynamic Cognitive Structures," is authored by Dr. Samuel Messick. The chapter begins by examining the question of what educational achievement tests are or ought to be. Both educational achievement and cognitive ability are viewed as constructs. The distinction between theoretical definitions and practical reality of assessment instruments is a major theme. Messick posits that educational achievement is a compound of developed abilities and knowledge structures. He then contrasts his view of what educational achievement tests are with that presented by Ebel, Anastasi, and others. Messick's conclusion is that theory, not empiricism, should guide the conceptualization and process of test development. He maintains that, to serve both theory and practice, new approaches to achievement measurement that are complex, dynamic, and cognitive need to be developed.

Each author in Section II conceptualizes aptitude and achievement testing differently. Anastasi elects to present aptitude and achievement testing on a single continuum, the distinction between them being one of specificity of a task and antecedents to the task. Ebel, on the other hand, considers aptitude as a special case of achievement and vice versa, establishing that intelligence, aptitudes, abilities, and achievements are synonymous. Messick believes the conceptual distinction between aptitude and achievement tests is flawed due to a reliance on empirical results obtained from using imperfect and variously contaminated tests. Thus, he discards the approach taken by Ebel, Anastasi, and others. His implication is that new approaches to appropriate measurement of aptitudes and achievement, which should be dynamic, cognitive, and complex, will enable a better assessment of what role cognitive abilities play or ought to play in educational achievement testing.

In summary, the authors in Section II focus on aptitude and achievement testing and debate social and technical issues pertaining to their application,

meaning, and usage. Dr. Anastasi points out that, despite attempts by psychometricians to defeat the distinctions popularly held by the public about aptitude and achievement tests, the “strawperson” remains indestructible and hence is an excellent example of social influence on test interpretation and usage. Technical issues that influence test construction and usage are central to Ebel’s and Messick’s chapters, with Ebel postulating the existence of an item-writing technology and Messick imploring test developers to use a theoretical, not empirical, basis for test construction.

CONCLUSIONS

The assessment of human abilities and qualities by tests has become an integral part of decision making in modern society. Nearly everyone has taken or will take a test that has the potential to influence his or her life significantly. The public is becoming more aware of and concerned about testing, test quality, and appropriate test usage. Testing and measurement cannot be treated in isolation. They are not immune from criticisms and influences from the very people their work affects most—society. To survive and thrive, measurement and testing must continue to develop through both improved technology and interactions with society.

REFERENCES

- Franzen, R. H. The accomplishment quotient, a school mark in terms of individual capacity. *Teachers College Record*, 1920, 21 432–440.
- Franzen, R. H. The accomplishment ratio: A treatment of the inherited determinants of disparity in school product. *Teachers College Contributions to Education*, 1922, No. 125.
- Kelley, T. L. *Interpretation of educational measurements*. Yonkers, N.Y.: World Book Co., 1927.

