

# Qualities that will influence testing and assessment practices with children and youth toward the beginning of the twenty-first century: international perspectives<sup>1</sup>

Thomas Oaldand

*University of Florida*

## Summary

Conditions that have impacted test development and use, as well as are those that are likely to influence continued test development and practice, are reviewed. Emphasis is placed on test use with children and youth in the international arena.

**Palavras-chave:** testes, avaliação, práticas de avaliação.

## Qualidades que influenciarão as práticas de teste e avaliação de crianças e jovens no começo do século XXI: perspectivas institucionais

## Resumo

São revistas as condições que tiveram impacto no desenvolvimento e uso, bem como aquelas que possivelmente continuarão a influenciar o desenvolvimento e a prática de uso dos testes. A ênfase recaiu no uso de teste com crianças e jovens no cenário internacional.

**Key words:** tests, evaluation, assessment practices.

Conditions that characterize testing and assessment activities toward the beginning of the twenty-first century will reflect, in part, extensions of conditions that historically have characterized *these* activities. Thus, a brief review of some historical events and trends as well as current and emerging qualities are needed before speculating as to conditions we may influence test development and use preeminently in the future.

The history of educational and psychological assessment is long and its development uneven. The birth of assessment in the behavioral sciences occurred in China at least 3.000 years ago. Measures of problem-solving, visual spatial perception, divergent thinking, creativity and others qualities that reflect important talents and behaviors were used somewhat commonly. Later, under the Sui dynasty (581-618), a civil service examination system consisting of three parts was initiated: regular examinations stressing classical cultural knowledge, a committee examination before the emperor stressing planning and administrative features, and a third examination on martial arts (Wang, 1993). Forms of this assessment system continued in China until 1905.

Following these initial efforts in China, subsequent developments in testing generally are unremarkable through the middle of the nineteenth century. Apparently few if any other countries initiated civil service examinations similar to those utilized in China. In addition, formal educational in most countries was limited to the sons of the wealthy and often provided by tutors,

---

<sup>1</sup> Some of the information utilized in this article was drawn from Oakland and Hambleton (1995). Readers are encouraged to consult this publication for a more extended discussion of these related issues.

thus minimizing the need for educational testing.

Various social and economic conditions occurred during the period from the 1870 to the 1920s within Western Europe and North America that were to significantly impact testing children and youth. Four critical needs emerged during this period for which public support often was provided: to educate more children at higher levels, to ensure quality control of education, to ensure children and others were cared for socially and were not abused, and to provide special services for those who exhibited severe disorders. Educators, psychologists, judges, social workers, politicians, and others increasingly utilized tests as an important source of information while addressing these needs.

During this same period the discipline of biology was evolving, spinning off the emerging science of psychology. Wundt and others in Germany and Galton in England developed a number of pioneering measures used largely for research purposes. Tests developed by Binet and other pioneers in France had a different goal, to classify children in need of special education services and to provide services to them. Their test development activities and those of others coincided with the emergence of the previously identified needs.

### **Traditional general purposes of test use with children and youth**

The success of these early efforts to develop tests provided confidence that reliable and valid measures of important qualities could be developed. Professional working in the behavioral sciences found tests and other methods to be useful for various purposes.

The uses of tests to facilitate research and to classify behaviors remain important. Over the years additional purposes have been added: the use of tests to describe current characteristics and attainment, screen for special needs, assist in guidance and counseling, diagnose disabling disorders, place students in special programs, evaluate progress, and determine whether students should be admitted, retained, or promoted. Tests also are used for various administrative and learning purposes.

### **Testing as a universal enterprise**

The amount of literature on test use with children and youth is considerable in the United States (e.g., Aiken, 1988; Anastasi, 1988; Cattell & Johnson, 1986; Cohen et al., 1988; Cronbach, 1984; Mitchell, 1985; Salvia & Yesseldyke, 1988; Sattler, 1988; Sweetland & Keyser, 1986; Reynolds & Kamphaus, 1990) and available from many other countries including Australia (Richie, 1989), China (LaVoie, 1989a, 1989b; Zhang, 1988a, 1988b, 1988c), Germany (Sehorr), Greece (Fatouros, 1984), Hungary (Klein, 1980), Israel (Raviv, 1989), The Netherlands (Evers & Zaal, 1982; Laak & Raad), the Orient (Oakland & Hu, 1989), United Kingdom (Farrell, *et al.*, 1989), Russia (Holowinsky, 1984-85; Pambookian & Holowinsky, 1987), Turkey (Oner, 1992); Kagiteibasi & Oney, 1992) and others (Aikens, 1987).

The growth of educational and psychological tests and other forms of measures has been somewhat startling. There are an estimated 5,000 standardized tests, mainly in English and constructed in the United States and Western Europe. Most of these measures are not used widely.

An international survey of tests used with children and youth in 44 countries, not including the United States, identified 455 tests used frequently (Oakland & Hu, 1989; 1991; 1992, Hu & Oakland, 1991). Some of the more important and general findings are summarized below.

Measures of intelligence, personality, and achievement are used most commonly. About 50 percent of the tests were developed within other countries and imported for use. Foreign developed tests tend to be used more commonly than locally developed tests.

Validity studies are available on between 50 and 70 percent of the tests and reliability estimates on 50 to 60 percent. Local norms are available on 80 percent of achievement tests and about 60 percent of intelligence and personality tests. At least 16 professional groups commonly administer tests; their levels of post-secondary education correlate strongly with their levels of competency in test use.

Two-third of the countries report an especially critical need for both group and individual tests of achievement, intelligence, vocational interests and aptitudes, social development, and personality, as well as more moderate needs for entrance measures for primary, secondary, and tertiary school, perceptual, and motor development. Respondents also indicated the classifications of children for which tests are needed. Virtually all of the responding countries reported the need for tests that assess qualities important for those who are mentally retarded, blind, deaf, learning disabled, slower learners, emotionally and socially disturbed, physically impaired, and gifted. The need for tests for the learned disabled is most critical.

As expected, test use is not uniform throughout the world. Highest test use occurs within highly industrialized nations and three pre-1990 social nations (Yugoslavia, principally Slovenia, East Germany, and Czechoslovakia). The dismantling of the Berlin Wall was announced on November 9, 1989 and is used to mark the beginning of this period. In contrast, lowest test utilization was reported by the least developed countries. Foreign developed tests are used overwhelmingly in the Middle East and least developed nations.

## **Some Conditions that impact test use**

Conditions both external to professionals and internal to them impact test development and use. External qualities include a country's social, political, religious, and economic conditions, qualities that strongly influence values, needs, resources and that in turn impact testing. Testing is stronger in countries that value technology and individual differences, utilize resources based on notions of meritocracy instead of egalitarianism, have well developed educational systems, and have the commercial and technological resources needed to produce and distribute tests. Countries lacking these qualities generally develop and use fewer tests.

The presence of these qualities will strongly impact the status of testing, particularly in those countries that have few test resources. Countries lacking these to test development and use. Thus, current data suggest continued under-development in many countries, perhaps through most of the next century.

In addition, various qualities internal to our disciplines and professions impact testing. Thus, we have considerable control over the following needs that, if achieved, will further test development and use: to broaden our disciplines and professions beyond their current narrow perspective that emphasize theory and research largely from Western European and North American; to prepare additional psychometricians and other scholars, particularly in developing nations, with advanced skills and abilities in testing and assessment, and to promote their stable

employment within institute that encourage and supports both teaching and research; to advance our knowledge and use of measurement models and methods; to develop and promote the use of suitable standards for test development and use within and across various cultures; and to develop assessment methods that accurately assess qualities that are both modifiable and enhance child growth and development.

Although space does not permit a discussion of each of these conditions (see Oakland, 1995 for a more extended discussion), some of the above points are briefly discussed below in an attempt to emphasize qualities that will impact the futures of testing.

### **External Conditions that need our attention**

The future of testing depends, in part, on positive attitudes toward testing held by the public. Professions whose work is dependent on and enhanced by test use should not remain neutral with respect to general conditions within countries that impact testing. They must be on the vanguard, creating conditions that help ensure tests are used in ways that serve important social goals and foster positive attitudes.

For example, we should work to ensure tests are not used to restrict educational or vocational opportunities, to rigidly shape school curricula, to invade one's personal privacy, to place children in categories that are dehumanizing and developmentally restrictive, and to denigrate the dignity and pride of classes of people - typically minority groups.

Furthermore, public, opposition to test use is legitimate when test practices are inappropriate. Inappropriate practices include failure to test persons in their dominant language, using invalid and poorly normed tests, and making inferences about children based on meager information. Test specialists must demonstrate that tests serve those who are tested, not merely those who test or the institutions that employ them.

In addition, professions must strongly oppose efforts that limit or restrict the legitimate uses of tests when these restrictions and limitations are unjustified and based on anti-testing attitudes. To this end professionals need to properly educate the public and their colleagues as to the strengths and limitations of test use and to address false allegations that improperly restrict test use (e.g., tests are biased against minority children, falsely label children, and create false and negative expectations regarding their futures).

The increased availability of computer based testing, while expanding the range and flexibility of test use, poses some potentially ominous implications for societies and professions that rely on test use. Large scale test use by industry and institutions, as well as self-testing, will be encouraged, with attendant problems associated with test misused and incorrect interpretations.

Professionals will be less able to monitor test development and use; professional and industry standards are unlikely to be followed.

In addition, quality tests are developed within a commercial climate in which those who conceive, develop, and market tests can be expected to bear related costs and receive related benefits.

Conditions that impede these essential commercial conditions, including the wide-spread photocopying and the other forms of copyright infringement (e.g. the free distribution of commercially

developed tests through computers), prevent the development of quality tests and the provision of proper professional services.

### **Internal Conditions that will impact the future of testing**

Internationally developed and approved standards for test development and use are needed. Within most countries, issues impacting test development, estimating reliability and validity, the natures of test norms, test user qualifications, and the commercial distribution of tests are governed more by convenience and tradition rather than professional standards. If test standards are used, those developed within the United States (American Educational Research Association, 1985) and the United Kingdom (British Psychological Society, 19XX) are used most commonly. However, these standards were developed to address needs in their countries rather than to meet the needs of a broad international community. Furthermore, most professionals in most countries are unaware of these standards; thus, testing practices generally are not guided by them. The International Test Commission recently approved a project that, if successful, will lead to the development of international standards that address these and other critical issues.

Test use within and across cultures can be expected to increase. Therefore, professionally developed and validated standards are needed to guide test translations, to equilibrate a test's cognitive qualities, and to help ensure similarity in reliability and validity estimates. The International Test Commission is coordinating efforts to prepare and disseminate validated technical standards for translating test and establishing test score equivalence. These efforts have resulted in providing guidelines needed when translating tests for use in different cultures while helping to ensure their equivalency (see Hambleton, 1995 for the preliminary guidelines).

As previously noted, in those countries with well developed testing resources, tests generally are used widely and effectively to facilitate research, classify behaviors, describe current characteristics and attainment, screen for special needs, assist in guidance- and counseling, diagnose disabling disorders, place students in special programs, evaluate progress, determine whether students should be admitted, retained, or promoted, and for various administrative and planning purposes. In addition, we are being asked to do more. We are being asked to rethink some of the basic premises that have guided test development and use. In short, we are being asked to design assessment methods that focus on learning and other important developmental processes, uncover hidden talents and abilities, and enhance learning and other forms of personal growth and development (Table 1).

<b>ASSESMENT STRATEGIES</b>	
<b>TRADITIONAL</b>	<b>POSSIBLY EMERGING</b>
<i>Assumptions about academic behavior</i>	
Behavior is stable	Behavior is dynamic
<i>Assumptions about teaching</i>	
Instruction precedes testing	testing precedes instruction
Teachers solely responsible for guiding instruction	Pupil, peers, parents and teachers responsible for guiding instruction
Promote ability to solve problems	Promote ability to demonstrate suitable problem-solving methods
Emphasize learned abilities	Emphasize learning abilities
Emphasize assessment of memorized knowledge	Emphasize assessment of higher order cognitive applications
Rely on external regulation of achievement	Promote self-regulation of achievement
<i>Assumptions about learning</i>	
Teachers and tests evaluate degree of attainment	Student, along with other sources, evaluate degree of attainment
Reward attainment at high levels	Reward attainment at or above potential
Reward convergent thinking	Promote divergent thinking
<i>Assumptions about the purposes of testing</i>	
Assess outcomes or products	Assess the process used to achieve the product
Evaluation focused on past and present	Evaluation focused on present and future
Test to inform professionals	Test to inform students
Diagnose and label permanent barriers that restrict attainment	Identify (without labeling) temporary and improvable barriers to attainment
<i>Assumptions about the testing process</i>	
Asses achievement separately in content areas	Assess achievement across content areas
Exclude assessment of feelings and personal attitudes	Include assessment of feelings and personal attitudes
Use paper and pencil methods	Use multi-sources and multi-methods to measure multi-traits
Test simulated outcomes	Test authentic outcomes
Score tests quantitatively	Score tests qualitatively
Establish external standards	Promote internal standards
Judge attainment in light of behavioral objectives	Judge attainment in light of developmental outcomes
Rely on norm-referenced standards	Rely on criterion-referenced standards
Emphasize summative evaluation	Emphasize formative evaluation

**Table 1.** Some differences between traditional an emerging trends in student assessment

### **Test use within the context of education: an international view**

Economic, political, and social changes occurring in many countries are likely to be on-going for years. Change may be most evident among those going from command to free enterprise economies as well as those in Sub-Sahara Africa and Latin America. Changes also are apparent among nations within the European Union. Less apparent although no less important are changes in

educational systems in these and many other countries. Virtually every country experiencing educational changes is reviewing how assessment methods may enhance the effectiveness and efficiency of its efforts. The importance of education internationally was underscored by the World Conference on Education for All that met in 1990 in Thailand. The conference delegates developed and adopted a World Declaration on Education for All (Interagency Commission, 1990). Issues pertaining to assessment were reviewed. Article 4 of the declaration states, "It is, therefore, necessary to define acceptable levels of learning acquisition for educational programs and to improve and apply systems of assessing learning achievement" (p.5). In formulating guidelines for implementing the Declaration, the conference delegates identified assessing needs and planning action as its first priority.

Countries increasingly realize investment in education is one of the principal methods by which they can become or remain competitive and improve the welfare of their citizens.

Educational change and test use increasingly are seen as inextricable, leading to increased recognition as to the value of test use for research and practice, developing methods to improve test use, within and between cultures, and designing assessment systems to support educational initiatives. Two examples are provided.

The United Nations, the World Bank, national governments, together with many private foundations and non-governmental organizations are providing financial support for projects in which test use is important. Research on and evaluations of interventions designed to promote infant and preschool development, educational attainment, social behaviors, and other child-focused efforts often require suitable tests and other measures.

The use of achievement and ability tests to assist in research and evaluation efforts within developing countries can serve five main goals: to monitor progress toward national goals, to evaluate the effectiveness and efficiency of specific educational policies, to hold schools and governmental agencies accountable for performance, to identify students for further education and to certify their competence, and to facilitate teacher assessment of student performance (Lockheed, 1991).

Second, the nature of international programs to enhance development within developing countries is being revised. Efforts to promote international development typically focused on raw materials, labor, and capital investment to improve a country's abilities to manufacture and in other ways produce products for local and foreign consumption. Current and future efforts increasingly will focus on ways to promote and utilize cognitive abilities and knowledge as these qualities replace raw materials, labor and capital as the basis for production and a quality lifestyle. Knowledge is viewed as forming the foundation for a country's vitality. Knowledge must be produced, conveyed to others, adapted, and used effectively. A country's success during the next century will be determined by its ability to acquire and produce knowledge, convey it to others, adapt and use it efficiently. Test use is directly tied to these efforts.

## References

- Aiken, L.R. (1987). *Assessment of intellectual functioning*. Newton, MA: Allyn & Bacon.  
Aiken, L.R. (1988). *Psychological testing and assessment* (6th ed). Boston, MA: Allyn & Bacon.  
American Educational Research Association (1985). *Standards for educational and psychological testing*.

- Washington, D.C: Author.
- Anastasi, A. (1988). *Psychological testing* (6th ed.). New York: Macmillan.
- Cattell, R.B. & Johnson, R.C. (Eds.) (1986). *Functional psychological testing: Principles and instruments*. New York: Brunner/Mazel.
- Cronbach, L.J. (1984). *Essentials of psychological testing* (4th ed.). New York: Harper & Row.
- Evers, A., & Zaal, J. (1982). Trends in test use in the Netherlands. *International Review of Applied Psychology*, 31, 35-53.
- Farrell, P., Dunning, T., & Foley, J. (1989). Methods used by educational psychologists to assess children with learning difficulties. *School Psychology International*, 10,47-55.
- Fatouros, M. (1984). Research studies of testing in Greece: A review. *International Review of Applied Psychology*, 33,351-370.
- Hambleton, R. (1994). Guidelines for adapting educational and psychological tests: A progress report. Bulletin of the International Test Commission. Published in the *European Journal of Psychological Assessment*, 10, 3, 229-244.
- Holowinsky, I.Z. (1984-5). Assessment of cognitive skills in the USSR: historical trends and current developments. *Journal Special Education*, 18, 541-545.
- Hu, S. & Oakland, T. (1991). Global and regional perspectives on testing children and youth: An empirical study. *International Journal of Psychology*, 26, 329-244.
- Kagiticbasi, C. & Oney, B. (1992). *Evaluation of early development instruments in Turkey*, Bekek-Instanbul: Bogazici University.
- Klein, S. (1980, December). Uses of test methods in Hungary. *Newsletter of the International Test Commission*.
- Laak, J.T. & Road, B.D. (No date) *Psychological assessment In the Netherlands: Teaching, practice and research*. Utrecht, The Netherlands: Department of Developmental Psychology, University of Utrecht (unpublished paper).
- LaVoie, J.C., (1989a). School psychology research in the People's Republic of China. *Professional School Psychology*,4,137-145.
- LaVoie, J.C. (1989b). School psychology in the People's Republic of China. In P. Saigh & T. Oakland (Eds.) *International perspectives on psychology in the schools* (p. 165-175), Hillsdale, NJ: Lawrence Erlbaum Associates.
- Lockheed, M. (1992). *World Bank support for capacity building: The challenge of educational assesment*. Washington D.C.: The World Bank.
- Lockheed, M. & Verspoor, A. (1991). *Improving primary education in developing countries*. Oxford, England: Oxford University Press.
- Mitchell, J.V. (Ed.) (1985). *The ninth mental measurements yearbook, Vol. 2*. Lincoln: Buros Institute of Mental Measurement.
- Oakland, T., & Hu, S. (1991). Professionals who administer tests with children and youth: Ao international survey. *Journal of Psychoeducational Assessment*, 9 (2),108-120.
- Oakland, T. & Hu, S. (1992). The top 10 tests used with children and youth worldwide. *Bulletin of the International Test Commission* 19,99-120.
- Oakland, T. & Hu, S. (1989). Psychoeducational services for children and youth in the Orient: Current status, problems, and some remedies. *International Journal of Psychology*, 24, 237-250.
- Oakland, T. & Hu, S. (1994) International perspective on tests used with children and youth. *Journal of School Psychology*.
- Oakland, T. & Hambleton, R. (Eds.) (1995). *International perspectives on academic assessment*. Hingham, MA: Kluwer Academic Publishers.
- Oakland, T. Test use with children and youth: current status and future directions. In T. Oakland & R. Hambleton. (Eds.) (1995). *International perspective on academic assessment*. Hingham, MA: Kluwer Academic Publishers.
- Onir, N. (1992). *Toward a source book on psychological tests in Turkey*. Paper presented at the annual meeting of the Internal School Psychology Association, Istanbul, Turkey.
- Pambookian, H.S., & Rolowinsky, I.Z. (1987). School psychology in the USSR *Journal of School Psychology*, 25, 209-221.
- Raviv, A. (1989). School psychology research in Israel. *Professional School Psychology*, 4 (2) 147-154.
- Reynolds, C. & Kamphasu, R. (1990). *Handbook of psychological and educational assessment of children*.

- New York: Guilford Press.
- Ritchie, M.R. (1989). School psychology research in Australia. *Professional School Psychology*, 4, 129-135.
- Salvia, J. & Yesseldyke, J.E. (1988). *Assessment in special and remedial education* (4 th ed.). Boston, MA: Houghton Mifflin.
- Sattler, J. (1988). *Assessment of children* (3rd ed.). San Diego, CA: Jerome M. Sattler.
- Schorr, A. (1991). *Stand und perspektiven psychologischer diagnoslik in der praxis*. Unpublished paper.
- Sweetland, R.C. & Keyser, D.J. (Eds.) (1986). *Tests: A comprehensive reference for assessment in psychology, education, and business* (2nd ed.). Kansas City: Test Corporation of America.
- Vassof, G. (1982). Mental massacre. *School Psychology Intemational*, 3, 43-48.
- Wang, Z-M (1993). Psychology in China: A review. *Annual review of psychology*, 44, 87-116.
- Zhang, H. (1988a). The contribution of psychology to education in China. *School Psychology International*, 9, 3-11.
- Zhang, H. (1988b). Psychological measurements in China. *International Journal of Psychology*, 23, 101-117.
- Zhang, H. (1988c, October). Psychological testing and China's modernization. *Bulletin of the International Test Commission*, 27, 23-31.
- Zhang, H. (1992). Problems in the application of tests from overseas. *Bulletin of the International Test Commission*, 19, 173-180.
- 

#### Observation

Persons interested in tracking educational changes in European nations are encouraged to consult the *Education Newsletter Faits Nouveau*. The Newsletter can be obtained from the Council of Europe, F67075, Strausbourg France CEDEX.