

**Bibliothèque
et Archives
nationales**

Québec



Le présent fichier est une publication en ligne reçue en dépôt légal, convertie en format PDF et archivée par Bibliothèque et Archives nationales du Québec. L'information contenue dans le fichier peut donc être périmée et certains liens externes peuvent être inactifs.

Version visionnée sur le site Internet d'origine le 12 août 2013

Section du dépôt légal

MJE

The McGill Journal of Education

Special Issue

International Perspectives on School-Based Assessment

Philip A. Saigh, Guest Editor

L.B. Birch, Jean-Marc Chevrier,
Jack A. Cummings, Henry L. Janzen,
Marc C. Laquerre, Joseph C. LaVoie,
Donald H. Saklofske, Jerome M. Sattler,
Robert L. Thorndike, Moshe Zeidner

Winter 1990 Vol. 25 No. 1

Editorial

A Special Issue on School-Based Assessment

This issue of the *Journal* provides a description of the research in five nations, Canada, China, Israel, Lebanon, and the United States, relative to the many dimensions of school-based assessment. It is of interest to know that these countries share some common concerns, issues, and directions of development in this area, and yet at the same time all the countries have their own particular domains of research and development that emanate from their historical, political, and cultural roots.

The research reported here is by no means exhaustive, as several authors note, but it does give much insight as to the commonalities we share and the unique situations we face in school-based assessment as we move toward the year 2000. To comment further on a topic that has been, in the following pages, superbly examined would be redundant.

The editor and staff of the *McGill Journal of Education* extend their gratitude to Dr. Philip A. Saigh, of the Graduate Center of the City University of New York, for his unfailing attention to guest-editing this issue. To coordinate the writings of researchers in five different nations was no small task, and he is to be congratulated for bringing to completion an edition of the *Journal* that will, no doubt, be stimulating and useful to all who are interested in "International Perspectives on School-Based Assessment."

WMT

Guest Editorial

International Perspectives on School-Based Assessment Research

In recent years, information relating to international school psychology practices (Saigh & Oakland, 1989) and research (Saigh, 1989) has become increasingly apparent. Although these reports provide information regarding practice or current research in a number of countries, the extant literature has not focused on school-based assessment research as it relates to intelligence, achievement, and personality testing. In view of this and as an understanding of international research may facilitate new avenues of inquiry, this issue of the *McGill Journal of Education* is intended to provide readers with a selected overview of research developments in Canada, China, Israel, Lebanon, and the United States.

Initially, Donald Saklofske, of the University of Saskatchewan, and Henry Janzen, of the University of Alberta, provided a broad overview of school-based assessment research in Canada. The second article by Joseph LaVoie, who recently completed a Fulbright scholarship in China, provides a summary of research developments in the world's most populous country. The third paper, by Moshe Zeidner of the University of Haifa, addresses the current status and future directions of school-based assessment research in Israel. Philip Saigh, who served as a professor at the American University of Beirut from 1977 through 1986, provides a description of Lebanese research in the fourth article. The final article by Jack Cummings and Marc Laquerre, of the University of Indiana, describes psychodiagnostic assessment research in the United States from a sociocultural perspective.

Two commentaries (the first by Robert Thorndike and the second by Jerome Sattler and Philip Saigh) are also provided. Finally, L.B. Birch, a close professional acquaintance of Sir Cyril Burt, presents a review of Robert B. Joynson's book, *The Burt Affair*, and Jean-Marc Chevrier, of the Institut

de Recherches psychologiques, Inc. (Montréal) discusses the recent French intelligence test, *Épreuve individuelle d'habilité mentale*, developed in Quebec.

Examined *in toto*, it must be acknowledged that the aforementioned articles represent a selected review of the literature and that the contributors were compelled to restrict the length of their manuscripts due to space limitations. It is hoped, however, that this issue will provide readers with an introduction to international developments in the area of school-based assessment research. It is also hoped that this knowledge will facilitate the scholarly exchange of information and contribute to the progress of school-based assessment research.

Philip A. Saigh

Guest Editor

REFERENCES

- Saigh, A.P. (1989). School psychology research in Lebanon: A retrospective analysis and a look ahead. *Professional School Psychology, 4*, 201-209.
- Saigh, P.A., & Oakland, T. (Eds.). (1989). *International perspectives on psychology in the schools*. Hillsdale, NJ: Lawrence Erlbaum Associates.

D.H. Saklofske
University of Saskatchewan

H.L. Janzen
University of Alberta

School-Based Assessment Research in Canada

Abstract

School-based assessment practices in Canada share many similarities with the United States. The place of testing in the larger assessment process is still seen as very important and school psychologists are frequently requested to administer intelligence, personality, and achievement measures to school children. This has resulted in using tests developed in other countries and, at times, modifying or standardizing them for use here or developing Canadian tests. Teacher-made tests are most preferred by teachers for the assessment of student achievement. Emerging trends in education will require some changes in the way assessments are conducted and an increased recognition of the importance of cognitive processes being assessed by educational personnel.

Résumé

Les méthodes d'évaluation en milieu scolaire au Canada ont de nombreux points en commun avec celles des États-Unis. L'importance des tests dans le processus d'évaluation global est toujours perçue comme essentielle et les psychologues scolaires sont fréquemment tenus d'administrer aux écoliers des tests d'intelligence, de personnalité et de niveau. Cela explique qu'on se serve de tests mis au point dans d'autres pays et qu'on les modifie ou qu'on les normalise parfois pour le Canada ou que l'on conçoive des tests résolument canadiens. Les tests conçus par des professeurs ont la faveur des enseignants pour le contrôle des progrès des élèves. Les orientations futures de l'éducation nécessiteront d'apporter certains changements à la conduite des évaluations et de faire reconnaître au personnel scolaire l'importance des processus cognitifs qu'ils évaluent.

The task of describing school-based assessment practices in Canada is a rather complex one because of geographical size, social and cultural diversity, and the fact that educational matters are the jurisdiction of the provinces. Teachers are most involved in the ongoing measurement and evaluation of student learning through the administration of teacher-made tests (criterion-referenced and curriculum-based) and norm-referenced standardized achievement tests. The more specialized diagnostic assessments requested by classroom teachers may be performed by various specially trained professionals such as school psychologists, guidance counsellors, itinerant teachers, school social workers, and speech therapists. Dumont (1987) observed that the job descriptions for school psychologists obtained from "the various ministries of education in Canada always include assessment and diagnosis" (p. 106).

The purposes and methods of educational and psychological assessment in Canadian schools share a great many similarities to practices in the United States. So far, Canada has remained relatively free from the kinds of landmark court cases heard in the United States regarding the assessment and school program placement of minority children (e.g., *Larry P. v. Riles*, 1979; *Marshall v. Georgia*, 1984). Both countries endorse the right of access to education for all children. However, while the United States has enacted some comprehensive pieces of legislation that are intended to ensure appropriate educational services to all children with special needs (e.g., PL 94-142, PL 99-457), Canada has no similar national legislation of this kind that would have an impact on diagnostic and program services. *The Canadian Charter of Rights and Freedoms* (1982) certainly has some potential in this regard (Dickinson & MacKay, 1989; Fox, 1988; Kimmins, Hunter, & Mackay, 1985). As well, Canada has signed both the Universal Declaration of Human Rights and the Declaration of the Rights of the Child which support the right to appropriate education. Ray (1989) concluded that "Canadians may assert with confidence that their education deals more effectively with many rights than was customary, and it receives high marks in international reviews. . . and domestic opinion" (p. 155).

Assessment Methods

Canadian educational personnel employ multiple assessment approaches to aid in describing and understanding children in order to provide the best and most appropriate educational services. The "four pillars of assessment" discussed by Sattler (1988), including informal assessments, interviews, observations, and norm-referenced tests, form the basis for collecting the kind and amount of information necessary for defining and evaluating the psychological and educational needs of children. The techniques and tools that comprise these four assessment categories are further cast within a frame-

work that recognizes the necessity of measuring and evaluating a whole complex range of human factors and external conditions. In keeping with the theme of this journal issue, the following sections will briefly overview both nontest and test-based methods employed in the assessment of intelligence, personality, and achievement as they are generally practised in Canada.

Nontest Assessment

School personnel have a wide choice of tests that may be employed in the assessment of childrens' needs yet much of the valuable data leading to hypothesis generation or decision making are based on nontest methods. Indeed, such methods as interviews, task analysis, work samples, and observations frequently comprise not only the initial stages of the assessment process but also are utilized throughout it. While classroom teachers are continuously engaged in ongoing evaluations of students, and therefore routinely employ all of these methods, school psychologists usually become involved with a student once a referral/request is made for their professional services. The initial data collection should most likely utilize interviews, observations, and record reviews that will shed further light on the reason for referral and provide the kinds of background information that will help guide the assessment process.

Information focusing on home behaviours, the family system, social and developmental history, classroom or instructional factors, and student perceptions of the "problem" are necessary in formulating hypotheses relevant to determining the nature of the problem and what further actions may be required. For example, Thompson (1985) reported that home environment factors (e.g., parental behaviours, home literacy, educational ambition, socio-economic efficiency) are important determiners and predictors of educational performance and that this effect increases with age. Cummins (1982) has further shown that the length of residency and age of arrival for children born outside Canada can lower scores on certain cognitive and language-based tests during the first five years of enrollment in an English-language curriculum.

Interviews are important in the construction of a model of the student which is the central purpose of assessment. Although much has been written about interviews, there is some indication that we are still concerned about whether we can trust what we see, hear, and feel. It is still most common to read test-based journal articles written on student assessment by school psychologists. While the assessment reports generated by school-based specialists may be heavily weighted by test scores, most will also contain information gleaned from either structured or unstructured interviews with

the student. Although different interview techniques may be selected, the need to "fit" the interview to the client is particularly highlighted in cross-cultural assessment, and Massey (1988) has outlined a useful set of guidelines for conducting such interviews.

While assessment is sometimes viewed as being synonymous with testing, it is much more than the administration of tests. During these days of fear of litigation, diagnosticians rely heavily on the safety of tests and test scores. Furthermore, school systems frequently request test data for purposes of designation and funding for special education. However, blindly interpreting such data without the full use of background, interview and self-report information, and current observations is not professionally defensible. Systematic behavioural observations are required to support and supplement obtained test data, perhaps even enhance it. Observational recordings may be used to describe both test and nontest behaviours, support the validity of test scores, aid in explaining test score variance, and add important information necessary to formulate a particular diagnosis and program recommendation.

Observation permits the opportunity to gather data of relevance to understanding and describing the child in school. Observation has an advantage in monitoring behaviours that change rapidly, exhibit particular growth or performance curves, and may not be typical or frequent. Of particular importance in the assessment process is the observation of children in their everyday environments such as the classroom. It is difficult to create tests that are both robust and sensitive enough to yield information about the frequency, intensity, duration, and situational specificity of a wide range of psychological constructs and human behaviours. For example, careful observation and recording procedures may aid in determining whether a child's poor performance in school is related to inability, lack of prerequisite skills, unwillingness, anxiety, or inappropriate instructional techniques.

Observation, like an interview, is both a technique and a skill. Knowing how to systematically observe is as important as knowing what to observe. Various observation schedules in the form of checklists and rating scales that focus on physical variables, affect, activity level, speech and language, adaptive behaviours, and social skills are commercially available but may also be created by teachers and psychologists as required. Two Canadian studies may be cited in this latter regard. Whyte (1984) developed a student self-report inventory that differentiated between learning-disabled and regular achieving adolescents on the basis of visual-spatial development, language and communication skills, memory and sequencing, social and emotional adjustment, and attentional skills. Whyte commented that the use of these techniques may permit the identification of learning-disabled stu-

dents and “make the more expensive and time consuming psychological testing more effective and efficient” (p. 24). At the other end of the educational ladder, Simner (1987) constructed a five-item questionnaire for identifying preschool children who may be at risk for early school failure. *The Teacher School Readiness Inventory (TSRI)* requires teachers to rate their observations of student in-class distractibility/attention, verbal fluency, interest and participation, letter identification, and printing skills. Results from Simner’s study suggest that the TSRI had high interrater reliability, was significantly related to performance across the curriculum, and had a “hit-rate” of 86% in identifying children who were experiencing academic problems in the first year of school.

Educational and Psychological Testing

The history and growth of psychological and educational test use in Canadian schools is fairly similar to the American experience. However, the disproportionate population size between the two countries has resulted in it sometimes being more economical to “import” tests developed and standardized in the United States and other countries for use in Canadian school systems rather than attempt to create them here. This situation can sometimes present rather major problems in the assessment process especially when norm-referenced tests are employed or the product being measured is tied to specific and unique experiences as in the case of curriculum-based assessment (Janzen & Saklofske, 1990). Thus, it is not always such a straight forward matter of simply using well constructed American tests in Canada. Since tests and measurements are such essential tools in assessment, some American instruments that are brought into Canada may be renormed or modified following the accumulation of data from research and clinical use. As an alternative, a number of tests have been created in Canada, especially in those areas that involve school achievement that is tied to local and provincial curriculums.

Assessment of Intelligence

The measurement of children’s intelligence is considered to be an important component in a comprehensive and multilevel assessment program. Despite the ongoing controversy surrounding the use of intelligence tests and Mercer’s (1988) contention that the intelligence testing debate is “evidence of growing intellectual crisis in measurement psychology that may be the prelude to a scientific revolution in that field” (p. 1), these are the most frequently administered of all individual tests requested of school psychologists. Many school boards also routinely administer group intelligence tests. IQ scores may be compared with achievement data to determine the extent of

student underachievement and learning difficulty, often with the aid of "mathematically-based formulae" (McLeod, 1988). Intelligence is viewed as an important variable in an analysis of school achievement and the findings from Canadian studies (e.g., Hardman & Oldridge, 1985; Janzen *et al.*, 1989 a,b) of the predictive validity of tests such as the *Weschler Intelligence Scale for Children-Revised* (WISC-R) and the *Stanford-Binet Intelligence Scale: Fourth Edition* (SBIS) are in line with American results.

Wilson and Humphries (1986) provide a not uncommon description, based on events in Ontario, of the demand for intellectual assessment as a major data source used for the designation and placement of students in special classes; "there is widespread dependence in school boards on norm-based instruments of which the most frequently used is the WISC-R" (p. 3). This widespread and frequent use of the WISC-R in assessment has generated a fairly large amount of research, in part, because of the controversy that Canadian children may perform differently from the American standardization sample and that some parts of the test may be biased against Canadian children. Since Vernon (1977) recommended the substitution of certain Canadian items on the Information subtest, a number of studies have been conducted to determine the efficacy of such changes (e.g., Cyr & Atkinson, 1987; Marx, 1984; Peters, 1976) culminating most recently in Beal's (1988) assertion that there is no solid evidence for changing the item content on the WISC-R. While national norms have not been developed for the WISC-R in Canada, some local or provincial norms do exist (e.g., Holmes, 1981). It is generally agreed that the American norms for this test are appropriate for use here although Canadian children, on average, score slightly above the subtest and IQ mean scale scores. This is partly due to the fact that the test was standardized 15 years ago. The WISC-R will soon be replaced by a new version which is concurrently being validated in Canada (Beal, 1989).

The WISC-R has also been employed in validity and clinical studies of exceptional children including retarded, learning disabled, and gifted students (see Janzen & Saklofske, 1990). Some of this research parallels that carried out in the United States; for example, WISC-R profile and pattern analysis related to achievement and learning disabilities (Bellemare, Inglis, & Lawson, 1986; Schmidt, Kuryliw, Saklofske, & Yackulic, 1989; Schmidt & Saklofske, 1983; Walsh, Marx, & Sudmant, 1983). A major concern relates to the use of this and other intelligence tests with culturally different groups of children, especially with Canadian aboriginal children. There is a divergence of opinion about the continued use of the WISC-R in the assessment of these children yet a fairly consistent pattern of average performance scores but below-average verbal scores has been reported by several researchers (Common & Frost, 1988). As well, a recent study of aboriginal children

living in villages and cities in British Columbia found that experiential background, not ethnicity, was a major determinant of information processing ability and the development of reading skills (Williams, 1988). These kinds of findings have resulted in school psychologists developing guidelines for the assessment of Canadian aboriginal children, including the examiner's responsibilities (Goldstein, 1988).

It appears that Canadian school psychologists and educators will continue to rely on individual intelligence tests developed and standardized in the United States. While intelligence is a universal construct, its measurement and phenotypic expression can vary as a function of culture, experience, language, and so forth unless one chooses to measure more basic biological processes such as reaction time or evoked potentials. This is clearly realized in a multicultural country like Canada where research investigations of new tests with the potential for widespread use are usually initiated fairly quickly following their introduction. Such is the case, for example, with the *Stanford-Binet Intelligence Scale-Fourth Edition* (Janzen et al., 1989 a,b; Kline, 1989), *Kaufman Assessment Battery for Children* (Gardner, 1986, 1988; Saklofske & Jedlicki, 1985), *Matrix Analogies Test-Short Form* (Saklofske & Murray, 1989), and *Draw A Person: A Quantitative Scoring System* (Saklofske & Braun, 1989).

It should be noted that several intelligence tests have been specifically developed for use in Canada although these are mainly group tests that are most often administered by classroom teachers as part of a school district or even province-wide testing program. *The Canadian Cognitive Abilities Test* (CCAT) is one of the most often used group tests of intelligence and was recently standardized along with the *Canadian Test of Basic Skills* (CTBS). The CCAT covers the K-12 grade range and yields measures of verbal, quantitative, and nonverbal ability. While not aimed specifically for the Canadian context, the work of Dr. J.P. Das of the University of Alberta (see Das & Naglieri, 1990) in developing tests for the assessment of attention, processing, and planning abilities must certainly be recognized.

Finally, it is interesting to note a contemporary study (Paul & MacLeave, 1989) that examined the perceptions of educators about intelligence and intelligence testing. The study was conducted within the Gander Bonavista-Connaigre Roman Catholic school district. Caution is necessary in generalizing to other areas of Canada. The reported findings included: (1) intelligence testing and the use made of these data were viewed more positively by educators who were high versus low "hereditarians" and by educators with university training beyond a Bachelor's degree in contrast to those only with an undergraduate degree, (2) secondary school teachers compared with ele-

mentary teachers supported the increased use of test scores, and (3) educators across all positions surveyed "appeared to be satisfied with current testing practices" (p. 380). These results and also the knowledge that intelligence tests are so often used in the psychoeducational assessment of children suggest that their place is assured for the foreseeable future.

Personality Assessment

The assessment of school children by psychologists and counsellors may include various personality measures depending on the nature of the referral and the decisions to be made. The personality characteristics of relevance to school personnel in understanding and providing for a child's personal and educational needs are the same as those of general interest in child and adolescent clinical psychology. Thus, as part of a diagnostic assessment, school psychologists may administer one or more comprehensive objective personality questionnaires (e.g., *Children's Personality Questionnaire*; *Junior Eysenck Personality Questionnaire*), projective techniques (e.g., *Children's Apperception Test*, *Kinetic School Drawings*, *Kahn Test of Symbol Arrangement*) or briefer instruments that measure very specific personality variables (e.g., *Children's Depression Scale*, *Nowicki-Stickland Locus of Control Scale for Children*).

A glimpse of Canadian educational and psychological journals (e.g., *The Alberta Journal of Educational Research*, *Canadian Journal of Behavioural Science*, *Canadian Journal of School Psychology*, *Canadian Journal of Special Education*) gives an indication of the investigations undertaken to examine personality factors in school children and how they may be related to school performance. Byrne's (1986) investigation of the self-concept of high school students replicated previous findings regarding the multidimensional, hierarchical structure and stability of self-concept as well as its relationship with academic achievement. Studies of learning disabled (LD) children have shown that they can be distinguished from non-LD students through an analysis of the graphic elements contained in the *Kinetic School Drawing Test* (Andrews & Janzen, 1988). Rogers and Saklofske (1985) found that LD students had lower self-concepts, a more external locus of control, and lower performance expectations than normally achieving elementary school children. Furthermore, measures of general and academic locus of control and academic self-concept contributed significantly in predicting the extent to which LD children were rated by teachers as experiencing academic success in their special education programs.

Turning to studies of attitudes, Nyberg and Clarke (1982) reported that the responses of grade 5 and 8 students on the *School Subjects Attitude Scales*

were related to teachers' perceptions of student attitudes, student choice of most and least liked subject, type of school program (e.g., academic), and cultural and sex differences. Following the suggestion that attitudinal and personality variables are important in the reading process, Summers (1980) has provided evidence for the validity of the *Estes Reading Attitude Scale* as a measure of school-oriented attitude toward reading in the intermediate grades. Furthermore, attitudes toward reading were not changed by the opportunity for sustained silent reading (Summers & McLelland, 1982).

In another area, self ratings of social behaviour by secondary school students were observed to be related to school achievement and gender (Loranger, Poirier, & Gauthier, 1983). As a final example, Rampaul, Singh, and Didyk (1984) reported significant positive correlations between self-concept, academic achievement, and teacher expectations among grade 3 and 4 Canadian aboriginal students in northern Manitoba. However, the authors observed "academic retardation with increasing age" among this sample of students (p. 213). These and other studies of personality in school children provide the school psychologist with a research base that will enhance the assessment process. Few school psychologists routinely administer selected personality tests to all referred children but rather will determine what to measure and what test(s) to employ based on the individual child's presenting needs. This may be suggested by the teacher's referral of a child for assessment; the child "appears nervous and anxious and doesn't participate in class," "has a low self-concept, difficult to motivate," or "has problems staying on task, is somewhat aggressive." Of course, school psychologists will continuously be "on the lookout" for personality factors that may support or alter their hypotheses about a child and guide data collection and diagnostic interpretation. Massey (1988, p. 29) and other Canadian researchers (e.g., Berry, 1975) have stated that attitudinal, motivational, and "other personality characteristics such as the amount of eye-contact, reaction to physical proximity, male-female interaction and lack of spontaneous speech also affect the interpretation of results" in the assessment of culturally different children.

It is very rare that personality-type questionnaires would be routinely administered to large numbers of children as part of a screening assessment program, as occurs with group intelligence and achievement tests, unless they are part of a more research-oriented project. An example of such a situation may be found in two studies of depression in Canadian elementary school children conducted by Paananen and Janzen (1986) and Saklofske, Janzen, and Paananen (1987). In order to determine the incidence and characteristics of depression among grade 4-6 children, over 600 children completed measures of self-esteem, locus of control, and depression; achievement and ability data were obtained from cumulative records. These results suggested, among

other things, that less than 10% of these children suffer from a depressive syndrome, depressed children have lower self-esteem and an external locus of control, and that the *Children's Depression Inventory* and *Children's Depression Scale* provided a similar index of depression.

The greatest number of personality measures used in Canada are most often American in origin and range from comprehensive, multifactor standardized tests to brief scales published in various journals. In recent years, there has been some encouragement by test developers to simultaneously standardize new instruments in both Canada and the United States. For example, a new battery called the *Behaviour Assessment System for Children* (BASC), authored by Kamphaus and Reynolds and to be published in 1990 by the American Guidance Service, is intended to provide a psychometrically sound approach to assessing the emotional and behavioural problems of children and adolescents. The authors advertised for Canadian sites so that appropriate norms may be developed. In this regard, Saklofske, Yackulic, and Kowalchuk from the University of Saskatchewan have recently completed the data collection phase for a provincial standardization study of the BASC with school children in grades 2-8 and a second study of conduct disordered boys with the expectation that this multisource measure (child self-report personality profile, parent rating scale, teacher rating scale, parent personality profile, classroom observation scale, developmental questionnaire) will prove especially useful in the assessment process.

In the event that personality tests from other countries are employed in the assessment of Canadian school children, it may be necessary to renorm or even modify the instrument. The *Junior Eysenck Personality Questionnaire*, developed in England, may be administered to Canadian children without scoring changes in the scales but the Canadian data show higher mean scale scores for the neuroticism and tough-mindedness scales, lower scores on social desirability, and almost no difference in extraversion (Eysenck & Saklofske, 1983). In contrast, the *Junior Impulsiveness Inventory*, also from England, did require some item changes when introduced in Canada (Saklofske & Eysenck, 1983).

Finally, it is important to note that various children's personality measures have been developed and standardized in Canada. Alberta has been the venue of some of this work and includes: the *Frost Self-Description Questionnaire: Extended Scales* (Frost, 1979) which measures various forms of anxiety and aggression; the *Canadian Self-Esteem Inventory for Children* (Battle, 1976, 1979); the *Culture-Free Self-Esteem Inventory for Children* (Battle, 1981); and the *Student's Perception of Ability Scale* (Boersma & Chapman, 1985) which is now being revised. This latter instrument follows

from research that has shown a strong relationship between self-perceptions of ability and school achievement.

Assessment of Achievement

The assessment of student learning and achievement can be initiated at different levels (classroom quiz to board and province-wide testing programs) and in various different ways (teacher-made tests of specific instructional content, standardized group achievement tests, diagnostic tests, homework assignments, project reports, etc.). Educational specialists are less involved than other staff in the everyday assessment of school achievement, but they are consulted by the teacher when additional information and possible program changes may be required for a student.

In this area of assessment, tests are much less portable from country to country and even across regions within a country like Canada. School performance and the learning outcomes of main importance for a given grade or subject area are most specific to the defined curriculum and objectives of school districts and provinces and may further vary to some extent across teachers and schools. Thus, very few of the comprehensive standardized achievement test batteries used in American schools are employed in Canada. Some briefer instruments and diagnostic tests prepared outside of Canada are useful as screening devices or for providing an analysis of particular learning problems. The *Wide Range Achievement Test* (WRAT) and the most recently revised version are sometimes used to obtain a general indication of reading, arithmetic, and spelling ability. The WRAT has not been normed in Canada and there is some difference of opinion as to the utility of this test for use here (Siegal, 1984; Snart, Dennis, & Brailsford, 1983). Another example of this kind is reflected in efforts to develop guidelines for the effective use of non-Canadian developed instruments such as the *Boder Test of Reading-Spelling* for screening decoding and encoding problems in French immersion students (Wiss & Burnett, 1988).

Several carefully standardized individual and group achievement tests have been specifically constructed for use in Canadian schools. The *Canadian Achievement Tests* (CAT) and the *Canadian Tests of Basic Skills* (CTBS) are two instruments frequently used by school systems as a broad achievement measure across all school grades. For example, the CTBS measures reading, language, work-study, and mathematics skills. Both tests were normed on large numbers of Canadian school students which permits the comparison of obtained test scores with local, provincial, and national results. The *British Columbia Quick Individual Educational Test*, yielding measures of spelling, word identification, passage comprehension, and arithmetic, is based on the

British Columbia curriculum and was normed in that province (Wormeli, 1983). McDonald (1985) concluded that this test is effective in the identification of students with learning difficulties. From a different perspective, the *Canadian Readiness Test* is a battery of seven subscales measuring four general areas of prereading behaviour and has been found to have predictive validity for reading and mathematics achievement (Collis, Ollila, & Yore, 1986; Evanechko, Ollila, Downing, & Braun, 1973). As well, a large number of achievement tests are constructed by provincial education departments and local school boards as a means of assessing learning outcomes based upon specific objectives.

Classroom teachers base their evaluations of student learning less often on standardized testing programs alone, but rather employ a variety of assessment methods and procedures. Teacher-made tests appear to be the most often used source of achievement data for student evaluation. Anderson (1989) reported that teacher-made tests were viewed by British Columbia science teachers as the most important information source with the least emphasis being given to standardized objective tests and other techniques such as oral tests, work contracts, and student self-reports. In an earlier study, Wahlstrom and Danley (1976) reported that Ontario elementary teachers prefer classroom observations of student performance but secondary teachers favour tests.

While classroom teachers are more inclined to use their own tests to measure student achievement, a number of Canadian educational researchers have expressed concern about the "gap" between test theory and research in the measurement and evaluation of student performance on the one hand, and the evaluation practices of teachers on the other (McLean, 1985). Wilson (1989) concluded that "it is still unclear whether teachers view evaluation activities in a coherent, holistic manner" (p. 143) and suggested two possible reasons for this situation:

The ready-made instruments may perform one part of the teacher's purpose for evaluation but not another. . . similarly, the professional's concern with reliability of scoring, item analysis, and systematic content converge may also seem of only esoteric interest, and even an imposition, to teachers practising the delicate balancing acts required by the different, sometimes contrary, forces operating on them already. (p. 143)

This separation of measurement theory and research from actual classroom evaluation practices has led to the recommendation that "applied edu-

cational measurement should center its attention on the classroom" (Anderson, 1989, p. 133). Evidence of a positive response to this plea is now being seen across Canada. To cite only a few examples of this new direction, McLean (1988) has described how functional language theories, which view achievement as performance, and as task and situation dependent, can provide the basis for effective teaching and achievement measurement. McLean has extended this "functional approach" to the areas of mathematics and science in the middle school. A recent issue of *Research Forum* (1989, No. 4), titled "Evaluation: Facing the Challenge", contained a number of articles that describe and examine the different kinds of measurement and evaluation practices employed by teachers including curriculum-based, criterion-referenced, naturalistic, and multidimensional evaluation procedures.

An emerging trend involves the identification of deficient cognitive functions using techniques such as Feuerstein's *Learning Potential Device* (Watts, 1985) and the analysis of both process and product in student learning, such as in spelling (Evans & Smith, 1989). The movement toward cognitive education programs will certainly require further modifications to the way teachers and psychologists view and measure student learning.

A most exciting four-year research project evaluating Feuerstein's *Instrumental Enrichment* (IE) and Mulcahy's *Strategies Program for Effective Learning and Thinking* (SPELT) has just been completed at the University of Alberta (Mulcahy, 1989). While the first draft of this report to the Alberta government is not yet available for circulation at this time, the author's have been given permission to cite this work. In brief, a sample of approximately 900 students experienced one of the IE, SPELT, or conventional instruction programs. Gifted, learning disabled, and normally achieving students in grades 4 and 7 were followed through to the completion of grades 6 and 9, respectively, permitting the comparison of a variety of achievement indices and cognitive strategies gained from the out-of-content versus in-content approaches. Generally, SPELT instruction was superior to IE and regular instructional methods, most notably in reading comprehension and related cognitive strategies. The learned cognitive strategies were maintained over time, together with improvements in achievement. The efficacy of the SPELT program was further observed in student gains in self-concept, locus of control, problem-solving strategies, metacognitive reading awareness, and comprehension monitoring skills. Learned reading and comprehension skills such as skimming, rereading, paraphrasing, inferring, and checking were successfully maintained. This study certainly provides evidence for the potential of such cognitive education programs but also points to the need for new and different assessment strategies.

Conclusions

There are many similarities in the school-based assessment practices of Canadian and American teachers, school psychologists, and other educational personnel. Although some school psychologists are still employed to administer tests (mainly individual intelligence tests), this narrow role appears to be on the verge of an even more widespread decline. Assessment is more than the administration of tests; it requires input from the various sources representing the "four pillars of assessment" (Sattler, 1988) but must be further viewed as a process including referral, screening, classification, planning, and program modification (Salvia & Ysseldyke, 1985).

While classroom teachers are continuously assessing student learning, the kinds of assessment information obtained by educational specialists frequently extends beyond data from a single test and is not limited to only a measure of, say, intelligence if it is to have utility in understanding and providing for a child's needs. There is also the recognition that the assessment must include information about the child and the child's situation or environment. This position is especially relevant in a multicultural mosaic such as Canada. While there is no equivalent to Mercer's (1979) *System of Multicultural Pluralistic Assessment* in Canada, recent descriptions of cross-cultural assessment have provided important guidelines for data collection, interpretation, and diagnosis (Massey, 1988; see also Samuda, Kong, Cummins, Pascual-Leone, & Lewis, 1989).

Many different tests may be employed to measure children's intelligence, personality, and school achievement. The need to examine the psychometric characteristics of instruments developed outside of Canada is recognized as is the importance of Canadian standardization of these norm-referenced tests. Achievement tests, such as the CTBS and curriculum-based achievement tests developed at the local and provincial levels, provide useful information but the fact that teachers show a preference for constructing their own measures needs to be addressed. The teacher requires assessment data that have instructional relevance, and educational measurement and evaluation specialists can certainly support the teacher's in-class assessment roles. Finally, emerging trends in cognitive education will require the development of new assessment techniques that will more effectively describe learning processes, strategies, and outcomes.

REFERENCES

- Anderson, J.O. (1989). Evaluation of student achievement: Teacher practices and educational measurement. *The Alberta Journal of Educational Research*, 35(2), 123-133.
- Andrews, J., & Janzen, H. (1988). Objectifying the Kinetic School Drawing (KSD): A comparison of LD and non-LD grade 5 students. *Canadian Journal of School Psychology*, 3, 27-51.
- Battle, J. (1976). Test-retest reliability of the Canadian Self-Esteem Inventory for Children. *Psychological Reports*, 38, 1343-1345.
- Battle, J. (1979). Self-esteem of students in regular and special classes. *Psychological Reports*, 44, 212-214.
- Battle, J. (1981). *Culture-free self-esteem inventory for children and adults*. Seattle: Special Child Publications.
- Beal, A. Lynne. (1988). Canadian content in the WISC-R: Bias or jingoism. *Canadian Journal of Behavioural Science*, 20(2), 154-166.
- Beal, A. Lynne. (1989). *Wechsler Intelligence Scale of Children III: A study to determine its validity for Canadian children*. A paper presented at the Canadian Association of School Psychologists' Conference, Edmonton, Alberta.
- Bellemare, F.G., Inglis, J., & Lawson, J.S. (1986). Learning disability indices derived from a principle component analysis of the WISC-R: A study of learning disability and normal boys. *Canadian Journal of Behavioural Sciences*, 18, 86-91.
- Berry, J.W. (1975). Psychological testing across cultures. In A. Wolfgang (Ed.) *Education of immigrant students*. Toronto: Ontario Institute for Studies in Education.
- Boersma, F.J., & Chapman, J.W. (1985). *Manual for the Student's Perception of Ability Scale*. Unpublished manuscript. University of Alberta, Edmonton, Alberta.
- Byrne, B.M. (1986). Self-concept/academic achievement relations: An investigation of dimensionality, stability, and causality. *Canadian Journal of Behavioural Science*, 18(2), 173-186.
- Collis, B., Ollila, L.O., & Yore, L.D. (1986). Predictive validity of the Canadian Readiness Test and selected measures of cognitive development for grade one reading and mathematics achievement. *The Alberta Journal of Educational Research*, 32(1), 2-11.
- Common, R.W., & Frost, L.G. (1988). The implications of the mismeasurement of Native students' intelligence through use of standardized intelligence tests. *Canadian Journal of Native Education*, 15, 9-30.
- Cummins, J. (1982). Tests, achievement, and bilingual students. *Focus, National Clearing House for Bilingual Education*, 1-7.
- Das, J.P., & Naglieri, D.H. (1990). Assessment of attention, simultaneous-successive coding and planning. In H.C. Haywood & D. Tsurie (Eds.), *Interactive assessment*. New York: Springer-Verlag.
- Cyr, J.J., & Atkinson, L. (1987). Test item bias in the WISC-R. *Canadian Journal of Behavioural Science*, 16, 30-35.

- Dickinson, G.M., & McKay, A.W. (1989). *Rights, freedoms and the education system in Canada*. Toronto: Edmond Montgomery.
- Dumont, F. (1987). School psychology in Canada: An evolving profession. *McGill Journal of Education*, 22(2), 101-116.
- Evanechko, P., Ollila, L., Downing, J., & Braun, C. (1973). An investigation of the reading readiness domain. *Research in the Teaching of English*, 7(1), 61-78.
- Evans, P., & Smith, A.L. (1989). Spelling errors: grade 1 to university-analysis of process and product. *The Alberta Journal of Educational Research*, 35(1), 20-33.
- Eysenck, S.B.G., & Saklofske, D.H. (1983). A comparison of responses of Canadian and English children on the Junior Eysenck Personality Questionnaire. *Canadian Journal of Behavioural Science*, 15(2), 121-130.
- Fox, Tricia A. (1988). Prisoner and students rights in regard to education: Are they implemented in Canada? *McGill Journal of Education*, 23(3), 265-276.
- Frost, B.P. (1979). *The Frost Self Description Questionnaire: Extended Scales*. Department of Educational Psychology, University of Calgary, Calgary, Alberta.
- Gardner, J. (1986). A comparison of the K-ABC and WISC-R for Cantonese, English and Punjabi speaking Canadians. *Canadian Journal of School Psychology*, 2, 45-54.
- Gardner, J. (1988). Factor structure of the K-ABC for Cantonese, English and Punjabi speaking Canadians. *Canadian Journal of School Psychology*, 4, 39-59.
- Goldstein, C.R. (1988). Cautions concerning assessment of Native Indian students. *Canadian Journal of School Psychology*, 4, 17-20.
- Hardman, S., & Oldridge, O. (1985). Predictive validity of WISC-R on the Woodcock-Johnson Achievement Battery, the B.C. Quick Individual Education Test and teacher rankings of student underachievement. *Canadian Journal of School Psychology*, 1, 31-38.
- Holmes, B. (1981). *Individually administered intelligence tests: An application of anchor test norming and equating procedures in British Columbia*. Unpublished doctoral dissertation. University of British Columbia.
- Janzen, H., Andrews, J., Fritzke, B., Clark, L., & Boersma, F. (1989a). *Relationship between the Stanford-Binet, 4th Edition and the WISC-R: A concurrent validity study*. Unpublished manuscript, Department of Educational Psychology, University of Alberta.
- Janzen, H., Andrews, J., Fritzke, B., Clark, L., & Boersma, F. (1989b). *The discriminant validity of the Stanford-Binet, 4th Edition*. Unpublished manuscript. Department of Educational Psychology, University of Alberta.
- Janzen, H.L., & Saklofske, D.H. (1990). Testing for diagnostic purposes. In L. Stewin, R. Short, & S. McCann (Eds.), *Psychology for the classroom*. Toronto: Copp Clark Pitman, (in press).
- Kamphaus, R.W., & Reynolds, C.R. (1990). *Behavior Assessment System for Children*. Circle Pines, MN: American Guidance Service, (in press).
- Kimmins, R., Hunter, W., & McKay, A.W. (1985). Educational legislation and litigation pertaining to the practice of school psychology in Canada. *Canadian Journal of School Psychology*, 1(1), 1-10.
- Kline, R.B. (1989). *The new Stanford-Binet is not a four factor test*. Paper presented at the 97th Annual Convention of the American Psychological Association, New Orleans, Louisiana.

- Loranger, M., Poirier, M., & Gauthier, D. (1983). Questionnaire d'auto-évaluation des conduites soules a l'école. *Canadian Journal of Behavioural Science*, 15(3), 280-300.
- Massey, Don S. (1988). Cross cultural psychological assessment: A clinical model of service delivery. *Canadian Journal of School Psychology*, 4, 21-37.
- Marx, R.W. (1984). Canadian content and the WISC-R Information Subtest. *Canadian Journal of Behavioural Science*, 16, 30-35.
- McDonald, D.D. (1985). Evaluation of the B.C. Quick Individual Educational Test. *Canadian Journal of School Psychology*, 1, 57-60.
- McLean, L.D. (1985). *The craft of student evaluation in Canada*. Toronto: Canadian Education Association.
- McLean, L.D. (1988). Achievement measures made relevant to pedagogy. *McGill Journal of Education*, 23(3), 243-252.
- McLeod, J. (1988). Individual differences from the perspective of computer-guided psycho-educational diagnosis. In D.H. Saklofske, & S.B.G. Eysenck (Eds.), *Individual differences in children and adolescents* (pp. 10-26). London: Hodder and Stoughton.
- Mercer, J.R. (1979). *System of Multicultural Pluralistic Assessment technical manual*. New York: Psychological Corporation.
- Mercer, J.R. (1988). Scientific revolutions and the psychomedical paradigm: The growing crisis in measurement psychology. *Canadian Journal of School Psychology*, 4, 1-16.
- Mulcahy, R.F. (1989). *Alberta Cognitive Education Program*. Department of Educational Psychology, University of Alberta, Edmonton.
- Nyberg, V.R., & Clarke, S.C.T. (1982). School Subjects Attitudes Scales. *The Alberta Journal of Educational Research*, 28(2), 175-287.
- Paananen, N., & Janzen, H. (1986). Incidence and characteristics of depression in elementary school children. *Canadian Journal of School Psychology*, 2(1), 7-19.
- Paul, J.F., & MacLeave, A.M. (1989). Educators' perceptions of intelligence and intelligence testing. *Canadian Psychology*, 30(2a), 380.
- Peters, H.D. (1976). The validity of the Wechsler Intelligence Scale for Children-Revised. *Canadian Journal of Behavioural Science*, 8, 414-417.
- Rampaul, W.E., Singh, M., & Didyk, J. (1984). The relationship between academic achievement, self-concept, creativity, and teacher expectations among Native children. *The Alberta Journal of Educational Research*, 30(3), 213-225.
- Ray, D. (1989). Human rights: Is Canadian schooling a privilege, a right or an obligation? *McGill Journal of Education*, 23(2), 143-155.
- Rogers, H., & Saklofske, D.H. (1985). Self-concepts, locus of control and performance expectations of learning disabled children. *Journal of Learning Disabilities*, 18(50), 273-278.
- Saklofske, D.H., & Braun, S. (1989). *A psychometric investigation of the Draw-A-Person: A Quantitative Scoring System utilizing a Canadian sample*. Paper presented at the Canadian Association of School Psychologists' Conference, Edmonton, Alberta.

- Saklofske, D.H., & Eysenck, S.B.G. (1983). Impulsiveness and adventuresomeness in Canadian children. *Psychological Reports, 52*, 147-152.
- Saklofske, D.H., Janzen, H.L., & Paananen, N. (1987). Predictors of childhood depression: Literature review and empirical findings. *Canadian Journal of Special Education, 3*(1), 1-14.
- Saklofske, D.H., & Jedlicki, G. (1985). *Concurrent and content validity of the K-ABC*. Paper presented at the 17th Annual Convention of the National Association of School Psychologists, Las Vegas.
- Saklofske, D.H., & Murray, W. (1989). *A normative and validity study of the Matrix Analogies Test-Short Form with Canadian children*. Paper presented at the Canadian Association of School Psychologists' Conference, Edmonton, Alberta.
- Salvia, J., & Ysseldyke, J.E. (1985). *Assessment in special and remedial education* (3rd ed.). Boston: Houghton Mifflin.
- Samuda, R.J., Kong, S.L., Cummins, J., Pascual-Leone, J., & Lewis, J. (1989). *Assessment and placement of minority students*. Toronto: Hogrefe and Huber.
- Sattler, J.M. (1988). *Assessment of children* (3rd ed.). San Diego: Jerome M. Sattler.
- Schmidt, H.P., Kuryliw, A.J., Saklofske, D.H., & Yackulic, R.A. (1989). Stability of WISC-R scores for a sample of learning disabled children. *Psychological Reports, 64*, 195-201.
- Schmidt, H.P., & Saklofske, D.H. (1983). Comparison of WISC-R patterns of children of average and exceptional ability. *Psychological Reports, 52*, 539-544.
- Siegal, L.S. (1984). On the adequacy of the WRAT: A reply to Snart, Dennis and Brailsford. *Canadian Psychology, 24*, 75-76.
- Simner, M. (1987). Predictive validity of the Teacher's School Readiness Inventory. *Canadian Journal of School Psychology, 3*(1), 21-32.
- Snart, F., Dennis, S., & Brailsford, A. (1983). Concerns regarding the Wide Range Achievement Test. *Canadian Psychology, 3*, 21-32.
- Summers, E.G. (1980). Validity of the Estes Reading Attitude Scale for intermediate grades. *The Alberta Journal of Educational Research, 26*(1), 36-43.
- Summers, E.G., & McLelland, J.V. (1982). A field-based evaluation of sustained silent reading in intermediate grades. *The Alberta Journal of Educational Research, 28*(2), 100-112.
- Thompson, W.W. (1985). Environmental effects on educational performance. *The Alberta Journal of Educational Research, 31*(1), 11-25.
- Vernon, P.E. (1977). Final report on modifications of WISC-R for Canadian use. *Canadian Psychological Association Bulletin, 5*, 5-7.
- Wahlstrom, M., & Darley, R.R. (1976). *Assessment of student achievement*. Toronto: Ontario Ministry of Education.
- Walsh, J., Marx, R.W., & Sudmant, W. (1983). WISC-R factor score profiles and academic achievement: A test of Kaufman's profile Analysis procedures. *Canadian Journal of Behavioural Science, 15*(3), 203-210.
- Watts, W.J. (1985). An error analysis on the Raven's using Feuerstein's Deficient Cognitive Functions. *The Alberta Journal of Educational Research, 31*(1), 41-53.
- Whyte, L.A. (1984). Characteristics of learning disabilities persisting into adolescence. *The Alberta Journal of Educational Research, 20*(1), 14-25.

- Williams, D.G. (1988). Developmental trends in information processing and reading: A cross-cultural comparison. *Canadian Journal of School Psychology, 4*, 39-47.
- Wilson, R.J. (1989). Evaluating student achievement in an Ontario high school. *The Alberta Journal of Educational Research, 35*(2), 134-144.
- Wilson, R.J., & Humphries, T. (1986). The impact of recent legislation on the role of Canadian school psychologists. *Canadian Journal of School Psychology, 2*(1), 1-6.
- Wiss, C.A., & Burnett, W. (1988). The application of the Boder Test of Reading-Spelling Patterns to the assessment of learning disabilities in French immersion poor readers. *Canadian Journal of School Psychology, 4*, 61-70.
- Wormeli, C.T. Jr. (1983). *Manual for the B.C. QUIET*. University of British Columbia Education Clinic. Vancouver, BC.

Donald H. Saklofske received his Ph.D. from the University of Calgary and is currently a professor in the Educational Psychology Department at the University of Saskatchewan. He has published papers on intelligence, personality and assessment and recently co-edited a book on individual differences.

Henry L. Janzen received his Ph.D. from the University of Calgary. He is a professor in the Department of Educational Psychology at the University of Alberta, Director of the Psychological Testing Center at the Faculty of Education, Clinical Services, and a consulting school psychologist. Dr. Janzen is active in many psychological associations, has published extensively, and is co-host of the radio program *That's Living* at CJCA 930.

Donald H. Saklofske a obtenu son doctorat à l'Université de Calgary et il enseigne actuellement au département de psychologie éducationnelle à l'Université de Saskatchewan. Il a publié des articles sur l'intelligence, le caractère et l'évaluation et a récemment coédité un livre sur les différences individuelles.

Henry L. Janzen a obtenu son Ph.D. à l'Université de Calgary. Il est professeur au département de psychologie éducationnelle de l'Université d'Alberta, directeur du centre de tests psychologiques des services cliniques de la faculté des sciences de l'éducation et psychologue scolaire-conseil. Il est membre de nombreuses associations de psychologie, a publié de nombreux ouvrages et est le coanimateur de l'émission *That's Living* diffusée sur les ondes de CJCA 930.



Joseph C. LaVoie
University of Nebraska at Omaha

School-Based Assessment Research in the People's Republic of China

Abstract

School-based assessment, as we know it, does not occur in the People's Republic of China. Assessment procedures, primarily achievement tests, are used for placement purposes in select schools. Some psychological testing is done with young children to detect learning disabilities. Several psychological tests are available, but they are not widely used. Research in the areas of intellectual, personality, clinical, and achievement assessment is reviewed in this paper. Given the lack of psychological services for schools, the role of teachers in the evaluation process is discussed with respect to educational philosophy in the People's Republic of China. The impact of recent events in China on the educational system is also discussed.

Résumé

L'évaluation en milieu scolaire telle que nous la connaissons, ne se pratique pas en République populaire de Chine. Les modalités d'évaluation, essentiellement les tests de niveau, servent à placer les élèves dans certaines écoles. Certains tests psychologiques sont administrés aux jeunes enfants pour dépister les troubles d'apprentissage. Plusieurs tests psychologiques existent, mais leur emploi est loin d'être généralisé. L'auteur du présent article analyse les recherches qui se font dans les domaines des tests d'intelligence, de personnalité, de niveau et domaines des tests d'intelligence, de personnalité, de niveau et des tests cliniques. Compte tenu de l'absence de services psychologiques dans les écoles, le rôle des enseignants dans le processus d'évaluation est analysé par rapport aux principes de l'éducation en République populaire de Chine. L'auteur aborde également l'impact des récents événements qui se sont produits en Chine sur le système éducatif.

Progress in school-based assessment in the People's Republic of China has been very slow. According to Catterall's (1982) review of international school psychology, reports from the People's Republic of China* indicated that there were no psychologists providing services to the schools, although psychology was a rapidly expanding field in the country at that time, and there was a focus on education of the developmentally disabled. The situation has not changed greatly, according to LaVoie (1989a, 1989b) who found that school psychology is not a recognized field, although some psychologists working in the field of education publish articles that pertain to school psychology, and certain ongoing activities reflect assessment concerns of school psychology. For example, educational psychologists have been involved in test development for the purposes of assessment and placement, and developmental psychologists have been interested in the child's cognitive development, while medical doctors have focused their assessment mostly on the first five years of life (LaVoie, 1989a, 1989b).

There are several reasons why school psychology has not emerged as a separate discipline. First, as LaVoie (1989a) noted, educational decisions are made by political, not educational, leaders. While the need for psychological services in the schools is recognized, such services are still a luxury in a country with so many other demands for psychologists, particularly in the clinical and other mental health settings where the needs are so urgent. According to Qicheng (1988), the primary focus in education is on the practical use of knowledge, such as children's use of cognitive strategies to solve problems in mathematics, determining the manner in which Chinese characters are formed by people so that a computer system that uses these characters can be developed, and the effect of the one-child family. Qicheng contends that it is necessary to be very selective in what is studied because the problems are so numerous. Psychological services in the schools have to compete with such other demanding areas in psychology as industrial, clinical, and health.

Not to be overlooked is the underlying educational philosophy in China that all children have the potential to learn the curriculum. Given this mindset, it follows that the potential for change has to lie within the child. Senkowski, Corser, and McLean (1989) found that the teacher's job is to reduce individual differences in the classroom by working with the entire class so that they all perform at the same basic level. Therefore, the purveyor

*Throughout the remainder of this paper the People's Republic of China will be referred to simply as China in order to facilitate smoother reading.

of psychological services in the schools is the teacher. LaVoie (1989b) noted that teachers serve some of the functions that would be delegated to school psychologists in the United States. If the services are already provided by teachers, it then becomes less critical to assign psychologists to the schools.

Given this overview of the Chinese educational philosophy, what is the current state of school-based assessment in China? To answer this question, one first needs to determine what school-based assessment consists of in the United States so that we have some basis for making a comparison. According to Sattler (1988), assessment usually involves an evaluation of the child's learning abilities, speech and language, visual-motor ability, and identification of behaviour problems. The child's developmental history and how the child interacts with others are also considered. Assessment in China focuses on the characteristics of individual children, which reflects the medical model. School psychology in the United States originally adopted this model, but has since changed (Reynolds, Gutkin, Elliott, & Witt, 1984). The primary use of assessment in China is for screening and placement, but mostly the latter. For the Chinese, the test score is important, because it indicates the child's level of intellectual functioning or determines the school that the child will attend. Factors other than test scores, such as developmental history or current life circumstances, are not considered in placement or screening decisions in China.

Before reviewing some of the representative research related to school-based assessment, it needs to be recognized that the Chinese were among the first people to use assessment procedures to evaluate performance. Bowman (1989), in her comments on testing in China, cited evidence to show that testing procedures for determining individual differences in mental ability were being used 2000 years ago. Speed of writing and speaking were used as measures of intelligence as early as 1 A.D., according to a review of assessment in China by LaVoie (1989b). He also noted that personality assessment was carried out in ancient China, and sensorimotor tests of infant development were used in the sixth century A.D.

According to Hawkins (1983), China's fascination with testing has spread throughout the educational system so that the curriculum at the primary and secondary levels is skewed to prepare students for the secondary school and university entrance exams. Therefore, the curriculum does not have the necessary breadth to educate both the college track student and those students who cannot gain entry to college. Further, teachers are evaluated on the basis of the number of students who pass the entrance exam for secondary schools and universities.

Intellectual assessment

The *Binet-Simon Scale* was first translated into Chinese in 1922, and adopted for use in assessment in 1924, at which time norms were established. The *Stanford-Binet Scale* was revised in 1981 and renamed the *China Binet Scale* (Wu, 1985). Western tests, such as the *Wechsler Intelligence Scale for Children - Revised* (WISC-R) and the *Peabody Picture Vocabulary Test* (PVT) are the instruments that have been most often used to assess intellectual performance. D. Li, Jin, Zhu, and Tang (1987) have validated and normed the Chinese translation of the WISC-R, and their reported subscale and full-scale means and standard deviations are similar to those in the American standardization group. However, as noted by LaVoie (1989b), D. Li *et al.* found performance differences between the Chinese and American standardization groups. The Chinese were five years ahead of their U.S. age-mates on math and digit span scores, four- to five-years advanced on block design, and two- to three-years ahead on coding, but they scored two- to three-years behind the U.S. comparison group on picture completion. Some of the subtests in the verbal subscale did not discriminate well, according to D. Li *et al.* After 10 years of age, the vocabulary and math items did not discriminate significantly among low and high IQ children, and the comprehension, block design, and maze subtests were less discriminative between low and high IQ children after 12 years of age. But the relationship between WISC-R IQ and student GPA was reasonably high (.76, *Phi* coefficient). D. Li *et al.* conclude from their findings that while the revision seems to fit reasonably well for the Chinese culture, there are some problems in the accuracy of the WISC-R for the Chinese (LaVoie, 1989b).

The role of environmental factors in IQ scores has also been considered by the Chinese. Using a longitudinal study of children from birth to 36 months, Y. Mao (1986) found that preschool attendance and gender were related to IQ scores. Those children who had attended preschool at an early age had higher IQ scores. Occupation of the family and being an only child, as well as certain personality factors, were also predictive factors. Children who were curious, persistent, and nonimpulsive had higher IQ scores. Mao concludes that adaptive behaviour and cognitive ability are the main indices of IQ at this age period.

Intellectual assessment has been used mostly to identify mental retardation, and to assess intellectual performance among the hearing impaired. The PPVT, for example, has been used to screen young children for learning disabilities. Gong and Gua (1984) reported that the PPVT correlated .65 with Chinese language score and .61 with arithmetic score in their sample of 7-8-year-old children. The Chinese are interested in gifted children, or what they

call "super-normal," but formal assessment has not been used to identify this group of children, although some attempts have been made to use such tests as analogical reasoning, problem solving, and mathematics to detect gifted children. However, intelligence testing has generated much controversy about such issues as the use of western tests with their cultural bias, and the problem of low correlations with certain measures of academic performance, particularly grade point average and course-work exam grades. LaVoie (1989b) cites a 1981 study by Wei and Shi who reported that their measure of IQ (national entrance exam scores) was minimally correlated with college grades.

Personality assessment

Among the personality measures translated into Chinese are the *Minnesota Multiphasic Personality Inventory* (MMPI), *California Psychological Inventory* (CPI), and the *Eysenck Personality Inventory* (EPI) (LaVoie, 1989b). However, the MMPI has been used more extensively than the other instruments. Clinical studies constitute most of the reported research, although a few projects involving personality assessment have been conducted in the schools. A study by Yang and Shun (1986) is representative of the assessment research. They used a trait-anxiety inventory to select high and low anxiety middle school students who were then given the MMPI. The two anxiety groups did not differ significantly on any of the MMPI scales. The MMPI has also been used to examine differences in parents of minimal brain-damaged children. Zhu (1987) reported that parents of these children scored higher on the Pd (psychopathic deviate) and PT (psychasthenia) subscales than parents of normals, which suggests a potential interactive factor, in that the behaviour problems of the children may, in part, be associated with the psychological adjustment of their parents.

Temperament in elementary and middle school children has been studied by G. Lin (1986), who developed a 40-item scale based on Galen's four classification theory of temperament – sanguine, melancholic, choleric, and phlegmatic. Lin found that children could be classified into one of the four categories. Independence in early adolescents (ages 11 – 14) is a concern of some psychologists and school personnel in China. In one such study (H. Mao, 1984), young adolescents identified their independence needs to include respect from peers, recognition as adults, and the admission of parents and teachers that they sometimes can be wrong. These independence needs were then paired with such adolescent personality traits as radicalism, stubbornness, exhibitionism, and noncompliance. From this conceptualization, Mao developed an instructional program for teachers of this age group so that they could provide guidance for adolescents in their independence seeking. Among the objectives of the program were to socialize the adolescent to

develop self-control, reduce conflict with parents, and to enhance moral values.

Clinical assessment

Much of the interest in this area can be attributed to the adoption of the medical model. Currently, clinical assessment is conducted in medical clinics where psychologists are located. There is a great need for behavioural assessment within the school setting, but the shortage of trained psychologists makes this service impossible at the present time. The *Luria-Nebraska Neuropsychological Battery* has been revised and normed for Chinese use by Xu and Gong (1987). The discriminant validity of the test was .90, based on the correct assignment of 70 of 78 subjects who had been previously diagnosed as normal or having some type of neurological disorder.

Child behaviour problems, particularly frustration, have received some attention in the Chinese psychological literature (e.g., Shun, 1986). Elementary and middle school students were asked to describe an event that angered/frustrated them most. Students identified such events as preparation for entrance exams for middle schools, interpersonal relations, and specifically, inability to form friendships, difficulty in getting along with teachers, parent relations, personal inadequacies, and failure to gain respect. Some age and gender differences in frustration were found by Shun. Elementary school children reported that parent relations and personal inadequacies were the most frustrating, whereas, middle school students indicated that interpersonal relations and failure to gain respect were the most frustrating. Boys were more likely to react to frustrating events with aggression, while girls responded with avoidance.

Chinese psychologists have also been concerned with the problem of school dropout. Shao (1986), in his study of middle school (i.e., secondary school) dropouts, found that 17% of the friends of dropouts were still in school, but 20% had already dropped out. Interestingly, 60% of the dropouts' friends were classified as conduct disorders. Personal characteristics associated with this group of dropouts included less time spent in reading, lower motivation and poor study habits, more time loitering, and greater incidence of serious delinquent activities, than in-school students. These characteristics are similar to those describing school dropouts in the United States (e.g., Steinberg, 1989). When the dropouts were asked about their future goals, 44% responded that they were free to do whatever they wanted, 45% were concerned and did not know what they would or could do, 8% said that they would stay at home and wait for the government to help them, and 2% said that they had ruined their life.

Achievement assessment

More attention has been given to achievement assessment than any of the others because it impacts directly on educational placement. The Chinese are concerned with selecting the most capable students to attend their "key" (i.e., highly selective) junior and senior high schools. Achievement tests are used throughout the country to select these students. The entrance examination for admission to middle schools is a standardized test developed by the Educational Bureau of the government and consists of the following sections: Chinese language, mathematics, history, natural science, geography, moral education, and physical training. The minimum score for admission to a middle school is 60%, but all students are ranked for admission purposes. Students in the top 10% of the ranking are admitted to "key" middle schools, which are the best schools in the country. Other than this evaluation program, achievement testing in the classroom, as used in the United States, is not practiced in China.

Although the Chinese education system has committed major resources to the development of the middle school admission examination, little or no research has been conducted on the reliability or validity of the test, but some insight into achievement assessment in China can be gained from cross-cultural studies. Hess, Chang, and McDevitt (1987) investigated the relation between sixth-grade-children's mathematics performance and the attributions of their mothers about this performance. Boys and girls and their mothers from the following cultural groups, China (47 families), Chinese-American (51 families), and Caucasian-Americans from the United States (47 families), were involved in the study. Each mother was asked to indicate how well her child had performed in mathematics during the sixth grade, and then to provide explanations for the child's success (if the child had performed well) or the child's failure (if the child had not performed well). Children were also asked to explain the reasons for their level of performance.

Maternal explanations for the failure of their child to perform at a higher level were similar among the three cultural groups in that the mothers emphasized the child's level of effort, but the Chinese mothers cited lack of effort, an internal factor, as the primary cause for poor performance. Chinese-American mothers placed the blame on lack of ability, inadequate training at school, and insufficient help at home. American mothers, however, tended to spread the responsibility across a number of factors, most of which were not controllable by the family or the child. Mothers of American children were more likely to place responsibility for poor performance on the school, whereas both groups of Chinese mothers were more likely to attribute responsibility to training in the home. Mothers also differed in the action that they

would take in responding to failure performance in their children. American and Chinese-American mothers indicated that they would try to determine the reason for the failure, but Chinese mothers stated that they would express anger and perhaps punish the child.

Responses of the children were similar to those given by their mothers. Children from China chose controllable causes (i.e., effort) for their poor performance; whereas, both Chinese-American and American children were less likely to attribute poor performance to controllable factors, and American children were more likely to perceive luck as a factor in poor performance. But all children from the three cultures more often identified controllable factors (i.e., effort) in their attributions for poor performance than did their mothers.

Mothers' attributions for their child's success in mathematics also differed among the cultures. Mothers from China more often gave credit to the school, while Chinese-American mothers saw home training as most important, and American mothers selected ability, effort, and school training as the critical factors. Perhaps the most revealing finding was the mothers' responses to high performance by their children. Chinese-American and American mothers were more likely to reward their child with praise for good performance. Mothers from China, on the other hand, were more likely to set higher standards for their child, and less likely to offer praise for high performance.

These findings have significance for school-based assessment research on achievement. As Hess *et al.* (1987) suggest, the results indicate that motivational factors are important sources of differences in achievement, and that cultural factors have to be considered. It is apparent from the Hess *et al.* study that certain attributions about academic performance are stable, and these beliefs appear to be incorporated into the socialization process.

Further insight into factors that impact on school-based assessment of achievement in China emerges from a second series of cross-cultural studies involving first- and fifth-grade children (22 classrooms at each grade level) from Beijing (China), and Chicago (USA). In one study, Lummis (1989) administered curriculum-based achievement tests of vocabulary and mathematics to the students. Analyses of the test scores showed that the American sample had a higher mean score on vocabulary, but the Chinese students scored higher on the math tests. The data analyses also revealed that the score variance on both vocabulary and mathematics was much larger for the American sample than for the Chinese sample, indicating greater individual differences among the American students than among the Chinese students.

One likely explanation for the cultural differences is teaching. Uttal and Lee (1989) found some major differences among the two groups of teachers. Chinese teachers devote more time to lesson preparation and other activities associated with instruction. Further, nearly all of their instructional time is used in classroom teaching. There is no individual instruction as we know it. The teaching objective of Chinese teachers is to bring slow students to the level of the class, so they spend more time teaching the slow student. Other beliefs held by Chinese teachers include: it is important to get the correct answer quickly, therefore speed in problem solution is emphasized; one works very intensely in the classroom; and a good teacher is one who can explain a lesson clearly. That is, one needs to teach well at all times. These findings suggest that school-based achievement assessment in China has a different meaning, which should result in a different plan of action than similar assessment in the United States.

Conclusions

It becomes evident from this discussion that school-based assessment is in its infancy in China. Various assessment measures are available, but the primary objective of assessment appears to be limited to placement rather than diagnosis or evaluation. Sattler (1988) makes an important distinction between psychological assessment and psychological testing, which fits this discussion quite well. Psychological testing involves the use of tests, which provide findings, whereas psychological assessment provides an interpretation of these findings given the child's situation. Some psychological testing is done in China, particularly for placement. But psychological assessment as practised in the United States, is not occurring at the present time in China. However, teachers in China are conducting school-based assessment to the extent that they not only perform ongoing evaluation, but they also modify their teaching to correct the identified problems. Unlike the United States, where the focus is on individuality, ability grouping is not used in China. Rather, the teacher strives to increase the performance of the entire classroom. This approach fits with the educational philosophy in China which assumes that if a problem exists, one focuses on teaching because that is where the problem is likely to reside.

Most psychologists would agree that assessment measures are merely tools to sample behaviour. We need to know more about the child to determine the why of the child's behaviour. When a child in China does not meet the expected standards in the classroom, Chinese teachers help the child with the deficient work and focus on their teaching to increase uniformity of performance in the classroom, whereas in the United States we tend to assess why the child has a problem.

Recent events in China will probably mandate changes in the educational system. Political studies have always been part of the curriculum at all levels, and the student's understanding of political philosophy has been assessed in entrance exams to secondary schools and universities. Starting in the Fall of 1989 increased emphasis has been placed on political philosophy in the schools. Assessment of political understanding will probably occur at all levels of schooling. The Chinese are adamant in their belief that the solution to problems, such as alienation, disruptive behaviour, and disrespect is to reeducate people and insist on accountability.

REFERENCES

- Bowman, M.L. (1989). Testing individual differences in ancient China. *American Psychologist*, *44*, 576-577.
- Catterall, C.D. (1982). International school psychology: Problems and promises. In C.R. Reynold & T.B. Gutkin (Eds.), *The handbook of school psychology*. New York: Wiley.
- Hess, R.D., Chang, C.M., & McDevitt, T.M. (1987). *Cultural variations in family beliefs about children's performance in mathematics: Comparisons among the People's Republic of China, Chinese-American, and Caucasian-American families*. Unpublished manuscript, Stanford University.
- Gong, Z., & Gua, D. (1984). An intelligence screening test for preschool and primary school children – Peabody Picture Vocabulary Test. *Acta Psychologica Sinica*, *16*, 393-401.
- Hawkins, J.N. (1983). The People's Republic of China (mainland China). In R.M. Thomas, T.N. Poselathwaite (Eds.), *Schooling in East Asia*. New York: Pergamon Press.
- LaVoie, J.C. (1989a). School psychology in the People's Republic of China. In P.A. Saigh, & T. Oakland, Eds.), *International perspectives on psychology in the schools*. Hillsdale, NJ: Earlbaum
- LaVoie, J.C. (1989b). School psychology research in the People's Republic of China. *Professional School Psychology*, *4*, 137-145.
- Li, D., Jin, Y., Zhu, Y., & Tang, C. (1987). Report on the norm of the Weschler Intelligence Scale For Children-China Revised (WISC-CR) in Shanghai proper. *Acta Psychologica Sinica*, *19*, 136-144.
- Lin, G.Z. (1986). Test of temperament for elementary and middle school. *Popular Psychology*, *4*, 22-23.
- Lummis, M. (1989). *The effects of parents and schools on academic achievement of Chinese, Japanese, and American children*. Paper presented at the Biennial Meeting of the Society for Research in Child Development, Kansas City, April 27-30.

- Mao, H.L. (1984). Development of independence in adolescent personality. *Journal of Qui Qui Ha Normal University*, 4, 124-128.
- Mao, Y.Y. (1986). Child's intellectual development from birth to 36 months. *Acta Psychologica Sinica*, 18, 248-255.
- Qicheng, J. (1988). *Psychology in the developing nations of the world: The People's Republic of China*. Paper presented at the Annual Meeting of the American Psychological Association, Atlanta, August 12-16.
- Reynolds, C.R., Gutkin, T.B., Elliott, S.N., & Witt, J.C. (1984). *School psychology: Essential of a theory*. New York: Wiley.
- Sattler, J.M. (1988). *Assessment of children*. San Deigo: J.M. Sattler.
- Senkowski, R., Corser, J., & McLean, C. (1989). *Education for excellence: The Asian experience*. Paper presented at the Biennial Meeting of the Society for Research in Child Development, Kansas City, April 27-30.
- Shao, D.S. (1986). Dropping out of school influences student's socialization. *Acta Psychologica Sinica*, 18, 190-195.
- Shun, L.M. (1986). Child behavior problems and frustration. *Journal of Nanjing Normal University*, 2, 13-17.
- Steinberg, L. (1989). *Adolescence*, 2nd ed. New York: Alfred A. Knopf.
- Uttal, D.H., & Lee, S.Y. (1989). *Teacher beliefs and practices: A comparison of China and the U.S.* Paper presented at the Biennial Meeting of the Society for Research in Child Development, Kansas City, April 27-30.
- Wu, L.G. (1985). On intelligence tests. *Journal of Beijing Normal University*, 4, 103-104.
- Xu, Y., & Gong, Y.X. (1987). A preliminary revision of the Luria-Nebraska Neuropsychological Battery. *Communication on Psychological Science*, 3, 28-35.
- Yang, L.N., & Shun, Z.X. (1986). MMPI analysis of test anxiety in middle school students. *Communication on Psychological Science*, 3, 36-38.
- Zhu, D.S. (1987). Correlation of MBD children with parent characteristics. *Communication on Psychological Science*, 3, 46-48.

I would like to acknowledge the contribution of those persons who assisted me with this manuscript. Lai Qun, University of Nebraska at Omaha, assisted with the translations of the articles from Chinese journals. Aron Armfield, University of Nebraska at Omaha, and Qicheng Jing, Institute of Psychology, Academia Sinica, were consulted about matters of applied psychology in the People's Republic of China.

Joseph C. LaVoie (Ph.D., University of Wisconsin-Madison) is a professor of developmental psychology in the Department of Psychology, University of Nebraska at Omaha. Professor LaVoie's research interests are in social and personality development in children and adolescents, specifically peer relations, social competence and its relation to academic performance, and individuation processes in adolescence. His interest in school psychological services in the People's Republic of China is the result of a six-month assignment in China as a visiting scholar, and his doctoral studies in cultural anthropology.

Joseph C. LaVoie (Ph.D., Université du Wisconsin-Madison) enseigne la psychologie du développement au département de psychologie de l'Université du Nebraska, à Omaha. Ses recherches portent sur le développement social et le développement du caractère des enfants et adolescents, notamment sur les relations entre pairs, les compétences sociales et leur corrélation avec les résultats scolaires et les processus d'individuation chez les adolescents. Son intérêt pour les services psychologiques scolaires en République populaire de Chine est le fruit d'un séjour de six mois qu'il a effectué en Chine à titre de chercheur invité et de ses études de doctorat sur l'anthropologie culturelle.

School-Based Assessment Research in Israel: Current state and future directions

Abstract

The major aim of the present paper is twofold: (1) to survey and evaluate the present "state of the art" in Israeli school-based assessment research; and (2) to delineate potential future directions, foreseeable developments, and needed research. The paper focuses primarily on the assessment of aptitude, achievement, and personality in the schools and highlights a number of key issues engaging the interests and concerns of the Israeli school psychology and measurement community (i.e., group differences in test performance; test bias and equity; and testing the selection policy in the schools). There is currently in Israel a paucity of well developed, soundly normed, and psychometrically validated (or adequately adapted) instruments available for assessment purposes – particularly in the area of non-cognitive assessment. Furthermore, there is a need for the development of criterion-referenced measures in a variety of school subjects for student evaluation and achievement tracking purposes in both the Jewish and Arab sectors. This paper reveals the need for a major concerted effort to improve both the quantity as well as the quality of available school-based assessment instruments, procedures, and practices.

Résumé

L'objectif principal de cet article est double: 1) sonder et évaluer la situation actuelle des recherches sur l'évaluation en milieu scolaire en Israël; et 2) délimiter les orientations futures, les développements prévisibles et les recherches nécessaires. L'article porte essentiellement sur l'évaluation des aptitudes, des résultats et du caractère dans les écoles et il fait ressortir un certain nombre de problèmes clés qui présentent de l'intérêt pour les psychologues scolaires israéliens (différences de groupe dans l'exécution des tests; partialité et équité des tests; et évaluation de la politique de

sélection dans les écoles). On constate actuellement en Israël une pénurie d'instruments d'évaluation intelligemment conçus, aux assises solides et validés selon des moyens psychométriques, notamment dans le domaine de l'évaluation non cognitive. En outre, on constate le besoin d'établir des mesures à base de critères dans diverses matières enseignées à l'école pour évaluer les élèves et contrôler leurs résultats dans les communautés juive et arabe. Cet article fait ressortir le besoin d'un effort concerté pour améliorer la quantité et la qualité des instruments, des modalités et des pratiques d'évaluation en milieu scolaire.

Israel, much like the United States, may be aptly depicted as a test-oriented and test-consuming society, with tests widely employed by the educational system, military, industry, and government as an aid to making decisions about people. Tests have a particularly formidable presence in the school system and are employed for a variety of functions, including: (a) research and program evaluation; (b) vocational guidance and counselling; (c) student selection, classification, and placement (e.g., entrance, streaming, homeroom class-formation, within-class academic tracking); (d) screening and diagnosis; and (e) identification and selection of children for special programs (e.g., gifted children's program, special education classes).

General historical background information

Prior to the establishment of the State of Israel in 1948, student populations were fairly homogeneous, composed largely of children from Jewish middle-class families of European origin (about 90%). Since the inception of the State of Israel, the student body has changed radically in content and scope, with the student population growing seven-fold since 1943. Mass immigration from over 70 countries, including post-war Europe and the Middle East, doubled the population over the first three years of statehood and tripled it over the first twelve (Adler, 1984).

During the early years of statehood, the educational system in Israel was charged with accommodating an increasing percentage of Jewish children from Eastern cultures (e.g., Morocco, Algeria, Yemen, Egypt, Persia), many of whom presented the classical symptoms of cultural deprivation, including inadequate mothering and bonding; low degree of mediated learning experiences; and reported sensory, linguistic, and conceptual deficits (Feuerstein, 1980; cf. Raviv, 1989). As the State absorbed more and more immigrants, the school system faced the problem of greater heterogeneity in its student body. This variety manifested itself in both family structure and cultural-educational orientation, traditional customs, and behaviour patterns, as well as students' achievement profiles and motivation. Whereas immigrant

students of European extraction adapted quite readily to modern Israeli society and its Western-oriented school system, students of Eastern background evidenced considerable difficulty in the school acculturation process; they tend to regard attainment of education in a more instrumental way and were prepared to invest less in education as a means of acquiring status (Kfir, 1988). Since students of Eastern origin had lower levels of achievement aspiration compared to their Western counterparts, they were also destined to be less mobile within a society that holds modern status-attainment norms (Adler, 1984; Kfir, 1988).

It was the educational system that assumed the main burden and responsibility for equalizing opportunities and integrating culturally different groups into Israeli society, no mean task (Adler, 1984). Furthermore, it was largely school psychologists – whose original function in Israel was to assess and diagnose failing students – who were charged with assessing and placing the masses of immigrant children. Unfortunately, few standardized measures were available at the time for assessment functions; the exigencies of the times dictated the use of measures that were often not adequately translated, adapted, normed, or validated for their culturally different target populations. Only in the past two decades or so have serious efforts been made to construct standardized culturally-indigenous instruments appropriate for the Israeli scene or in adequately norming conventional measures on representative school populations in Israel.

Major goals and scope of this paper

School-based assessment and evaluation is currently one of the key research areas in Israeli school psychology (Raviv, 1988). Indeed, the history of school psychology, both in the United States and in Israel, is in many ways synonymous with the history of efforts to assess and evaluate students' cognitive, affective, and psychomotor behaviours. The bulk of school-based assessment research is carried out in academic settings by university faculty rather than by school practitioners. Unfortunately, the development of school-based assessment in Israel has not been systematically documented over the years and little integrative work has been devoted to historical, current, or future perspectives in school-based assessment research (Gina Ortar, personal communication, June, 1985). Consequently, this paper is intended to fill a needed gap by evaluating the present state of the art of Israeli school-based assessment research and by suggesting some potential future trends and possible directions.

The author of this paper relies primarily on three sources: (a) relevant papers identified through conventional bibliographic research as well as a

computerized search of the Israeli periodical literature for assessment-related material published over the past 15 years; (b) documents, papers, and tests provided to the author by a number of key Israeli scholars who were personally approached for relevant material; and (c) data bearing on school-based assessment in Israel, gathered from knowledgeable Israeli informants, as part of a cross-national study of testing and assessment practices among children and youth (Thomas Oakland, personal communication, December, 1988).

The paper is restricted in its scope to assessment measures (i.e., cognitive, achievement, personality) as well as assessment issues and procedures relevant to school-aged populations in Israel; preschool assessment or college-age and adult assessment themes will not be addressed. The choice and selection of material to survey was clearly subjective and by no means purports to cover the entire domain of Israeli school-based assessment research in the past or present. Due to the disparate and fragmentary sources for garnering information about school-based assessment research in Israel, a full and complete description would have been an insurmountable task. Moreover, given the author's own research interests in the areas of ability testing and test bias, coupled with the fact that some of the better assessment instruments, procedures, and research endeavours in Israel have been associated with ability and intelligence measures, there is a distinct "positive bias" in the overrepresentation of the ability-test domain in this paper.

In addition, it is clearly preferable to present a more detailed and in-depth coverage of cardinal instruments and assessment issues and high quality research and evaluation studies, as opposed to aiming for an exhaustive coverage of all relevant issues.

Cognitive Assessment

Notwithstanding the growing trend in Israeli school psychology over the past years to assist children through working with parents and their children's teacher – reflected in the decline in individual diagnosis and increase in prevention services – a good portion of the school psychologist's time is still devoted to assessment (Raviv, 1989). Individual diagnosis, including diagnosis of pupils suspected of retardation or learning and behavioural problems, is generally performed with the aid of a conventional test battery, of which the *Wechsler Intelligence Scale for Children-Revised* (WISC-R) constitutes a primary component (Raviv, 1988). In addition, a variety of group aptitude tests are used in the schools for research, evaluation, vocational guidance, and screening purposes.

Individual intelligence measures

The Israeli versions of the WISC-R. The standardization of the Israeli version of the WISC-R was begun in 1972 and was carried out on a nationwide representative sample of both Jewish and Arab populations. A total of 2700 children from 99 schools were individually tested in their native languages; the sample was found to be representative of the Israeli population with respect to sociocultural parameters. The Hebrew and Arabic versions of the WISC-R are used for diagnostic, guidance, and school classification purposes among normal, gifted, and retarded Israeli populations aged 6–16.

The Hebrew version of the WISC-R (Lieblich, Ben-Shachar, & Ninio, 1976) is currently the most widely used and perhaps best constructed diagnostic tool in Israel today for assessing Hebrew-speaking Israeli children's intelligence and cognitive difficulties. The guiding principle followed by the authors in the development of the Hebrew version of the WISC-R was to stay as close as possible to the original, unless there were strong reasons otherwise. Therefore, only a few changes were made in the general framework and structure of the tests, or the scoring system. The WISC-R consists of a Verbal (Information, $k=30$; Similarities, $k=16$; Arithmetic, $k=18$; Vocabulary, $k=24$; Comprehension, $k=17$; Digit Span, $k=14$) and Performance part (Picture Completion, $k=26$; Picture Arrangement, $k=12$; Block Design, $k=10$; Mazes, $k=7$; Digit Span, $k=93$; [Digit Span for children under eight, $k=45$]; and Object Assembly, $k=3$).

The Hebrew version of the WISC-R was normed on a representative sample of "normal" Israeli children with 100 children chosen from each of 11 age groups (6 to 16). About 8 to 16 children were tested per school (in grades 1 to 8), and children aged 12–16 were sampled from a list of elementary school graduates. The sample was about evenly distributed by sex. About 65% of the children were of Eastern extraction, 33% were of Western extraction, and about 2% third generation Israelis. Within each of the age groups, the Full Scale, Verbal, and Performance raw scores were standardized (following normalization) to a mean of 100 and SD of 15. The effective range of scores for Full Scale IQ was from 40 to 160 (about 4 SD s on both sides of the mean). The Digit Span and Object Assembly tests were not used in the calculation of scores.

As reported in the manual (Lieblich, Ben-Shachar, & Ninio, 1976), average internal consistency estimates for the WISC-R across 11 age groups

show coefficients as high as .96 for the Full Scale, .92 for the Performance, and .95 for the Verbal IQs. Subtest reliabilities are also satisfactory, ranging from .74 to .93. Stability estimates based on a small scale study among fifth grade children (for a 15-day lag) was over .90 for both the Full Scale score and the Verbal and Performance sections.

With respect to the test's construct validity, a number of studies have found the SSA space for the WISC-R to be three-dimensional and to conform to the lawfulness of a "cylinder" (Guttman & Levy, 1980; Lieblich, Ben-Shachar, & Ninio, 1976). As Guttman (1970) pointed out, if a battery of tests is constructed or selected according to three basic facets – language of presentation (verbal, numerical, geometrical); mental operation (rule inference, rule application, rule learning); and modality of expression (oral expression, manual manipulation, and use of paper and pencil) – the expected correspondence between the tri-faceted definitional system of intelligence and the empirical aspect of the correlation matrix is that of a cylinder (Levy, 1985). As reported by Levy (1985), this lawfulness has been replicated for WISC-R scores of children in Israel and the United States. Furthermore, a number of studies provide evidence for the diagnostic validity of the WISC-R profile in differentiating learning disabled from normal children aged 8 to 12 (e.g., Raviv, Margalith, Raviv, & Sade, 1981).

Group ability tests

The *Milta Group Verbal Intelligence Test* (Ortar, 1966) was for many years the most popular group verbal ability test administered in the school system. The test is modeled after the *Lorge-Thorndike Intelligence Test* and was designed as a group-administered verbal scholastic aptitude test for Hebrew-speaking students aged 9-18. It appears in parallel forms for three different grade categories (grades 4-6; 7-9; 10-12). The Milta subtest composition and number of items vary from form to form. The Milta total score is reported to be of acceptable reliability (about $r_{xx} = .90$) and has satisfactory criterion validity (r_{xy} of about .60) (Ortar, 1966).

Despite its lack of reliable Israeli norms, the *Raven Progressive Matrices* (1960) is probably the most widespread group measure of nonverbal ability used among Israeli children aged 8 to 13. The adequacy of the test's validity is generally deemed modest compared to verbal tests such as the Milta (cf. Zeidner, 1988b).

Vocational aptitude measures

In order to optimize the probability of a student's success in a particular track and to supply the school system with reliable information about

students' vocational aptitudes and interests, a good percentage of eighth and ninth grade Israeli students undergo vocational testing either at the Hadassah Guidance Institute, in Jerusalem, or at a number of other privately owned companies offering such services.

The *Hadassah Vocational Aptitude Battery* administered to eighth graders is typically composed of 12 tests assessing the following four domains: Verbal ($k=3$), Numerical ($k=3$), Figural Perception ($k=3$), and Psychomotor ($k=3$). The test battery given to ninth graders is composed of 14 tests along with measures of differential aptitudes (e.g., technical-mechanical, comprehension). The specific tests included in the battery are usually specifically tailored to the needs of the student or the school making the referral. The psychological narrative report, based on representative norms for each age group in the Jewish population, is generally relayed to the school career counsellor who interprets the results to the student and/or the student's parents. A number of large-scale prospective studies reported in the literature have shown that the aptitude test battery, when taken in grades 8-9, does a reasonably good job of predicting twelfth grade matriculation exam scores (Meier & Adler, 1985).

Perceptual tests

Based on the reports of expert educational psychologists (primary results of an international survey on testing) and the personal impression of the author, the *Bender Visual Motor Gestalt Test* (Bender, 1948) is often used for diagnostic, guidance, and placement purposes of Israeli children aged 4 to 6. The test is both group and individually administered and used for the full range of populations by schools, psychological clinics, and medical centres. Little systematic information is available regarding the test's psychometric properties or normative performance. Some psychologists also use the *Frostig Developmental Test of Visual Perception* (Frostig, Maslow, Lefevre, & Whittlesey, 1963) to help in the diagnosis of reading problems.

Aptitude tests in the works

A number of cognitive tests are currently in the process of being adapted and normed for psychoeducational diagnostic purposes in the Israeli school system. The *Kaufman Assessment Battery for Children* (K-ABC) (Kaufman & Kaufman, 1983) is in its early stages of adaptation and norming (Nevo, personal communication, June, 1989) as is the *Luria-Nebraska-Manitoba Neuropsychological Test Battery for Children* (cf. Rieck, 1988) and the *McCarthy Scales of Children's Abilities* (Kaufman & Kaufman,

1977; Miriam Rieck, personal communication, January, 1989). There are also plans for norming the Stanford-Binet. (The Hebrew translation of the Binet that does exist dates back to the 1960s and is devoid of reliable norms.)

Learning Potential Assessment Device (LPAD)

It would seem deficient to end this section without at least mentioning Feuerstein's interesting work in the development of dynamic learning potential assessment procedures. The LPAD procedure is based on a test-instruct-retest paradigm and has been used for assessing the abilities of culturally disadvantaged and learning disabled children in particular. This procedure aids at transforming the test situation from a static to dynamic one and from a product- to process-oriented one. The LPAD is offered as a viable alternative to conventional testing practices, designed to assess the educability and modifiability of an examinee's cognitive ability through active learning of thinking and test-taking strategies. Rather than focus on total scores alone, this procedure uses the peaks in the pattern of results as indicators of cognitive potential.

Based on the author's subjective impressions and knowledge of the Israeli testing context, the LPAD procedures are generally less well known than Feuerstein's internationally disseminated instrumental enrichment program (Feuerstein, 1980), and have yet to gain currency as accepted assessment procedures in Israeli psychoeducational assessment.

Scholastic Achievement Assessment

The largest national testing program ever implemented in the Israeli school system for any extended time period was the *National Achievement Test* (or *Seker*), used for guidance, resource allocation, and research purposes among eighth grade students; it was initiated in 1955 and discontinued in 1972. At present, standardized criterion- or norm-referenced achievement tests are not available, at the national level, to gauge Israeli students' attainments in various mastery domains. However, various achievement tests (e.g., Hebrew, reading comprehension, arithmetic) are administered on a partisan basis, particularly at the elementary school level. In addition, specific tests have been constructed for selection purposes by "absorbing schools," or prepared by municipal authorities for program evaluation and comparative purposes. Clearly, implementation of a large scale standardized achievement testing program at regular intervals in the schools would help improve the quality of student evaluation and allow systematic monitoring and tracking of students' academic performance.

As will be clearly evidenced in this section, a good deal of the evaluation research on student achievement has centred on sociocultural group

differences in performance. This concern surfaced in the early 1950s, when students of Eastern background were observed to fall below their Western counterparts in scholastic attainment at the very outset of their school experience, particularly in basic reading and language skills. Despite a wide array of social and educational programs designed to bridge the gap, the results have generally been disappointing and the differences have been very persistent over the years. For example, based on an index of "disadvantage" adopted by the Ministry of Education in 1974, about 44% of all Israeli students were defined as disadvantaged and, of these, 95% are of Eastern origin (Adler, 1984). It should be mentioned that a number of important internal publications of the Curriculum Division of the Ministry of Education on student achievement in various subjects (e.g., history, math, geography, Bible, writing proficiency, etc.), were unavailable to the author (at his sabbatical site at Stanford University) and will not be included in the survey.

Elementary school level

In the early 1970s a large scale study, modeled after the Coleman report in the United States (Coleman *et al.*, 1966), was launched to evaluate the scholastic achievements of Israeli elementary school children and identify student background and school variables that meaningfully contribute to success and failure in learning. The sample was composed of about 17,000 elementary school students – a representative sampling of 98 elementary schools in Israel – with particular focus on grades 1, 2, 4, and 6. Criterion-referenced achievement tests were composed in line with the demands of the specific curriculum for various subjects in each grade level. The items reflected both optimal and minimal demands of the curriculum, along with anchor items to allow comparison across grade levels. Data were collected on achievement in language, mathematics, science, geography, and Bible.

The highest mean achievement level observed, across grades, was in language; the lowest achievement level was in geography; and moderate degrees of success were reported for math, science, and Bible. Depending on the subject, low levels of student achievement were variously attributed to deficiencies in teaching methods (e.g., geography); lack of adequate facilities for experiments (science); complexity of material (e.g., math), and excessive amounts of material to be covered (e.g., Bible). A consistent rate of progress was discernible on all achievement tests, though significant differences existed between subjects, grade levels, and student groups.

Compared to the serious reading failures reported among disadvantaged students in the 1950s, the results of the language tests administered show that this particular problem has been more or less resolved. Further-

more, in contrast to the one *sigma* unit ethnic group differences found on the *Seker* qualifying exams in the 1950s and 1960s (Ortar, 1967), this report shows differences between Eastern and Western students ranging between .70 and .80 *SD*, implying a small decrease in the ethnic group disparity in achievement. Furthermore, no evidence emerged from this study to support the notion of a cumulative deficit in the achievement gap between Western and Eastern students or children of varying SES groups. Although cultural background and generation in Israel both had a significant relationship with student performance, the influence of cultural background was greater than that of the generational effect.

Comparable to the American results (Coleman *et al.*, 1966), more than one-third of the between-class variance in achievement is attributable to pupil background and less than one-tenth to schools and classroom variables. The data suggest that although the schools have contributed to the advancement of underprivileged students, the contribution is less than what the educational system could achieve and needs to achieve (Minkowitch, Davis, & Bashi, 1982).

The foregoing data are consistent with those collected as part of an international study on school achievement sponsored by the **International Association for the Evaluation of Educational Achievement** (Lewy, Rapaport, & Rimor, 1978). Meaningful differences were found between students aged 10 to 11 of Eastern and Western background – in favour of Western students – in reading comprehension, civics, and English. However, the absolute levels of achievement of Eastern students were closer to the norms of students in industrially developed than developing countries in the survey. (Meaningful ethnic group differences in student achievement in the same direction were also observed for students aged 14-15 and high school seniors). Furthermore, based on a stratified random sample of 69 elementary schools including about 3600 students, fourth to sixth grade students of Eastern extraction were reported to learn at a slower rate than their Western counterparts (Lewy & Chen, 1977), with a one- to two-year difference in achievement among the groups. Students of European origin in all three grades (4-6) consistently scored higher than their Eastern counterparts in the school subjects assessed (e.g., math, reading, English, geography), with close to a standard deviation discrepancy among the groups. However, the ethnic gap in educational achievement was not found to be cumulative and did not increase as students progressed from grades four to six.

Junior high school level

Similar to what was reported for elementary school students, the analysis of the qualifying *Seker* test results for 27,000 pupils in 1962-63 showed that the scores of European students were significantly higher than

that of Eastern students in the range of about one *SD* on all subtests (cf. Smilansky & Yam, 1969; Ortar, 1967). About 50% of Eastern students did not even reach the test's minimal requirement (set at 60) and only 15% passed the test at a score of 80. By comparison, only 10% of Western students received a grade lower than 60 while 40% managed to pass the test with a score of 80 or above.

The *Mechkar Ha-Chativot* junior high school evaluation study (Chen *et al.*, 1978) set out to evaluate the structural reform in the Israeli educational system in its transition from the traditional eight-year elementary school (8:4) to the modern six-year elementary school and three-year junior high and high school (6:3:3). The study was based on a sample of 3200 students drawn from 14 junior high schools, with 12 achievement tests in key subjects administered in each grade level. The test reliabilities ranged from .82 to .93.

On the whole no substantial differences were found in student progress for junior high school and traditional eight-year elementary schools, nor was the junior high school particularly advantageous to the scholastic performance of disadvantaged students. Levels of student attainment in a variety of subjects were disappointing, with only about 40 to 70 percent of the students reaching some reasonable level of mastery (set at 84%). Similar to the data generated by the *Seker*, about a two-year difference in achievement of students of European and Eastern origin was reported. Accordingly, students of Eastern background at the end of the ninth grade achieved at about the same level as Western students at the end of the seventh grade. Furthermore, about 50% of Western students were found to be in the highest tracks and ability groupings in key subjects compared to only about 28 – 30% of Eastern students.

With respect to gender differences in scholastic achievement, additional analysis based on the *Mechkar Ha-Chativot* junior high school evaluation data base (Kfir, 1988) showed that girls have no serious advantage over boys in objective achievement at the beginning of junior high school (grades 7-8). However, girls in each ethnic group succeeded more than boys when judged by other criteria, such as teacher evaluation in grades 7-9, higher ability group placements in junior high, and so on.

These results are partially consistent with those of another study (Dar, Resh, & Erhard, 1989) based on a national sample of about 9000 eighth and ninth graders and designed to explore gender and SES group inequalities in science and reading comprehension achievements. Adolescent boys revealed a small advantage over girls particularly in science – anywhere from .20 to .25 *SD*. The gender group differences in favour of boys was greater in grade

9 than in grade 8 and in understanding than in informational knowledge. The results bearing on ethnic group differences are consistent with previous data (though somewhat less extreme) showing an achievement gap of about .67 *SD* between students of Western and Eastern descent (Dar, Resh, & Erhard, 1989). Comparable to what was found in the Minkowitch report (Minkowitch, Davis, & Bashi, 1982), the achievements of mixed-origin students were closer to those of western origin, with the student's generation in Israel having a positive effect on achievement. SES was shown to have a stronger relationship with achievement than either ethnic origin or generation in Israel, with a gap of about one *SD* between upper and lower SES students. Multivariate analysis showed that SES mediates a considerable part of the ethnic and generational effects on achievement (Dar, Resh, & Erhard, 1989).

High school level

Matriculation exams have become the primary basis for evaluating achievement in Israeli high schools. They are required for high school credentials and university admissions in Israel and are also used for monitoring of curricula and teacher accountability as well as for research purposes (Nevo & Ben-Shakhar, 1985). Tests are given in a variety of school subjects (English, math, Hebrew, Bible, history), with each test assessed by two independent judges. The interjudge reliability of the matriculation tests varies by test content and has been reported to range anywhere from about .50 to .90 (Nevo & Ben-Shakhar, 1985). Matriculation exams are often preferred to teacher grades as achievement measures since the latter are typically less reliable, are difficult to interpret, and need to be calibrated on a common scale across schools.

Although a substantial ethnic gap still persists in matriculation exam success, it appears to be narrowing. In 1956-57 the ratio of Eastern to Western students who obtained matriculation diplomas was about one in three; by 1976-77 it was about 3 in 4. In the 10-year period from 1966-77, Eastern students who successfully passed the matriculation exam (from 17-year cohort) doubled in these years from 3.7% to 7.4%. In 1972-73, 17.4% of Eastern and 31.7% of Western student 17-year-olds obtained matriculation diplomas (as reported by Adler, 1984). On the whole, a survey of the studies available (Adler, 1984) suggests that the educational gap between sociocultural groups in Israel, despite its persistence, is in flux and heading toward a substantial decrease in the future.

A large scale evaluation study (Kfir, 1988) reports that Western students at ages 17-18 perform better than their Eastern student counterparts, both with respect to objective achievement measures, as well as number of

years of schooling and teacher evaluations. Furthermore, with respect to gender group differences, Kfir (1988) found that 18-year-old adolescent girls enjoyed an advantage over boys in the number of years of schooling, curricular track, ability group, school grades, and teacher evaluations, whereas boys had an edge on objective achievement tests.

Assessing achievement in the Arab sector

At present, there are over 1,350,000 children in the Israeli school system – about 200,000 of whom are in the Arab sector (Raviv, 1989). Israeli Arabs are a minority group in the true sense of the term, comprising over 18% of the total population and generally socially disadvantaged relative to their Jewish counterparts (Smoocha & Hofman, 1976/77). Jews and Arabs differ on a wide array of sociocultural parameters, including mother tongue, religion, nationality, norms, customs, and family structure (Smoocha & Hofman, 1976/77). Due to the wish of both Jews and Arabs in Israel to preserve their own cultural symbols, values, and norms, the two educational systems are kept apart. Arab schools, as a rule, have suffered from a wide array of deficiencies, including understaffed personnel, poor facilities, ill-trained teachers, and poor parental support – all interacting to contribute to the generally low levels of student achievement in the Arab sector (Bashi, Kahn, & Davis, 1981).

A comprehensive survey conducted in 1974 showed that in several areas of science and math, the average achievement of Arab elementary school students was substantially lower than that of Jewish students (Bashi, Cahan, & Davis, 1981). Interestingly, the average attainment of Arab students is remarkably similar to that of lower class first-generation Eastern Jewish students, particularly in geography and science. Although the formal curriculum is basically the same in a number of subjects (e.g., math, English, sciences), due to major differences between the Jewish and Arab school systems in language of instruction and certain curricular elements, serious limitations arise in attempting to compare student achievements in the two sectors.

Furthermore, only 35% of the Arab students who sat for matriculation exams in 1973 passed, compared with 50% of the Jewish candidates. However, in recent years, the success rate of Arab candidates in matriculation exams has been steadily going up. For example, in 1983 most Arab students who took higher level math in matriculation exams (4-5 units) passed (Saad, 1984). Comparably, a recent comparison of matriculation scores for 1778 Jewish and 1017 Arab student candidates in Israel showed an average difference of only about .25 *SD* across the subjects: math, English, and Hebrew languages (Zeidner, 1987). Also, results of the biology matriculation exams for 1983 and 1984 imply that while Jewish students obtain significantly

higher scores on the exams, the gap is narrowing, especially in tasks which require lower cognitive abilities or applications of inquiry skills (Tamir, 1986).

Educational integration and bridging the gap

Any extensive survey of the vast Israeli literature dealing with the effects of ethnic integration reforms and experimental projects on students' achievements and morale is beyond the scope of this study. (The interested reader is referred to Adler [1984].) However, it should be mentioned that, as in the United States, virtually all the Israeli summative evaluations of desegregation lead to the conclusion that at the microlevel of analysis, desegregation in itself yields little, if any, positive psychological, social, or academic results for minority pupils in the first few years following implementation (Chen, Levy, & Adler, 1978). By the same token, studies in Israel (Gutman, 1972) have shown that ability grouping does not produce a major beneficial effect for the achievement of disadvantaged children. However, Dar & Resh (1986) recently reported that classroom intellectual composition positively affects students' academic achievement, with low-resource students more sensitive than high-resource students to class compositional quality.

Personality Testing

A variety of projective measures have traditionally been part and parcel of the psychodiagnostic arsenal of Israeli school psychologists, and various standardized inventories have been employed to assess students' basic personality constructs for research and evaluation purposes. However, based on face value, many of these instruments have been indiscriminately adapted from abroad and would fail to meet systematic checks for: (a) indigenous Israeli cultural context; (b) tests for differential item-functioning; (c) tests for satisfactory validity and reliability on Israeli samples; and (d) adequacy and representativeness of available norms.

Commonly used projective measures

School psychologists in Israel administer a variety of individual projective measures for diagnostic, screening, and clinical purposes among both normal and clinical school-aged populations. Two of the more popular measures are the *Thematic Apperception Test* (TAT) (Murray *et al.*, 1938) and the *Children's Apperception Test* (CAT) (Bellak, 1975). Whereas the TAT is typically used for diagnostic and guidance purposes in Israel among both normal and disturbed children aged 8 to 16, the CAT is used primarily for younger children aged 5 to 10 for placement, guidance, and diagnostic purposes. There is little systematic data bearing on either test's psychometric

properties (e.g., reliability, generalizability, and construct or predictive validity), and both tests lack representative national norms.

The *Rorschach* (1942) is an individually administered projective measure used by school and clinical psychologists among children aged 5 to 16. The test is used for placement, guidance, and diagnostic purposes among both normal and disturbed children in the schools as well as in psychological centres and medical clinics. Here again, there is little evidence for the test's reliability or validity among national representative groups of school-aged children. In addition, various versions of the *Draw-a-Man-Test* and *Sentence Completion* projective measures are also used.

The *Bar-Ilan Picture Test for Children* (Itskowitz & Strauss, 1982) is one of the few examples of a semi-projective personality test both constructed and validated with Israeli children in mind, and was designed to elicit children's attitudes about roles in class and peer group. Information based on both test-retest and internal-consistency reliability indices show satisfactory reliability coefficients and a number of construct validity studies support the test's inferences in various areas including school anxiety, preparation for school entry, and family communications (Itskowitz & Strauss, 1982).

Standardized paper and pencil measures

The past decade has witnessed a flurry of Hebrew versions of standardized paper and pencil inventories for assessing key personality constructs adapted to the Israeli scene (e.g., *State-Trait Anxiety Inventory*, *Test Anxiety Inventory*, *16PF*). Unfortunately, a detailed survey of the various measures is beyond the scope of this study.

Sociocultural and Gender Group Differences in Cognitive Abilities

Overall, research in Israel shows that the social and educational environment throughout the school years appears to be more beneficial to the intellectual development of Western, middle-class, and male students compared to their respective Eastern, lower-class, and female student counterparts. The concern of Israeli educationists and psychologists surrounding group differences in ability and achievement is understandable in view of the acknowledged importance of cognitive performance for future academic success and social mobility in modern Israeli society, coupled with the national goal of fully integrating culturally different groups in Israel by providing them with equal educational and social opportunities. Indeed, the

documented differences in sociocultural and gender group attainment are in direct contrast to the basic values of Israeli society – first and foremost those of equality and national identity (Adler, 1984) and its common ideology which views all children as having equal potential and deserving equal treatment (Blass & Amir, 1984).

Sociocultural differences in the Jewish sector

Major sociocultural group disparities have been documented in mean levels of aptitude test performance from the very inception of intelligence testing in Israel in the late 1940s (cf. Ortar, 1953). Numerous studies over the past 40 years continued to report marked differences between Jewish examiners of Western and Eastern (Asian/African) background in intelligence test performance. The group disparities range anywhere from about one-half to one-and-half standard deviations and are documented at both the preschool (Lieblich, 1983), school (Lieblich, 1983; Minkowitch, Davis, & Bashi, 1982; Ortar, 1963; Zeidner, 1985), and university level (Zeidner, 1987).

Although there is little evidence for meaningful variations in the profiles of multiple-ability measures, by cultural group (Lieblich, 1983), Eastern examinees generally score further below their Western counterparts on verbal relative to nonverbal measures of ability (Lieblich, 1983; Minkowitch, Davis, & Bashi, 1982; Zeidner, 1985). Furthermore, middle-class children are reported to outperform their lower-class counterparts in total WISC-R performance by about .5 *sigma* units – in each ethnic group (Lieblich, 1983). A similar trend has been reported for group verbal ability tests (cf. Zeidner, 1985, 1988a). The IQ difference among SES groups starts out at about a .7 *SD* discrepancy at ages 6-7 and increases to about a 1.1 *SD* discrepancy at ages 15-16. Thus, there appears to be a cumulative increase of IQ difference on the WISC-R with age for middle- and lower-class students. By the same token, there appears to be a cumulative increase in ethnic group difference with age, rising from about .75 *SD* at ages 6-7 to a discrepancy of about two *SDs* at age 15-16. Ethnic group is reported to have a more sizable effect than SES on students' WISC-R performance, though there is a tendency for the ethnic gap to narrow among second generation Israelis.

The observed group differences have been attributed mainly to socio-cultural group differences in socialization patterns, home language, and differential schooling experiences. Furthermore, in view of the relatively more marked cultural group disparities commonly found on verbal relative to nonverbal tests, a number of Israeli scholars have argued that verbal deficit may be the main locus of the poor showing of minority groups on psychometric ability tests (Feuerstein, 1980; Minkowitch, Davis, & Bashi, 1982).

Sociocultural differences in the Arab sector

Partly because of difficulties involved in adapting conventional tools to the Arabic language and culture (Lieblich, 1983), and partly due to political considerations (Saami Mari, personal communication, May 1983), research on cognitive abilities and achievements in the Arab sector has been relatively sparse over the years. For many years psychoeducational assessment tools were simply not available for purposes of diagnosis, selection, or educational placement and intervention in the Arab sector.

In the 1970s a number of pioneering attempts were made to adapt and develop instruments for studying the cognitive development of the Arab child. In one study (Bashi, 1976; cf. Bashi, Cahan, & Davis, 1981), a subtest of *Raven Matrices* as well as the test of "associative ability" constituted part of the test battery administered to a representative group of Arab school children. Analysis of group differences at various grade levels showed that religion has a consistent effect on test performance, with Christian students performing meaningfully higher than their Moslem or Druze counterparts. Also, boys scored higher than girls on both tests and higher SES groups generally performed better than their lower SES counterparts. Religious affiliation was a better predictor of achievement than SES or sex in the Arab sector.

Overall, the few studies examining the effects of religious subgroup affiliation on intelligence test scores within the Arab population (Bashi, 1976; Lieblich, 1983; Lieblich & Kugelmass, 1981) tend to concur that Christian Arabs outscore their Moslem and Druze counterparts, whereas the latter two groups are not reliably differentiated in mean test performance. However, much like in the Jewish sector, subgroup affiliation is correlated with social background, with most of the Moslems belonging to the lower class and the Christians to the upper class. Indeed, the rank order of groups by religion conforms to what would be expected on the basis of the respective groups' socioeconomic status. In a large-scale study (Lieblich & Kugelmass, 1981) in which the WISC-R was specifically adapted to the Arab culture and standardized on a representative group of Arab children, Arab students evidenced significantly lower WISC-R total scores (about one *sigma* unit) compared to a Jewish comparison group; cultural group mean differences were greater on the performance than verbal section. In addition, several other studies conducted over the past 20 years have shown that Israeli Jewish groups outscore their Arab counterparts by about one standard deviation on standardized tests of intelligence and scholastic ability at the preschool (Kugelmass, Lieblich, & Bossik, 1974; Lieblich, 1983) and college entrance level (Zeidner, 1988b) as well.

Gender group differences

Israeli boys and girls appear to enter elementary school similarly equipped with necessary intellectual functions, as evidenced by nonsignificant gender group differences in intelligence test scores at the preschool level (Lieblich, 1983; 1985; Safir, 1986). However, boys evidence consistent superiority on intelligence tests from about the third grade, with boys beginning to outscore girls at age 9. The gender group differences grow through age 16, culminating in a 12-point difference in IQ. Similar trends are observed for both the Verbal and Performance sections. Remarkably parallel findings are reported for the Arab sector (Lieblich, 1985), with differences of about a third *SD* in favour of adolescent boys in total IQ. At age 16, the differences between gender groups are maximal – at about 10 IQ points (Lieblich, 1985).

As Safir (1986) has pointed out, gender group differences in IQ in Israel are considerably greater than in the United States from an early age on. Significant gender group differences are evident on both verbal and analytical tests among both Jewish and Arab college candidates as well (Zeidner, 1986a, 1986b). In fact, sex is the third demographic factor in magnitude of effect on intelligence in both the Jewish (ethnicity > SES > sex) and Arab (religion > SES > sex) population. These data are in contrast to much previous research in the West showing that adolescent boys are usually better in spatial and numerical abilities, while girls are superior in verbal abilities. These reported gender differences in ability are generally attributed to gender differences in life experiences, including: unequal intellectual opportunities; differential expectations of parents and teachers for boys and girls; differential peer group pressures; and exposures to different role models (Lieblich, 1985; Safir, 1986).

Test Bias and Equity

The widespread reliance on measures of scholastic ability for assessment and placement of students from diverse cultural backgrounds in the Israeli school system has raised serious concern regarding cultural test bias and fairness (Zeidner, 1987; 1988a; 1988b). The recent Israeli “antitest movement” and public outcry is noteworthy, since the chronology of litigious events closely parallels those leading up to the enactment of PL 94-142 in the United States. Indeed, elements in the anti- and pro-testing debate are nearly identical to those heard in the American educational system over the last 20 years. Among the striking cross-cultural parallels in the Israeli and American antitest movement are:

(a) *Antecedents of public concern with standardized testing.* These are: sizable sociocultural differences in aptitude and achievement test scores; popularity of the cultural difference position in the mid-1970s (Stahl, 1977); contentions of test bias, and adverse impact for cultural minority groups in the population; severe overrepresentation of lower-class Eastern children in special education programs and low-status tracks; and the sequence of Israeli attempts to deal with ethnic and SES divergence in abilities and school performance at a structural level (Miller, 1984).

(b) *Manifestations of antitest sentiments.* These are: vehement attack on scholastic aptitude tests in the media; appeal to the courts by concerned parents to ban the usage of aptitude tests in elementary schools for student selection and classification purposes (Cahan, 1986); an Israeli version of the "antitest syndrome" (cf. Jensen, 1980) surrounding the testing melodrama, characterized by an intense emotional involvement in the test debate on the part of both test protagonists and antagonists, contamination of reason by affect and ideology, and armchair speculation concerning the degree of bias and equity in testing (cf. *Israeli Psychological Association Bulletin*, 1986).

(c) *Consequences of the antitest campaign.* These include: a professional committee mandated to examine the goals, usages, problems, costs, and ethical implications of group ability test usage in the schools (Ministry of Education, 1985); ban of widespread and massive administration of group ability tests for selection and placement purposes in elementary schools by the Ministry of Education as of January 1985.

(d) *Results of test-bias research.* Research in the Israeli scene setting out to empirically test for cultural fairness in the predictive validity of both verbal and nonverbal tests for majority and minority group (Zeidner, 1986b) lend generalizability to much previous research conducted in the United States, providing some evidence of intercept bias (scores were overpredictive of minority group students) and little evidence in predictive slope bias by sociocultural group. Furthermore, a series of tests of the "situational bias" hypothesis conducted by this author among school children revealed that sociocultural group membership does not interact with test atmosphere (Zeidner, 1985) or with examinees' test attitudes and dispositions (Zeidner, 1988a) in affecting ability test performance. In view of the Israeli evidence, the cultural bias hypothesis is not vindicated.

Rather than acknowledge the persistence of sociocultural group differences in test performance and focus additional efforts on raising the level of minority group performance, many simply preferred to "kill the messenger" by attacking the validity and equity of the tests themselves (cf. Lerner, 1989).

Future Trends and Directions

Future developments

This review has shown that notwithstanding remarkable advances and developments in the field of school-based assessment in Israel over the years, there is much left to be desired. The current state clearly reveals the need for a major and concerted effort on the part of Israeli psychologists and measurement experts in improving both the quantity and, even more important, the quality of school-based measurement instruments and procedures.

Additional school-based assessment instruments are needed for diagnostic and evaluation purposes for a host of special student populations characterized by learning disabilities, attention disorders, mental retardation, emotional disturbance, physical disorders, and intellectual giftedness. Among the school-based measures that appear to be most desperately needed are individual and group tests of intellectual, perceptual, and emotional development, not to speak of standardized norm- and criterion-referenced measures of achievement.

The Israeli school system needs to direct systematic effort towards the development of criterion-referenced measures of achievement for the various school subjects taught in the Israeli school system — all the way up from elementary to high school. These measures would preferably include items indicative of both minimal and optimal levels of mastery (for formative and summative evaluation purposes) as well as anchor items for future tracking of student achievement across time (cf. Minkowitch, Davis, & Bashi, 1982). In particular, there is a crying need for developing standardized achievement tests in the various school subjects, at all grade levels, for use in the Arab sector.

Among the foreseeable developments and trends in assessment practices as a consequence of more global and general developments in school-based assessment are: decreased emphasis on norm-referenced measurements and increased emphasis on criterion-referenced achievement tests for student assessment and placement decisions; multifaceted assessment and use of a greater variety of assessment tools — supplementing standardized tests with other measures (interviews, direct observations); increased use of tests for program evaluation and decision-making purposes and explicit consideration of the costs of various types of errors in the decision-making process based on tests (e.g., banning ability tests for selection purposes may decrease the number of false negatives, but increase the number of false positives); and increased use of behavioural assessment technologies to

assess learning problems and to monitor progress in educational programs and school settings.

In addition, more emphasis will probably be placed on establishing and implementing elements of sound and ethical testing practices – a relatively neglected issue in Israeli testing-practice in the past. In specific, considerable attention will most likely be given to the development of ethical testing standards; specification of technical guidelines for test development; guidelines for test administration and usage; credentials and licensure; implementing informed consent procedures; securing test data and storing test material; and confidentiality of test or assessment information. Although very little *a priori* thought has been given to systematically establishing testing standards and ethical codes for test construction, usage, interpretation, and decision making, this situation is fortunately changing. The “Nevo Committee” (Nevo, 1988), mandated by the Ministry of Education, has recently suggested some very important, timely, and practical suggestions towards improving testing and selection practices in the schools.

Furthermore, school-based assessment in Israel would most likely profit from the application of recent advances in testing technologies, including item-banking of ability and achievement test items; computer adaptive testing programs; computerized test scoring and interpretation (with some headway already being made in the area of mathematics testing); and the use of modern test theory in the construction and scaling of items (e.g., logistic models and item-response theory and techniques).

Needed research

Future developments in testing and assessment in Israel will most likely be accompanied by systematic research in a wide variety of areas. Some of the potential areas for much needed school-based assessment research are (not necessarily in order of importance): (a) the feasibility of item-banking and computerized testing, scoring, and interpretation for school-based aptitude, achievement, vocational, and personality testing purposes; (b) examination of student trajectories and growth curves in achievement, abilities, and school-related personality variables over time; (c) psychometric properties of school-based instruments administered conventionally and by electronic means; (d) test bias for special subgroups in the population (e.g., racial, gender, handicapped), including tests for invariance of tests' structural components, differential item functioning, and differential predictive validity; (e) examinee test attitudes and dispositions; (f) effects of coaching on standardized aptitude tests; (g) contextual and personal variables in the assessment situation (test anxiety, test motivation, self-efficacy);

(h) test-taking strategies and skills of various subgroups in the population; (i) cognitive skill analysis of test performance; (j) specific strengths and weaknesses in test-taking and performance in various subgroups in the population; and (k) specific loci of low test performance of ethnic minority groups.

Concluding remarks

After an initial spate of enthusiasm and widespread adoption of a variety of tests in Israel during the 1960s and 1970s, the antithetical reaction to testing appeared in full blown force in the 1980s. According to the Hegelian triad, we are now bound to reach some sort of resolution and arrive at a consensus with respect to valid test usage or alternative assessment practices in the schools in order to secure a satisfactory synthesis between the opposing test and antitest forces at play.

In the first few decades of its existence, the Israeli educational system has almost miraculously managed to come to grips with pressing problems of both a quantitative nature (i.e., accommodating the waves of immigrant children and helping them assimilate to Israeli society) and of a qualitative nature (raising minimum levels of achievement of disadvantaged groups). The future demands a concerted effort at raising the level of functioning and standards of the educational system as a whole. A major effort directed at upgrading and improving school-based assessment procedures may play a key role in helping to achieve this important national goal.

Assessment tools and procedures used for educational decision-making purposes have important pedagogical, social, and economic implications for social systems (cf. Wainer & Braun, 1988). However, given the uniquely vulnerable and particularly difficult economic, social, political, and military situation of modern-day Israel, the importance of valid educational assessment, facilitating the optimal utilization of Israeli society's human resources, indirectly affects the welfare of the nation. Thus, the importance of valid assessment tools goes beyond their importance in sheer maximization of utility for the decision-maker; it may have critical consequences for the future social, economic, and military development and very survival of Israel in the years to come.

NOTE

This paper was completed while the author was on sabbatical at the Program for Psychological Studies, School of Education, Stanford University. The help of the host university in facilitating the completion of this manuscript is gratefully acknowledged. I would like to thank Thomas Oakland of the University of Texas at Austin for making available to me relevant Israeli research data gathered as part of a cross-national study of assessment practices and procedures. In addition, I would like to express my thanks to the following individuals for providing me with relevant documents and materials in the course of preparing this manuscript at my sabbatical site abroad: Surel Cahan and Uriel Last of the Hebrew University of Jerusalem; Baruch Nevo and Miriam Rieck of the University of Haifa; Rivkah Itskowitz of Bar Ilan University; and Elchanan Meier and Amiram Raviv of Tel-Aviv University. Thanks are also due to Dan Inbar of the Hebrew University of Jerusalem, Aryeh Lewy of Tel-Aviv University, Ozer Schild of Haifa University, and Yoav Cohen of the National Institute of Testing and Evaluation in Jerusalem for their useful comments and suggestions on an earlier draft of this manuscript. The responsibility for the content of this paper, however, lies with the author alone.

Correspondence regarding this manuscript may be addressed to: Moshe Zeidner, Ph.D., School of Education, University of Haifa, Mt. Carmel, 31999, Israel.

This paper was presented at the 97th Annual Convention of the American Psychological Association at New Orleans, Louisiana, August, 1989.

REFERENCES

- Adler, Ch. (1984). School integration in the context of the development of Israel's educational system. In Y. Amir & S. Sharan (Eds.), *Social desegregation: Cross-cultural perspectives* (pp. 21-46). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Bashi, Y. (1976). *Verbal and nonverbal abilities of students in grades four, six and eight in the Arab sector*. Jerusalem: School of Education, Hebrew University.
- Bashi, Y., Cahan, S., & Davis, D. (1981). *Achievement of Arab pupils in the elementary schools in Israel*. Jerusalem: School of Education, Hebrew University.
- Bellak, L. (1975). *The TAT, CAT, and SAT in clinical use* (3rd ed.). New York: Grune & Stratton.
- Bender, L. (1948). *Manual for instruction and test cards for Visual and Motor Gestalt Test*. New York: American Orthopsychiatric Association.
- Blass, N., & Amir, B. (1984). Integration in education: The development of a policy. In Y. Amir & S. Sharan (Eds.), *Social desegregation: Cross-cultural perspectives* (pp. 63-98). Hillsdale, NJ: Lawrence Erlbaum Associates.

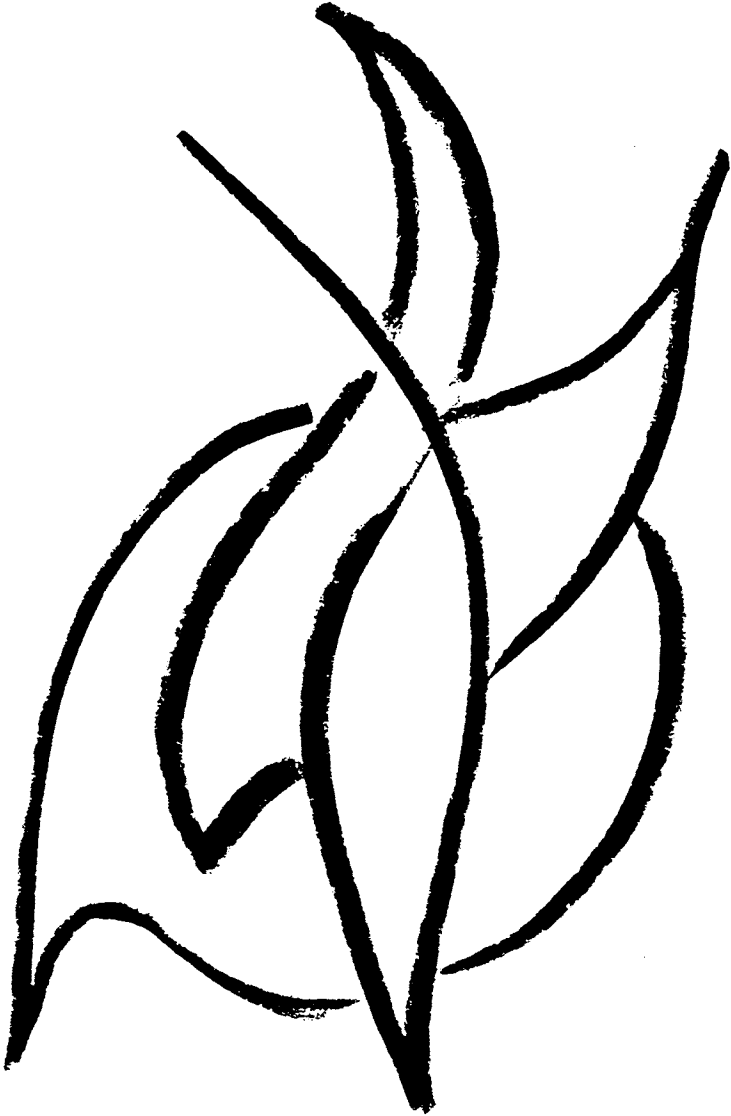
- Cahan, S. (1986). *The role of ability tests in educational decision-making: The case of tracking and assignment to homogeneous homeroom classes in the Israeli junior high school*. Paper presented at the 21st International Congress of Applied Psychology. Jerusalem, Israel, July 1986.
- Chen, M., Lewy, A., Adler, C., Inbar, D., Kfir, D., & Resh, N. (1978). *Process and result in education: Evaluating the contribution of junior high school to the educational system*. Tel-Aviv & Jerusalem: Tel-Aviv University and Hebrew University (in Hebrew).
- Coleman, J.S., et. al. (1966). *Equality of educational opportunities*. Department of Health, Education and Welfare, Office of Education, Government Printing Office, Washington, DC.
- Dar, Y., & Resh, N. (1986). Classroom intellectual composition and academic achievement. *American Educational Research Journal*, 23, 357-374.
- Dar, N., Resh, N., & Erhard, R. (1989). *Learning achievements in junior high schools in reading comprehension and science*. Jerusalem: School of Education (in Hebrew).
- Frostig, M., Maslow, P., Lefevre, D.W., & Whittlesey, J.R.B. (1984). The Marianne Frostig Development Test of Visual Perception. 1963 Standardization. *Perceptual and Motor Skills*, 19, 463-499.
- Feuerstein, R. (1979). *The dynamic assessment of retarded performers*. Baltimore: University Park.
- Feuerstein, R. (1980). *Instrumental enrichment*. Baltimore: University Park.
- Gutman, I. (1972). *An experimental investigation of the effects of various grouping methods on the cognitive and psychosocial development of elementary school pupils*. Jerusalem: Szold (in Hebrew).
- Guttman, L. (1970). Integration of test design and analysis. In *Proceedings of the 1969 Invitational Conference on Testing Problems*. Princeton, NJ: Educational Testing Service.
- Guttman, L., & Levy, S. (1980). Two structural laws for intelligence. *Megamot*, 25, 421-438 (in Hebrew).
- Hadassah Vocational Guidance Institute. (1982). *Bulletin of the Guidance Department* (in Hebrew).
- Itskowitz, R., & Strauss, H. (1982). *The Bar-Ilan Picture Test for Children* (Revised Ed.). Copenhagen, Denmark: Dansk psykologisk Forlag.
- Israeli Psychological Association Bulletin*. (1986). February issue (in Hebrew).
- Jensen, A. (1980). *Bias in mental testing*. New York: Free Press.
- Kaufman, A.S., & Kaufman, N.L. (1977). *K-ABC: Kaufman Assessment Battery for Children*. Circle Pines, MN: American Guidance Service.
- Kfir, P. (1988). Achievements and aspirations among boys and girls in high school: A comparison of two Israeli ethnic groups. *American Educational Research Journal*, 25, 213-236.
- Kugelmass, S., Lieblich, A., & Bossik, D. (1974). Patterns of intellectual ability in Jewish and Arab children in Israel. *Journal of Cross-Cultural Psychology*, 5, 184-198.
- Lerner, B. (1989). Intelligence and law. In R.I. Linn (Ed.), *Intelligence: Measurement, theory, and public policy*. (pp. 172-192). University of Illinois Press.

- Levy, S. (1985). Lawful roles of facets in social theories. In D. Canter (Ed.), *Facet theory: Approaches to social research* (pp. 59-96). New-York: Springer-Verlag.
- Lewy, A., & Chen, M. (1977). Differences in achievement: A comparison over time of ethnic group achievement in the Israeli elementary school. *Evaluation in Education, 1*, 3-72.
- Lewy, A., Rapaport, Ch., & Rimor, M. (1978). *Educational achievement in the Israeli school system*. Tel Aviv: Ramot (in Hebrew).
- Lieblich, A. (1983). Intelligence patterns among ethnic and minority groups in Israel. In M. Nisan & U. Last (Eds.), *Between education and psychology* (pp. 335-357). Jerusalem: Magnes Press (in Hebrew).
- Lieblich, A. (1985). Sex differences in intelligence test performance of Jewish and Arab school children in Israel. In M. Safir, M.T. Mednick, D. Israeli, and J. Bernard (Eds.), *Women's worlds: From the new scholarship*. New York: Praeger.
- Lieblich, A., & Kugelmass, S. (1981). Patterns of intellectual ability of Arab school children in Israel. *Intelligence, 5*, 311-320.
- Lieblich, A., Ben-Shachar, S., & Ninio, A., (Eds.). (1976). *Manual WISC-R: Wechsler Intelligence Scale for Children*. Revised. Institute for Psychological Development of the Hebrew University and Psychological Advisory Service of Ministry of Education and Culture, Jerusalem.
- Meier, E., & Adler, N. (1985). Prediction of success in matriculation exams. *Studies in Education, 41*, 35-50 (in Hebrew).
- Miller, N. (1984). Israel and the United States comparisons and commonalities in school desegregation. In Y. Amir & S. Sharan (Eds.), *Social desegregation: Cross-cultural perspectives* (pp. 237-253). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Ministry of Education. (1985). *Report of the Rashi Committee*. Jerusalem: Government Printing Office (in Hebrew).
- Minkowitch, A., Davis, D., & Bashi, Y. (1982). *Success and failure in Israeli elementary education*. New-Brunswick, NJ: Translation Books.
- Murray, H.C. et al. (1938). *Explorations in personality*. Cambridge, MA: Harvard University Press.
- Nevo, B. (1987). *Report of the Committee for Formulating Recommendations for Classifying Students in High School*. Jerusalem (in Hebrew).
- Nevo, B. (1988). Tests in Israel and the US: The state of the art. In B. Nevo & Y. Cohen (Eds.), *Selected problems in evaluation and measurement*. Jerusalem: National Institute of Measurement and Evaluation (in Hebrew).
- Nevo, B., & Ben-Shakhar, G. (1985). Interjudge reliability of matriculation exams. *Megamot, 29*, 42-56 (in Hebrew).
- Ortar, G. (1953). The diagnostics of the Wechsler exam among individuals with different intelligence levels. *Megamot, 3*, 199-216 (in Hebrew).
- Ortar, G. (1963). *Is a verbal test cross-cultural? Scripta Hierosolymitana* (pp. 219-235). Jerusalem: Hebrew University.
- Ortar, G.R. (1966). *Mita Intelligence Test*. Jerusalem: Hebrew University

- Ortar, G. (1967). Thirteen years of "Seker": Student achievements. *Megamot*, 15, 220-230.
- Raven, J.C. (1960). *Guide to using the Standard Progressive Matrices*. London: Lewis.
- Raviv, A. (1988). *School psychology research in Israel*. Unpublished Draft Manuscript.
- Raviv, A. (1989). School psychology in Israel. In P. A. Saigh & T. Oakland (Eds.), *International perspectives on psychology in the schools* (pp. 11-124). Hillsdale, NJ: LEA.
- Raviv, A., Margalith, M., Raviv, A., & Sade, F. (1981). The cognitive patterns of Israeli learning disabled children as reflected in the Hebrew version of the WISC-R. *Journal of Learning Disabilities*, 14, 411-415.
- Rieck, M. (1988). Evaluation of children with learning disabilities on Luria's Neuropsychological Test Battery. *School Psychology International*, 9, 285-289.
- Rorschach, H. (1942). *Psychodiagnostics: A diagnostic test based on perception*. Berne: Huber (1st German Edition).
- Saad, S. (1984). The Arab educational system: Trends and problems. In A. Ramsberg & A. Moskovitz (Eds.), *The Arab citizens of Israel*. Jerusalem: Van Leer (in Hebrew).
- Safir, M.P. (1986). The effects of nature or of nurture on sex differences in intellectual functioning. *Sex Roles*, 14, 581-590.
- Smilansky, M., & Yam, Y. (1969). The relationship between family size, ethnic origin, fathers' education, and students' achievement. *Megamot*, 16, 248-273 (in Hebrew).
- Smoocha, S., & Hofman, J.E. (1976/77). Some problems of Arab-Jewish coexistence in Israel. *Middle East Review*, 9, 5-15.
- Stahl, A. (1977). *Language and thought of culturally deprived children in Israel*. Otzar Ha-Morch: Tel-Aviv (in Hebrew).
- Tamir, P. (1986). Achievement of Jewish and Arab students who studied inquiry oriented curriculum for several years. *Studies in Educational Evaluation*, 12, 191-195.
- Wainer, H., & Braun, H.L. (1988). *Test validity*. Hillsdale, NJ: Lawrence Erlbaum.
- Zeidner, M. (1985). A cross-cultural test of the situational bias hypothesis – the Israeli scene. *Evaluation and Program Planning*, 8, 367-376.
- Zeidner, M. (1986a). Sex differences in scholastic aptitude: The Israeli scene. *Personality and Individual Differences*, 7, 847-852.
- Zeidner, M. (1986b). Sex differences in scholastic ability in Jewish and Arab college students in Israel. *Journal of Social Psychology*, 7, 847-852.
- Zeidner, M. (1987). A test of the cultural bias hypothesis: Some Israeli findings. *Journal of Applied Psychology*, 72, 38-48.
- Zeidner, M. (1988a). Sociocultural differences in examinees' attitudes towards scholastic ability exams. *Journal of Educational Measurement*, 25, 67-76.
- Zeidner, M. (1988b). Cultural fairness in aptitude testing revisited: A cross-cultural parallel. *Professional Psychology: Research and Practice*, 19, 257-262.

Moshe Zeidner is a faculty member at the School of Education, University of Haifa. Research areas in which he has published include personality and individual differences, with particular concern for anxiety; cultural, sex, and age bias in psychological tests; sex, sociocultural, and age differences in ability; test attitudes and motivations. He serves as co-editor and methodological consultant for the new international journal *Special Education and Rehabilitation: A Multidisciplinary Journal* and methodological consultant and member of the review board for the American publication *Journal of Classroom Interaction*. He is currently a member of the Executive Committee of the International Society for Test Anxiety Research (STAR), chairs the Research Committee of the International School Psychology Association, and is a member of the American Psychology Association.

Moshe Zeidner enseigne à la faculté des sciences de l'éducation de l'université de Haïfa. Ses recherches portent sur les différences de caractère et les différences individuelles, notamment au niveau de l'anxiété, sur la distorsion des tests psychologiques au niveau de la culture, du sexe et de l'âge, sur les différences de compétences liées au sexe, aux éléments socioculturels et à l'âge et enfin sur les attitudes et les motivations des tests. Il a été coéditeur et consultant de la nouvelle revue internationale *Special Education and Rehabilitation: A Multidisciplinary Journal* et consultant et membre du comité de lecture de la revue américaine *Journal of Classroom Interaction*. Il est actuellement membre du comité exécutif de l'International Society for Test Anxiety Research (STAR), il préside le comité de recherche de l'Association internationale des psychologues scolaires et il est membre de l'American Psychology Association.



Philip A. Saigh
Graduate Center, City University of New York

School-Based Assessment Research in Lebanon

Abstract

Initially, descriptions of the recent history of Lebanon and the Lebanese systems of education are presented. Attention is subsequently focused on school-based assessment research involving cognitive, achievement, and personality testing. Attention is also afforded to research involving contextual factors that relate to the assessment process. Recommendations for additional research are proposed and the feasibility of these recommendations is discussed vis-a-vis the prevailing situation.

Résumé

L'auteur commence par donner une description de l'histoire récente du Liban et des systèmes d'éducation libanais. Il s'attache ensuite aux recherches sur l'évaluation en milieu scolaire, notamment sur les tests cognitifs, les tests de niveau et de personnalité. Il s'intéresse également aux recherches sur les facteurs contextuels qui ont trait au processus d'évaluation. Il recommande des recherches plus poussées et analyse la faisabilité de ces recommandations par rapport à la situation qui prédomine.

Recent History

Lebanon, the home of the ancient Phoenicians, became part of the Ottoman empire in the sixteenth century (Hitti, 1970). During the First World War, British and French forces invaded the area and expelled the Ottomans (Salibi, 1977). The country was subsequently administered by France under a League of Nations mandate from 1920 through 1941. A republic was

formally declared in 1941. At that time, the Lebanese Christian and Moslem communities agreed on a delicate balance of political power. The unwritten covenant provided for a distribution of legislative authority in a ratio of 6 Christians to 5 Moslems (Salibi, 1977). It was also formally concluded that the President of the Republic must be a Maronite Christian, the Prime Minister a Sunni Moslem, and the Speaker of the House a Shi'a Moslem (Europa Year Book, 1989).

This agreement was successful for a number of years and Lebanon prospered in an atmosphere of pluralistic harmony. However, with the establishment of Israel in 1948, thousands of Palestinian refugees (mostly Sunni Moslems) received asylum in Lebanon. These refugees coupled with the indigenous Moslem population's greater birth rate served to disturb the delicate balance between the country's religious communities. Under increased pressure from its Moslem community and the Palestinian Liberation Organization (PLO), the Lebanese government allowed the PLO to maintain its headquarters in Beirut and to initiate attacks against Israel through south Lebanon. This situation led to friction between Lebanon's Christians and the PLO. In 1974, the Christian Phalange party and the PLO openly clashed. The fighting rapidly escalated with Lebanon's Moslems openly siding with the PLO. Full scale warfare quickly enveloped the country and the Lebanese government collapsed in 1975. The Arab League subsequently convened and dispatched a 30,000-Arab Deterrent Force (ADF) to quell the hostilities. Although the ADF was able to temporarily contain the hostilities, it did not disarm the PLO or the various Lebanese militias. Since that time the country has, in effect, been partitioned along religious lines.

Continued hostilities against Israel by the PLO were associated with a large scale Israeli invasion in 1982. Although this action prompted the temporary withdrawal of Syrian and Palestinian troops from West Beirut and South Lebanon, these effects were ephemeral as Syrian and Palestinian units returned to their former positions in 1986. Since that time, Syria has assumed an increasingly confrontative role against Lebanon's Christians. In March of 1989, General Aoun, the Christian commander of the Lebanese army, closed the illegal ports that were being used to smuggle narcotics from the country. This action prompted Syria and its Moslem allies to shell the Christian enclave. Christian forces responded in kind, and six months of indiscriminate bombardment led to an evacuation of most of the capital's inhabitants. As this article goes to press, a cease-fire prevails and Lebanese legislatures have revised their constitution. (The new covenant calls for an equal distribution of legislative authority between Lebanese Christians and Moslems.) The agreement does not, however, call for the withdrawal of Syrian forces and

General Aoun has refused to honour any agreement that does not provide for a Syrian withdrawal.

System of Education

The official Lebanese curriculum was formulated according to the French model. Viewed along these lines, the first nine years of schooling offer a common curriculum for all students. By the age of 16 Lebanese students must decide if they want to pursue a literary or scientific major. As students cannot transfer high school marks and as Lebanese universities do not permit students to major in disciplines that are incongruent with their academic records, this decision is quite portentous. Following high school graduation, students may apply to a number of Lebanese universities (e.g., The American University of Beirut [AUB], Beirut University College, or the Université Saint Joseph). It should be noted in this context that two major universities train educators and psychologists in Lebanon: The American University of Beirut and the Université Saint Joseph. The former is a private institution that is largely subsidized by the American government, whereas the latter is a private institution that is subsidized by France and administered by Jesuits. Although *force majeure* precluded the collection of up-to-date educational statistics, it was estimated that 178,164 students enrolled during the 1984-1985 academic year (Europa Year Book, 1989). This figure is indicative of approximately 26.14% of the Lebanese population.

School-Based Assessment Research

Although the ongoing hostilities have drastically reduced research productivity, a considerable amount of applied and theoretical research has taken place. Given that school-based assessment research is a relatively orthogonal process, this paper will not review the research contributions that emanated in the fields of developmental (Diab & Prothro, 1957), social (Melikian & Prothro, 1955), experimental (Keehn, 1959; Walker, 1986), abnormal (Saigh, 1988), or clinical (Day & Sadek, 1982; Saigh, 1987) psychology. Attention will, however, be focussed on the research that has involved cognitive, achievement, and personality testing.

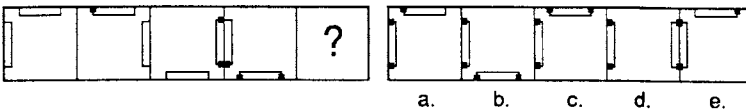
Cognitive tests

Two measures of mental ability were expressly developed for Lebanese students. Initially, Saigh (1984) constructed the *Lebanese Intelligence Scale for Adults* (LISA). The LISA presents a series of items relating to general knowledge, mathematics, analogies, logic, and schematic reasoning.

Figure 1 presents a selected sample of the aforementioned items. LISA test-retest reliability coefficients have ranged from .84 ($p < .01$) to .90 ($p < .01$) and moderate coefficients have been observed when LISA scores were correlated with secondary school semester grades in Arabic ($r = .45$, $p < .01$), civics ($r = .50$, $p < .01$), English ($r = .48$, $p < .01$), and mathematics ($r = .43$, $p < .01$) (Saigh, 1984).

Figure 1
Selected items from the Lebanese Intelligence Scale for Adults

15. The ancient temples of Balbeck were constructed under the rule of the
- a. Phoenicians b. Persians c. Romans
d. Greeks e. Egyptians
20. Samir must come from Womba because he has a red beard. This statement is logical if it is true that
- a. All adult Wombas have red beards
b. Few men aside from Wombas have red beards
c. All Wombas have red beards
d. Only Wombas have red beards
e. Red beards are traditional in Womba
34. The designs in the first box are arranged in a special way. Identify the design from the second row that belongs where you see the question mark.



39. Antibiotic is to disease as
- a. nurse is to doctor b. microbe is to infection
c. dark is to light d. vaccine is to typhoid
e. water is to drought
41. In a group of 30 people, 12 individuals traveled on the steamship Esperia and 10 people traveled aboard the steamship El Greco. Eight people at one time or another traveled on both ships. How many people out of the original 30 have not traveled on either ship?
- a. 30 b. 24 c. 22 d. 16 e. 20

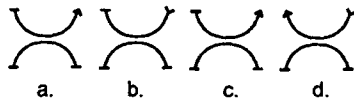
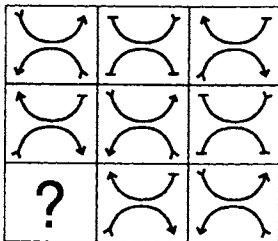
The *Lebanese General Ability Scale* (LGAS) (Saigh, 1986a) was developed in order to facilitate the selection of university scholarship applicants. The instrument was constructed with the view that subjective environmental experiences may prevent otherwise able applicants from succeeding on curriculum-based admission tests. As such, the LGAS provides items that involve analogies, logic, numerical concepts, and schematic reasoning. Figure 2 provides examples of the aforementioned items.

Figure 2
Selected items from the Lebanese General Ability Scale

1. The numbers are arranged according to a certain rule. Identify the missing number.

11	3	9
4	6	4
44	18	?

- a. 44 b. 1 c. 36 d. 38
2. Algebra is to mathematics as
- a. logic is to philosophy c. geometry is to triangle
 b. economics is to money d. experiment is to psychology
3. If Mahmoud eats twice as much as Yousef, who eats one-third as much as Ibrahim, who eats twice as much as Adel, who eats the most?
- a. Adel b. Mahmoud c. Yousef d. Ibrahim
4. The designs are arranged according to a certain pattern. Identify the missing figure.



Three alternate forms of the LGAS are available in Arabic. Each form is associated with a good deal of internal consistency (i.e., Kuder-Richardson formula 21 coefficients have ranged from .84 [$p < .01$] to .90 [$p < .01$]). Moderate coefficients were also observed when LGAS scores were correlated with undergraduate semester grades in organic chemistry ($r = .45$, $p < .01$), English ($r = .32$, $p < .05$), algebra ($r = .55$, $p < .01$), and architecture ($r = .69$, $p < .01$).

Examined from a cross-validated perspective, Subheyha (1983) administered Form R *Otis Lennon School Ability Test* (OLSAT) (Otis & Lennon, 1979) to 185 secondary school students whose ages ranged from 15 to 18 years. A test-retest procedure over a four-week interval evidenced a high degree of reliability ($r = .86$, $p < .01$). Subheyha also tested the alternate form reliability of the OLSAT by administering Forms R and S to the subject pool. He subsequently observed a coefficient of .76 ($p < .01$). OLSAT Form R scores were also correlated with semester grades in English ($r = .55$, $p < .01$), algebra ($r = .40$, $p < .01$), geometry ($r = .35$, $p < .01$), physics ($r = .55$, $p < .01$), chemistry ($r = .46$, $p < .01$), and social studies ($r = .23$, $p < .05$).

Achievement tests

The *English Entrance Exam* (EEE) (Miller, 1970) is the most frequently used achievement test in Lebanon. The EEE consists of 200 multiple-choice items involving English grammar, vocabulary, spelling, and comprehension. Test items are furnished by the faculty of the AUB's English Department and test development specialists from that university's Office of Tests and Measurements. The EEE appears in several alternate forms that are revised and up-dated on a yearly basis. The internal consistency of the EEE was established through the calculation of a series of Kuder-Richardson formula 21 coefficients. Viewed in this context, coefficients have ranged from .94 ($p < .01$) to .90 ($p < .01$) (Miller, 1970). Miller (1978) also established the test-retest reliability of the EEE by administering the instrument to 745 undergraduates and observing a coefficient of .79 ($p < .01$). The EEE presents moderate predictive validity given that coefficients of .54 ($p < .01$), .61 ($p < .01$), and .57 ($p < .01$) were observed when scores correlated with the first year GPA's of AUB undergraduates who were enrolled in the Faculties of Arts and Sciences, Engineering, and Agriculture. In light of these qualities, the AUB and Beirut University College rely on EEE scores as an integral part of their admission criteria.

The *Scientific Quarterly* (SQ) (Miller, 1980) has also enjoyed considerable utility in Lebanon. As in the case of the EEE, test items are prepared

by the faculty of the AUB's Departments of Biology, Chemistry, Mathematics, and Physics as well as specialists from the AUB's Office of Tests and Measurements. Miller (1980) observed Kuder-Richardson formula 21 coefficients of .91 ($p < .01$) and .94 ($p < .01$) with alternate forms of the test. SQ scores have also been moderately associated with undergraduate marks in biology ($r = .42$, $p < .01$), chemistry ($r = .45$, $p < .01$), mathematics ($r = .48$, $p < .01$), and physics ($r = .42$, $p < .01$).

Although a good deal of effort has been afforded to the development of endemic measures of achievement, less work has been ascribed to cross-validated studies. Moreover, the modicum of research in this area has produced rather unsatisfactory validity estimates. Viewed along these lines, Saigh and Khairallah (1984) attempted to cross-validate the *Diagnostic Analysis of Reading Errors* (DARE) (Gillespie & Shohet, 1979). According to the test manual, the DARE was developed in order to identify adolescents and adults with language-related learning disabilities and to provide diagnostic information about the nature of their disability. In so doing, the DARE provides an overall composite of the degree of impairment as well as subtests to identify specific error patterns that involve sound substitution, omission, and letter-sequence reversals. With these points in mind, Saigh and Khairallah administered the DARE to 40 post-secondary school students who were enrolled in an extensive English preparatory course and 79 secondary school students. Flanagan's (1937) formula was used to calculate the reliability of the test and coefficients of .60 ($p < .01$) and .65 ($p < .01$) were observed on the basis of composite scores of the two groups. The EEE scores of the post-secondary school students and the semester English grades of the secondary school students were correlated and coefficients of .31 ($p < .05$) and .32 ($p < .05$) were noted. Inverse associations were observed with the DARE substitutions ($r = -.27$, $p < .04$), omissions ($r = -.8$, $p > .05$), and reversals ($r = -.18$, $p > .05$) subtest when correlated with EEE scores. In a similar vein, these subtests were inversely associated ($r = -.18$, $p < .05$; $r = -.26$, $p < .05$; and, $r = -.21$, $p < .05$) with the English grades of the secondary school students. Saigh and Khairallah went on to suggest that these coefficients were indicative of the limited correspondence between the measures that were correlated. More specifically, it was indicated that the EEE and English semester grades consisted of test items involving English grammar, vocabulary, reading comprehension, and spelling whereas the DARE is made up of a 46-item spelling list.

Personality tests

Due to the importance that Lebanese parents and educators ascribe to scholastic achievement, high levels of test anxiety are frequently reported (cf. Saigh & Antoun, 1984). In light of the adverse effects of test anxiety and the

need to derive valid diagnostic and treatment outcome measures, two investigations were directed toward the cross-validation of self-report test anxiety inventories. Initially, Saigh and Mukallid (1983) administered the *Suinn Test Anxiety Behavior Scales* (STABS) (Suinn, 1979) to 192 English-speaking secondary school students. A test-retest reliability procedure yielded a coefficient of .89 ($p < .01$). Pearson product-moment correlations were calculated on the basis of the students' STABS scores and their semester grades in Arabic ($r = -.09$, $p > .05$), chemistry ($r = -.15$, $p > .05$), English ($r = -.06$, $p > .05$), geography ($r = -.06$, $p > .05$), mathematics ($r = -.09$, $p > .05$), and physics ($r = -.09$, $p > .05$). Given the apparent limited predictive power, and as more robust coefficients were reported in the West (cf. Suinn, 1979), Saigh and Mukallid conducted a series of *post hoc* interviews. It was subsequently determined that more than half of the selected subjects deliberately falsified their scores because they felt that their actual level of anxiety would augur against their university admission. In view of this, Saigh (1986b) administered the STABS to a similar sample (with the assurance that their scores would be kept in the strictest confidence) and moderate inverse correlations were observed between the STABS estimates and semester grades in chemistry ($r = -.41$, $p < .01$), English ($r = -.44$, $p < .01$), and mathematics ($r = -.40$, $p < .01$).

Saigh and Khuri (1983) subsequently translated the *Mathematics Anxiety Rating Scale for Adolescents* (MARS-A) (Suinn & Richardson, 1972) to Arabic. The cross-language equivalence of the translation was addressed by administering the Arabic and English MARS-A to two bilingual twelfth-grade classes according to a counterbalanced design. Correlational coefficients of .85 ($p < .01$) and .91 ($p < .01$) were noted. In an effort to avoid the problems that were encountered with the STABS, the school principal met with the subjects (133 Lebanese adolescents) and urged them to respond in a candid way with the understanding that their scores would remain in confidence. The Arabic MARS-A was subsequently administered and moderate inverse coefficients were observed between the students' mathematic anxiety scores and their semester grades in chemistry ($r = -.31$, $p < .01$), mathematics ($r = -.56$, $p < .01$), and physics ($r = -.30$, $p < .01$).

Given that all sectors of society have been affected by the war, a number of investigations have cross-validated or developed stress-related measures. Mathia (1982), for example, translated Wolpe and Lang's (1964) *Fear Survey Schedule* (FSS) to Arabic and administered the translation to 300 students. Cronbach *alpha* coefficients were reported in the range of .91 ($p < .01$) to .94 ($p < .01$) and FSS scores were moderately associated ($r = -.43$, $p < .01$) with clinically derived anxiety estimates.

In a similar vein, Saigh (1982a) translated the *State-Trait Anxiety Inventory* (STAI) (Spielberger, Gorsuch, & Luschane, 1968) to Arabic and administered the Arabic and English versions to 50 bilingual undergraduates according to a counter-balanced design. As coefficients of .90 ($p < .01$) and .85 ($p < .01$) spoke well for the cross-language equivalence to the Arabic STAI, Saigh went on to administer the Arabic translation to 128 undergraduates on two occasions over a five-day interval. Test-retest coefficients of .74 ($p < .01$) and .81 ($p < .01$) were observed for the state and trait scales. The state and trait scales were also moderately associated ($r = .51$, $p < .01$ and $r = .62$, $p < .01$) with FSS scores. Although the Lebanese psychometric properties of the FSS and STAI were satisfactory, it became apparent that these measures could not directly assess situationally specific fears that were indicative of the Lebanese crisis. In view of this, Saigh (1982b) developed the *Lebanese Fear Inventory* (LFI). In so doing, the incidence and type of war-related fear reactions that were observed over a two year interval at a counselling clinic were combined and a 15-item Likert-type inventory was constructed on a proportional basis to the observed frequencies. Table 1 (page 74) presents a representative sample of the LFI items.

The LFI was administered to 206 outpatients with a wide range of somatic disorders; a Cronbach *alpha* coefficient of .89 ($p < .01$) was noted. LFI scores were also moderately associated ($r = .62$, $p < .05$) with psychiatric ratings of war-related fears.

In an effort to establish a valid treatment-outcome assessment package for Lebanese children, Saigh (1986c) translated the *Revised Children's Manifest Anxiety Scale* (RCMAS) (Reynolds & Richmond, 1978), *Children's Depression Inventory* (CDI) (Kovacs, 1978), and the *Conners Teacher Rating Scales* (CTRS) (Conners, 1969) to Arabic. The English and Arabic versions of the RCMAS and CDI were administered to a sample of bilingual children; coefficients of .82 ($p < .01$) and .80 ($p < .01$) were noted. A bilingual sample of teachers also marked the Arabic and English version of the CTRS versions; a coefficient of .89 ($p < .01$) was observed. Saigh went on to administer the RCMAS and CDI to 691 children. The teachers of the selected sample also rated the students' accuracy according to the CTRS criteria. The large scale test administration yielded Cronbach *alpha* coefficients of .89, .92, and .92 for the RCMAS, CDI, and CTRS, respectively. The construct validity of the battery was established (Saigh, 1989a) by comparing the scores of three groups of subjects whose scores should have differed as a function of their psychiatric morbidity. The first group consisted of 231 children with posttraumatic stress disorder (PTSD), the second group consisted of 32 test phobia cases, and the third group was made up of 35

Table 1

Selected items from the Lebanese Fear Inventory

Directions: The items in this inventory reflect themes or experiences that may cause fear or unpleasant associations. Estimate the degree to which each of the following items disturbs you by placing a check mark (✓) in the appropriate column:

Item	Level of Disturbance				
	Not at all	A little	A fair amount	Much	Very much
3. Being asked to produce your identification card by an armed man	()	()	()	()	()
4. Explosions	()	()	()	()	()
6. Crossing from the "Christian" to the "Moslem" side of Beirut or <i>vice versa</i>	()	()	()	()	()
8. Being kidnapped	()	()	()	()	()
12. Snipers	()	()	()	()	()
14. Discussing politics with a stranger	()	()	()	()	()

nonclinical controls. As expected, the PTSD cases evinced significantly higher RCMAS, CDI, and CTRS scores than their phobic and nonphobic peers.

In a similar vein, Spielberger's (1973) *State-Trait Anxiety Inventory for Children* was translated to Arabic (Day, Knight, Nakadi, & Spielberger, 1986). The investigators administered the Arabic inventory to 266 children; test-retest coefficients of .60 ($p < .01$) and .67 ($p < .01$) were observed for the state and trait scales. In terms of construct validity, Day *et al.* (1986) determined that orphans presented higher state and trait scores than nonorphans.

Test-related contextual variables

Given that test-related contextual variables may contribute to error variance (Barber, 1976), a number of studies have addressed this concern. Saigh (1979), for example, argued that religious affiliations may have a direct bearing on test performance. In order to test this hypothesis, two intellectually matched groups of Lebanese Christian and Moslem school children were randomly assigned to one of three treatment groups. The first group was tested by a male examiner who wore a 3.5 cm by 4.5 cm gold crucifix about his neck, and the second group was tested by the same examiner as he wore a 3.5 cm by 4.5 cm gold symbol of the Koran. The third group was tested in the absence of religious symbols. Data analysis revealed that the *Wechsler Intelligence Scale for Children - Revised* (WISC-R) Digit Span scores varied as a function of the correspondence between the examinee's religion and the symbols that were worn. More specifically, the scores of Christians and Moslems significantly exceeded the scores of their peers when they were tested by an examiner who wore a symbol that was commensurate with their affiliation. Moreover, a reverse effect was observed when the subjects were examined when a symbol of a different religion was worn. It was subsequently indicated that the region's history of religious strife, in addition to the prevailing crisis, should be viewed as major determinants of the performance that was observed. More specifically, Saigh suggested that the religious symbols may have served to augment or reduce anxiety and that these variations were evinced through divergent scores.

In a similar vein, Saigh (1981) tested the assertion that the Lebanese assign precise and meaningful connotations to gestural language (Brewer, 1951). In so doing, 40 undergraduates were randomly assigned to one of two treatment conditions. Both groups went on to receive an administration of the Similarities and Digit Span subtests of the *Wechsler Adult Intelligence Scale* (WAIS). The students in the experimental condition were tested by an examiner who leaned forward in an attentive posture, established eye contact as the questions were presented, and simultaneously smiled and nodded after each response. On the other hand, the control group received neutral non-evaluative examiner feedback. In this case the examiner consistently leaned back and looked down at the WAIS manual as he questioned each subject. The examiner also maintained a bland facial expression throughout the assessment. Data analysis determined that the WAIS Similarities and Digit Span scaled scores of the experimental subjects significantly exceeded the scores on the controls. Saigh went on to suggest that the regularity of the experimental procedure and the close association between the examinee's responses and the examiner's nonverbal behaviour may have reinforced the

subject's efforts and this may have been reflected through enhanced performance. By the same token, it was also suggested that the noncommittal behaviour of the examiner in the face of increasingly more difficult questions may have induced anxiety and affected the performance of the controls.

Examined from a more operant perspective, Saigh (1985) observed that classroom testing is generally indicative of fixed-interval (i.e., announced and prearranged) or variable-interval (i.e., unannounced and unscheduled) schedules of reinforcement. In an effort to assess the effects of these schedules, two matched groups of undergraduates were selected from two learning and human development classes. The experimental subjects were told that they were going to receive a series of unscheduled assessments over a five-week interval and that their grades on these quizzes would reflect 20% of their semester grade. On the other hand, the controls were told that they were scheduled to receive a series of quizzes that would constitute 20% of their semester grade and that these quizzes were going to be administered on five consecutive Fridays. The fixed and variable regimens were subsequently effected. Pursuant to this, the subjects marked the STABS and a course evaluation questionnaire. Data analysis revealed that the experimental subjects were significantly more test anxious and dissatisfied with the course and the instructor. On the other hand, the quiz averages of the experimental and control groups were not significantly different. In view of these points, the social validity (Wolf, 1978) of variable-interval assessment regimens was challenged.

Recommendations

Examined *in toto*, it is apparent that a number of school-based intelligence, achievement, and personality tests have been developed. Although the majority of these measures evince adequate psychometric properties, additional research is in order. It should be observed in this context that the group intelligence tests that have been developed or cross-validated yield singular quotients and that these quotients are not well-suited for establishing specific school-related diagnoses (e.g., developmental reading disorder). In view of this and as measures like the LISA and LGAS present heterogeneous item pools, researchers should factor analyze these instruments and establish the validity of the ensuing factors *vis-a-vis* school grades. It is also recommended that research should be directed toward the development of multifaceted, individually-administered intelligence scales.

Viewed from the standpoint of achievement testing, it is important to note that well-validated tests like the EEE or SQ are not appropriate for

assessing elementary or junior high school students. As such, research should be directed toward the development of endemic norm-referenced measures that are indicative of the Lebanese curriculum. Research should also be directed toward the cross-validation of extant measures (e.g., the *Stanford Diagnostic Mathematics Test*, [Beatty, Madden, Gardner, & Karlsen, 1986]) inasmuch as a number of these measures appear to offer a good deal of content validity.

Given that all sectors of Lebanon's society have been affected by the war, it is not surprising to observe that a number of stress-related tests have been developed or cross-validated. Although the majority of these tests have yielded adequate reliability and validity coefficients, these tests are not suitable for formulating discrete psychiatric diagnoses (e.g., posttraumatic stress disorder). In view of this and the prevalence of childhood and adolescent psychiatric cases with war-related etiologies, it is incumbent on researchers to cross-validate structured clinical interviews (e.g., the Structured Clinical Interview for the DSM-III-R, [Spitzer, Williams, & Gibbon, 1987]) that are intended to provide psychiatric diagnoses.

Future Directions

As one might expect, the war in Lebanon has had a negative impact on all sectors of society. In terms of human suffering, Halande (1986) conservatively estimated that 100,000 people have been killed and that 860,000 were compelled to emigrate. Given an estimated pre-war population of 3,000,000, these estimates are especially daunting. Lebanon, once the "Switzerland of the Middle East," has been pummeled by the flight of capital, loss of tourism, and currency devaluations. Moreover, foreign and well-qualified Lebanese university faculty have left the country pursuant to the abduction or assassination of a number of their colleagues. The combination of these factors has drastically affected the quantity and quality of scientific inquiry.

The prospects for Lebanese research are, of course, inexorably linked with the future of the country (cf. Saigh, 1989b). Should the current state of unrest prevail, it is anticipated that research productivity will be virtually moribund. On the other hand, it is anticipated that a political settlement should lead to an economic recovery inasmuch as Lebanon has virtually no foreign debt and one of the largest gold reserves in the world. Economic and political stability would, over time, lead to the restaffing of university faculties, and the combination of these developments would serve to facilitate research productivity. Unfortunately, the Byzantine nature of Lebanese poli-

tics makes it difficult to predict which scenario will prevail. It is apparent, however, that the more positive prospectus will not prevail until the well-being of the country is placed above international and confessional self-interests.

REFERENCES

- Barber, T.X. (1976). *Pitfalls in human research: Ten pivotal points*. New York: Pergamon.
- Beatty, L.S., Madden, R., Gardner, E.F., & Karlsen, B. (1986). *Stanford Diagnostic Reading Test*. San Antonio, TX: Psychological Corporation.
- Brewer, W.D. (1951). Patterns of gestures among the Levantine Arabs. *Anthropologist*, 42, 232-237.
- Connors, C.K. (1969). A teacher rating scale for drug studies with children. *American Journal of Psychiatry*, 126, 884-888.
- Day, R.C., Knight, C.W., Nakadi, L., & Spielberger, C.D. (1986). Development of an Arabic adaptation of the State-Trait Anxiety Inventory for Children. In C.D. Spielberger, & R.D. Guerrero (Eds.), *Cross-cultural anxiety*, Vol. 3 (pp. 113-123). New York: Hemisphere.
- Day, R.C., & Sadek, S. (1982). The effect of Benson's relaxation response on the anxiety levels of Lebanese children under stress. *Journal of Experimental Child Psychology*, 34, 350-356.
- Diab, L.N., & Prothro, E.T. (1957). Cross-cultural study of some correlates of birth order. *Psychological Reports*, 22, 1137-1142.
- Europa Year Book. (1989). *Lebanon*. London: Europa Publishing Limited.
- Flanagan, J.C. (1937). A proposed procedure for increasing the efficiency of tests. *Journal of Educational Psychology*, 28, 17-21.
- Gillispie, J., & Shohet, J. (1979). *Diagnostic Analysis of Reading Errors Manual*. Wilmington, Delaware: Jastak.
- Halande, J.T. (1986, November). Lebanon: Facing economic ruin. *The Washington Report on the Middle East*, p. 19.
- Hitti, P. (1970). *History of the Arabs*. London: MacMillan.
- Keehn, J.D. (1959). On the non-classical nature of avoidance behavior. *American Journal of Psychology*, 72, 243-247.
- Kovacs, M. (1978). *The Children's Depression Inventory: A self-rated scale for school children*. Unpublished manuscript, University of Pittsburgh.
- Mathia, M. (1982). *Lebanese norms for the Fear Survey Schedule*. Unpublished master's thesis, American University of Beirut, Beirut.
- Melikian, L.H., & Prothro, E.T. (1955). Goals chosen by Arab students in response to hypothetical situations. *Journal of Social Psychology*, 46, 3-9.
- Miller, G.F. (1970). *A concurrent validity study of the AUB EEE and the English Usage Test for Nonnative Speakers of English*. Beirut: American University of Beirut, Office of Tests and Measurements.

- Miller, G.F. (1978). *Test-retest reliability of the AUB EEE*. Beirut: American University of Beirut, Office of Tests and Measurements.
- Miller, G.F. (1980). *The concurrent validity of the AUB SO*. Beirut: American University of Beirut, Office of Tests and Measurements.
- Otis, A.S., & Lennon, R.T. (1979). *Otis-Lennon School Ability Test*. San Antonio, TX: Psychological Corporation.
- Reynolds, C., & Richmond, B. (1978). What I think and feel: A revised measure of children's manifest anxiety. *Journal of Abnormal Child Psychology*, 6, 271-280.
- Saigh, P.A. (1979). The effects of perceived examiner religion on performance of Lebanese elementary school children. *Journal of Social Psychology*, 109, 167-173.
- Saigh, P.A. (1981). The influence on nonverbal examiner praise on the WAIS performance of Lebanese undergraduates. *Journal of Non-Verbal Behavior*, 6, 84-88.
- Saigh, P.A. (1982a) *The reliability and validity of the Arabic version of the State-Trait Anxiety Inventory*. Unpublished manuscript.
- Saigh, P.A. (1982b). The Lebanese Fear Inventory: A normative report. *Journal of Clinical Psychology*, 38, 352-355.
- Saigh, P.A. (1984). *The Lebanese Intelligence Scale for Adults*. Beirut: Privately Printed.
- Saigh, P.A. (1985). Unscheduled assessment: Academic achievement, test anxiety, and social validity. *Educational Research Quarterly*, 9, 6-11.
- Saigh, P.A. (1986a). *The Lebanese General Ability Scale (Forms A, B, and C)*. Beirut: American University of Beirut, Office of Research and Development.
- Saigh, P.A. (1986b). *The validity of the Suinn Test Anxiety Behavior Scale as a predictor of the academic achievement of Lebanese secondary school students*. Unpublished manuscript.
- Saigh, P.A. (1986c). Childhood psychology in Lebanon. In H. Armenian (Ed.), *The status of children in wartime* (pp. 227-235). Beirut: American University Press.
- Saigh, P.A. (1987). In vitro flooding of a childhood posttraumatic stress disorder. *Journal of Clinical Child Psychology*, 16, 147-150.
- Saigh, P.A. (1988). Anxiety, depression and assertion across alternating intervals of stress. *Journal of Abnormal Psychology*, 97, 338-341.
- Saigh, P.A. (1989a). The validity of the DSM-III posttraumatic stress disorder classification as applied to children. *Journal of Abnormal Psychology*, 98, 189-192.
- Saigh, P.A. (1989b). School psychology research in Lebanon: A retrospective analysis and a look ahead. *Professional School Psychology*, 4, 201-208.
- Saigh, P.A., & Antoun, F. (1984). Endemic images and the desensitization process. *Journal of School Psychology*, 22, 177-183.
- Saigh, P.A., & Khairallah, S. (1984). The concurrent validity of the Diagnostic Analysis of Reading Errors as a predictor of the English achievements of Lebanese students. *Educational and Psychological Measurement*, 43, 1149-1152.

- Saigh, P.A., & Khuri, A. (1983). The concurrent validity of the Mathematics Anxiety Rating Scale for Adolescents (MARS-A) in relation to the academic achievement of Lebanese students. *Educational and Psychological Measurement, 43*, 633-636.
- Saigh, P.A., & Mukallid, S. (1983). The validity of the STABS as a predictor of the academic achievement of Lebanese students. *Educational and Psychological Measurement, 43*, 633-663.
- Salibi, K. (1977). *A modern history of Lebanon*. Delman, NY: Caravan.
- Spielberger, C.D., Gorsuch, R.L., & Lushane, R.E. (1968). *Manual for the State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.
- Spielberger, C.D. (1973). *State-Trait Anxiety Inventory for Children*. Palo Alto, CA: Consulting Psychologists Press.
- Spitzer, R.L., Williams, J.B., & Gibbon, M. (1987). *Structured clinical interview for the DSM-III-R (SCID)*. New York: Psychometric Institute, Biometrics Research Department.
- Subheyha, R. (1983). *The reliability and validity of the Otis-Lennon School Ability Test as based on the performance of Lebanese students*. Unpublished master's thesis, American University of Beirut.
- Suinn, R.M. (1979). The STABS, a measure of test anxiety for behaviour therapy. *Behaviour Research and Therapy, 7*, 335-339.
- Suinn, R.M., & Edwards, R. (1982). The measurement of mathematics anxiety: The Mathematics Anxiety Rating Scale for Adolescents. *Journal of Clinical Psychology, 38*, 576-580.
- Walker, N. (1986). Direct retrieval from memory traces. *Memory and Cognition, 14*, 321-328.
- Wolf, M. (1978). Social validity: The case of subjective measurement or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis, 11*, 203-214.
- Wolpe, J., & Lang, P. (1964). A Fear Survey Schedule for use in behavior therapy. *Behaviour Research Therapy, 2*, 27-30.

Philip A. Saigh is a professor and head of the doctoral program in school psychology at the Graduate Centre of the City University of New York. He previously served on the faculty of the American University of Beirut from 1977 to 1986. Professor Saigh is a Fellow of the American Psychological Association (Division 16, School Psychology), and his research interests primarily involve the assessment and treatment of posttraumatic stress disorder. Many of his publications have been in that area as well as in the area of school-based assessment.

Philla A. Saigh est professeur et directeur du programme de doctorat en psychologie scolaire au Graduate Center de la City University of New York. Il a enseigné à l'Université américaine de Beyrouth de 1977 à 1986. Il est membre de l'American Psychological Association (division 16, psychologie scolaire). Ses recherches portent principalement sur l'évaluation et le traitement des troubles post-traumatiques. Bon nombre de ses publications portent sur ces questions et sur le problème de l'évaluation en milieu scolaire.

Jack A. Cummings
Indiana University

Marc C. Laquerre
Indiana University

Psychoeducational Assessment Research in the United States: From sociocultural context to extrapolations into the future

Abstract

Sociocultural factors are presented in order to provide a frame from which to view psychoeducational assessment research in the domains of cognitive ability, academic achievement, adaptive behavior, and social-emotional functioning. The role of theory is examined as it has had an impact on development and revision of recently published intelligence measures. Research on the complex issue of test bias is also briefly addressed. Two research trends in achievement are highlighted: the mismatch of local curriculum and nationally norm-referenced test content and curriculum-based assessment. The significant developments in the areas of adaptive behavior, social skills, and social-emotional assessment are explored. The paper is concluded with the authors' impressions and recommendations gleaned from the current analysis of research articles, test reviews, and technical manuals.

Résumé

Les facteurs socio-culturels servent de cadre à l'analyse des recherches sur l'évaluation psychopédagogique dans les domaines des compétences cognitives, des résultats scolaires, du comportement adaptatif, du fonctionnement socio-affectif, etc. Les auteurs analysent le rôle de la théorie et son impact sur la conception et le remaniement de mesures récemment publiées sur l'intelligence. Ils abordent également la question complexe des tests biaisés. Deux tendances se dégagent plus particulièrement: la disparité des programmes locaux et du contenu des tests nationaux et l'évaluation basée sur les programmes. Les auteurs analysent les progrès importants réalisés dans les domaines du comportement adaptatif, des habiletés sociales et de l'évaluation socio-affective. L'article s'achève sur les impressions et recommandations des auteurs qui sont le fruit de leur analyse d'articles de recherche, de tests et de manuels techniques.

Research and practice in psychoeducational assessment take place within a sociocultural context. Legislation, litigation over discriminatory aspects of testing, and public opinion have been powerful forces which have shaped psychoeducational assessment with school-age children. The intent of the first section of this paper is to provide a partial sociocultural perspective from which to view psychoeducational research. In subsequent sections two domains of psychoeducational assessment will be examined. These sections will focus on the research which addresses the use of intelligence, academic achievement, adaptive behavior, and social-emotional functioning measures. Each section will begin with a general discussion of the domain. Also included will be an analysis of present practices with emphasis placed on individual psychoeducational assessment to identify and remediate students who experience significant adjustment difficulties in the regular classroom. This will be followed by an examination of research trends within each area. A section on the authors' impressions of research across the three assessment domains will conclude the paper.

Sociocultural Context

In 1975 PL 94-142 (Education for All Handicapped Children Act) was passed by the U.S. Congress, with an effective date of September 1, 1978. It guaranteed all handicapped children the right to a free appropriate education and a nondiscriminatory evaluation. This legislation has institutionalized the assessment of large numbers of children (Salvia & Ysseldyke, 1988). If a child is referred by a teacher or parent, he or she has the right to a nondiscriminatory evaluation. PL 94-142 represents a categorical approach to special education. Definitions are given for deaf, mentally retarded, multihandicapped, orthopedically impaired, seriously emotionally disturbed, and learning-disabled categories of children. An evaluation is conducted to determine if a child is eligible for placement in the area of the suspected handicapped. If the child is judged by a multidisciplinary team to be eligible, he or she is placed in the qualifying category of a special education class.

As noted by Will (1986), PL-142 has had a tremendous impact on the provision of services to children who have difficulty coping with the demands of the regular class environment. The legislation led to the institutionalization of individualized instruction, increased the role of the parents in the education of their handicapped child, and made education possible for 500,000 previously unserved seriously handicapped children.

The Regular Education Initiative (REI) is a movement that suggests regular, not special, educators should assume responsibility for the education of students we would currently classify as handicapped (Reynolds, Wang, &

Walberg, 1987; Stainback & Stainback, 1984; Will, 1986). REI is a reaction to frustrations associated with the current dual system, where regular is segregated from special education. Reynolds *et al.* (1987) are critical of the existing approach in which programs are categorical, e.g., learning-disabled, mildly mentally handicapped, and seriously emotionally disturbed. The problem is that children with significant needs often do not fit a special education category. The debate of the appropriateness of this radical restructuring of special education is prominent in the special education and school psychology journals (Braaten, Kauffman, Braaten, Polsgrove, Nelson, 1988; Gerber, 1988; Keogh, 1988). Additionally, discussions on the merits of the approach are taking place at state and national conferences. The impact of the movement has yet to be realized, because the primary discussants are individuals associated with universities, rather than school-based practitioners (Davis, 1988). REI represents more than a restructuring of special education, because in order to succeed significant reforms will have to be made by regular educators. Liberman (1985) aptly points out that special educators will have a difficult time obtaining support for the regular education initiative from regular educators. While the movement has yet to have a pervasive impact on services, it does reflect a growing criticism of special education.

The assessment of intellectual functioning of minority children has been the focus of a storm of criticism. Plaintiffs in the Larry P. (1979) and PASE (1980) cases contended that intelligence tests were discriminatory and the major cause for disproportionate representation of minorities in special education classes. The federal judge in the former case banned the use of intelligence tests with minorities. The opinion of the court in the PASE case directly contradicted the ruling in Larry P. The conflicting rulings by two circuit courts increases the likelihood that the U.S. Supreme Court will hear the issue of intelligence tests and their alleged discriminatory impact.

Assessment of Intellectual Functioning

The sociocultural forces influencing psychoeducational assessment are clearly evident in the domain of intelligence. Recent landmark litigation and legislation have had a noticeable impact on current intelligence assessment and, undoubtedly, will continue to shape its future. However, not only is the practice of psychoeducational assessment affected by sociocultural forces, the construct of intelligence itself is conceptualized within a sociocultural context. Sternberg and Salter (1985) have argued that, ". . . all intelligent behavior occurs in a social context that includes goals, expectations, demands and a history of prior experiences" (p. 15). They continue stating that intelligence is defined as "goal-directed, adaptive behavior."

Within the last century, numerous theories have been proffered attempting to describe and explain intelligence. In fact, at times it seems as if there are as many definitions of intelligence as there are theorists. It was this state of affairs that prompted Sternberg to state, "Intelligence is among the most elusive of concepts. Certainly, there are few other concepts that have been conceptualized in as many different ways" (Sternberg, 1985a, p. 3).

Mayr (1982) suggested a dimension which distinguishes two broad categories of intelligence theories. Mayr proposed distinguishing theorists as either "lumpers" or "splitters." Theorists falling in the lumpers category are those who conceive of intelligence as a "general, unified capacity." Spearman's (1927) *g*-factor theory is an early example of this unitary capacity. Likewise, Wechsler's definition of intelligence proposes that intelligence is "the aggregate or global capacity of the individual to act purposefully, to think rationally, and to deal effectively with his environment" (Wechsler, 1944, p. 3). In contrast others suggest that intelligence is not a unitary construct, but rather, composed of a number of separate abilities that are fairly independent of each other, e.g., Guilford (1967). Most theories of intelligence have emerged from a psychometric approach. The structures of intelligence were the primary foci of these theories. In recent years, advances in cognitive psychology have offered new perspectives on the nature of human intelligence (Curtis & Glaser, 1984; Sternberg, 1986; Weinberg, 1989). Sternberg (1985b) suggests that these two approaches be viewed as complementary, with cognitive theories providing welcome insights regarding the cognitive processes that underlie intelligence and intelligent behaviors. Increases in scientific interest and research in information processing has caused the focus to shift from structure to the problem-solving process.

The *Kaufman Assessment Battery for Children* (K-ABC) (Kaufman & Kaufman, 1983) provides an illusion of an intelligence test where the authors emphasized process rather than content. Using Das, Kirby, and Jarman's (1975; 1970) successive and simultaneous model of information processing, Kaufman and Kaufman (1977) set out to build a measure which had as its base a grounding in theory. This stood in stark contrast to Binet's and Wechsler's approach to test development, both of which were atheoretical.

The success with which the K-ABC authors successfully operationalized a processing model has been questioned. Sternberg (1984) pointed out that a selected set of subtests is proposed to measure simultaneous processing, while another set is intended to assess sequential processing. Process is equated with tasks. If the focus had truly been on process of problem-solving, then the test authors would have assessed whether an individual worked at a given task in a successive or simultaneous fashion (Sternberg, 1984).

The role of theory has continued to play a part in the development/revision of cognitive tests during the latter half of the 1980s. As noted by Woodcock (1989), Horn and Cattell's *Gf - Gc* theory (Cattell, 1963; Horn, 1985, 1986) led to a reconstruction of the *Woodcock-Johnson Psycho-Educational Battery-Revised* (Woodcock & Johnson, 1989). In the fourth edition of the Stanford-Binet, Thorndike, Hagen, and Sattler (1986) used a three-level hierarchical model of cognitive abilities to guide the construction of the scale. See Figure 1 for a graphic representation of the theoretical model upon which the Revised Stanford-Binet was built.

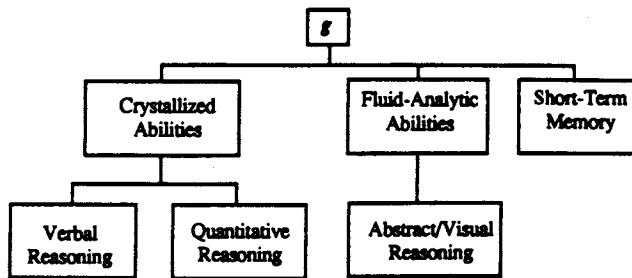


Figure 1
Theoretical Model of the Stanford-Binet: Fourth Edition

Reprinted with permission of The Riverside Publishing Company from page 4 of *Stanford-Binet Intelligence Scale Guide for Administering and Scoring the Fourth Edition*, by R.L. Thorndike, E.P. Hagen, and J.M. Sattler. The Riverside Publishing Company, 8420 W. Bryn Mawr Avenue, Chicago, IL 60631. Copyright 1986.

While recent test revisions have reflected major shifts in thinking about the theoretical construct of intelligence, the actual practice of assessment has shown signs of lagging behind. Federal and state definitions of mental retardation and learning disabilities reflect the unitary concept of intelligence. Likewise, an unfortunate tendency for practitioners has been to judge new measures against the past. Based on the assessment of hundreds or thousands of children, practitioners develop internal norms. When a new measure becomes available it is unfairly judged against the tried and true.

While acknowledging the importance of appropriate and accurate instrumentation, one must not diminish the importance of the psychologist in intelligence testing. A critical link in the chain of intelligence assessment is the examiner. Test scores remain to be interpreted. In elaborating the concept of "intelligent testing," Kaufman (1979) emphasizes the crucial role of the psychologist. He presents a well-founded and logical argument that test scores be interpreted within a more general context that includes observa-

tions, and cognitive and developmental theories. Such interpretation requires knowledge, flexibility, and insight on the part of the examiner. Technically adequate intelligence tests and well-prepared, observant psychologists are partners in the assessment enterprise and must not be seen as substitutes for each other.

In the 1970s and early 1980s, issues concerning the discriminatory nature of intelligence assessment measures received much public and professional attention. Measures were subjected to intensive scrutiny for evidence of test bias (Kaplan, 1985). Test bias is not a novel concept. It has received intermittent scholarly attention since the nascence of intelligence testing. In recent years, however, sparked by the controversial decisions of Larry P. (1979) and PASE (1980), the issue of test bias has been subject to renewed scientific inquiry and research.

At the heart of this controversy is the question of whether contemporary intelligence tests are biased against certain ethnic and cultural minorities. That certain minority groups typically score lower on tests of intellectual ability has been demonstrated for some time (see Jensen, 1980, for a review). This fact alone has, at times, been cited as evidence of test bias. It has also prompted emotional outbursts condemning intelligence tests as agents for the perpetuation of racism (Jackson, 1975; Williams, 1974). Unfortunately, these outbursts may have done more to obscure the issues than to clarify. Reynolds and Brown (1984) accurately characterize the question of test bias as an empirical one. As such, it demands evaluation within the rigors of scientific methodology.

Some have suggested that there is no *a priori* explanation for the existence of differences in performance on tests of intellectual ability across race (Alley & Foster, 1978; Hilliard, 1979). Reynolds and Brown (1984) replied that there is, likewise, no *a priori* basis for believing that differences should not exist. They state that, "Group differences on mental tests, or any other kind of test, *per se* give no directly applicable information regarding test bias" (Reynolds & Brown, 1984, p. 24). Jensen (1980) has suggested that the argument that simple group mean differences is evidence of test bias assumes the egalitarian fallacy (i.e., that all human subgroups are identical or equal in every trait) and reflects an unsophisticated understanding of test bias.

Distinctions have been made between test bias and test misuse (Reynolds, 1982; Reynolds & Brown, 1984). Test misuse concerns the use of a test or test results to make decisions that are unfair or biased. Test misuse can affect individuals from any group. Test bias, however, is typically defined in purely psychometric terms. Reynolds and Brown (1984) define test bias as "a

statistical term referring to a constant error of a measure in one specific direction as opposed to random error" (p. 2). Given this definition, test bias is essentially an issue of validity (Cole, 1981; Jensen, 1980; Reynolds, 1982). Numerous researchers have considered the implications and evaluation of test bias in the domains of content, construct, and predictive validity.

Another current in recent research of test bias is based on the Spearman hypothesis. Jensen (1980) has been a leading, contemporary proponent of this hypothesis. Generally, the Spearman hypothesis suggests that black-white differences in performance on intelligence measures are attributable to differences in *g*-factor loading. The magnitude of black-white differences is hypothesized to vary directly with *g*-factor loadings. Some subsequent research supported this hypothesis (Jensen, 1985). Tests with high *g*-loading yield greater black-white difference in performance. In a more recent study, Naglieri and Jensen (1987) compared the black-white differences in performance on the *Wechsler Intelligence Scale for Children – Revised* (WISC-R) and on the K-ABC. Their findings supported the Spearman hypothesis and indicated that the lower black-white difference demonstrated on the K-ABC was due, in part, to lower *g*-loadings. Jensen's work and the Spearman hypothesis, however, are not without critics. Humphreys (1985) explored black-white differences controlling for race and socioeconomic status. The results of his study suggested that the Spearman hypothesis more accurately applied to differences in socioeconomic status than to differences in race.

The controversy of bias has demonstrated that issues of test bias and such corollary concerns as racial discrimination and societal inequality can become conceptually entwined. Hopefully, a beneficial outcome of the intense scrutiny of test bias has been the development of a better and more accurate understanding. Certainly, recent test developers appear more sensitive to the issue (Kaufman & Kaufman, 1983). Cole (1981) concluded that while scientists can examine such technical aspects of psychometric tests as test bias, they cannot answer the more vexing questions of social policy.

Assessment of Academic Achievement

The focus of achievement testing is to assess a student's attainment of academic skills. These often include reading, written language, and mathematical functioning. Anastasi (1982) notes that, traditionally, academic assessment is distinguished from cognitive/intellectual assessment by the degree to which the measure is designed to assess uniform versus diverse antecedent experiences. Since first grade reading instruction is a fairly uniform previous experience, a test designed to measure reading attainment would be categorized as an academic achievement measure. In contrast, an intellectual assess-

ment measure attempts to measure diverse antecedent experiences, e.g., verbal reasoning, spatial skills, memory, etc. As Cummings (1987) states, "contemporary measurement specialists recognize that both achievement and aptitude tests assess acquired knowledge, but differ on the degree of specificity and abstraction" (p. 21).

A significant development in the assessment of academic achievement is the tendency of test publishers to promote batteries which include both cognitive and achievement measures normed on the same standardization sample. The *Kaufman Assessment Battery for Children* (Kaufman & Kaufman, 1983) and the revised Woodcock-Johnson exemplify this trend. Both of these batteries include cognitive sections to assess intellectual functioning as well as academic achievement.

There has been considerable controversy over the Kaufmans' decision to include what some have labeled measures of verbal intelligence on the achievement scale (Anastasi, 1984; Jensen, 1984; Keith, 1985, 1986; Keith & Dunbar, 1984; Sternberg, 1984). Placing the more cognitively complex subtests (Expressive Vocabulary, Faces & Places, and Riddles) on the achievement scale led to reduced black-white differences, but also removed highly g-loaded measures from the estimate of intelligence provided by the mental processing composite (a combined score from the simultaneous and successive processing scales).

A significant trend in research has been the examination of the mismatch of curriculum and norm-referenced test content when interpreting students' performances on norm-referenced measures (Good & Salvia, 1988; Jenkins & Pany, 1978; Leinhardt & Seewald, 1981; Mehrens, 1984; Mehrens & Phillips, 1986; Schmidt, 1983). At issue is the lack of overlap between the curriculum and the test. When a student's test performance is systematically influenced by a difference between the content of the curriculum and the test, it is referred to as curriculum (or content) bias. A study by Good and Salvia (1988) will provide an example. They administered four reading-achievement tests to 65 students who were exposed to a specific reading curriculum. They did a word analysis to quantify the overlap between the four reading measures and the reading program. As expected the students' reading-test performance could be predicted based on the degree of overlap between the test and the curriculum to which they were exposed.

Curriculum-based assessment (also referred to as curriculum-based measurement) has been proposed as an alternative to the norm-referenced approach embodied in both federal and state special education legislation. Curriculum-based assessment focuses on the curricula to which the student is

exposed. This reduces the mismatch between the student's own instructional experiences and the curriculum reflected in the selection of norm-referenced test items. Deno and his colleagues (Deno, 1985, 1986; Jenkins, Deno, & Mirkin, 1979; Marston, Lowry, Deno-Mirkin, 1981; Shinn & Marston, 1985) have described procedures for assessing student achievement in reading, spelling, and arithmetic. For instance, to assess a student's reading skills, he or she would be asked to read aloud for one minute, with the examiner recording the number of words read correctly and incorrectly. This type of reading sample would be repeated frequently, e.g., daily, so as to yield time-series data that could be used to judge the impact of a given instructional program (Deno, 1986). Another example of curriculum-based assessment, applied to written language, would be to use a story starter and allow a student to write for three minutes. Scoring can be as simple as counting correctly written word or letter sequences, or counting more complex, low frequency words.

Adaptive Behavior and Social-Emotional Assessment

Within the last 15 years significant strides have been taken in the domain of adaptive behavior assessment. Prior to the publication of Mercer and Lewis's (1977) *Adaptive Behavior Inventory for Children*, the best adaptive behavior measures failed to meet even minimal psychometric criteria. "Clinical judgement" was the best available practice for evaluating a child's adaptive behavior functioning. Despite the recognition of the importance of adaptive behavior in the legislation covering the identification of mildly mentally handicapped children, the IQ score often was the sole determining factor, with adaptive behavior being ignored. Since the publication of Mercer's adaptive behavior measure, the revised *Vineland Adaptive Behavior Scales* (Sparrow, Balla, & Cicchetti, 1984) and the *Scales of Independent Behavior* (Bruininks, Woodcock, Weatherman, & Hill, 1984) have been published. Both of these reflect careful attention to the test development process and have adequate nationally representative standardization samples (Cummings & Simon, 1988).

Harrison (1989) is correct in her observation that the bulk of research on adaptive behavior has focused on the psychometric properties of the scales. Despite this psychometric emphasis, Harrison (1989) notes that insights have been gained into the multifaceted nature of adaptive behavior through concurrent validation studies. As expected the communication domain of the adaptive behavior (as opposed to the self-help, socialization, or domestic skills domains) correlates most highly with intelligence (Bruininks & McGrew, 1987; Harrison & Kamphaus, 1984; Keith, Fehrmann, Harrison, & Pottebaum, 1987).

Several researchers have investigated the relationship between parents' and teachers' judgments of a child's adaptive behavior functioning (Heath & Obrzut, 1986; Meador & Richmond, 1980; Wall & Paradise, 1981). These researchers have reported that parents' ratings of their children's adaptive behavior are significantly higher than those scores obtained when teachers served as informants.

Depending on one's perspective, social skills may be subsumed as one domain of adaptive behavior functioning or, as Gresham and Reschly (1988) propose, both adaptive behavior and social skills are components of social competence. Although Gresham and Reschly's model is useful as a framework for a practitioner's conceptualization of an assessment, a unified theoretical approach to adaptive behavior/social skills has yet to be advanced and supported with empirical evidence. Notwithstanding, the technical sophistication of behavioral approaches to the assessment of social skills has improved dramatically due to the work of Elliott, Sheridan, and Gresham (1989), Gresham and Elliott (1984, 1989), and Gresham and Reschly (1988). Gresham and Reschly (1988) present systematic procedures for understanding social skill deficits, performance deficits, and self-control deficits. A strength of their work has been the emphasis on treatment validity. Assessment is judged by the degree to which it provides information relevant to the development and implementation of an intervention. Witt and Elliott (1986) have conducted an impressive array of treatment-acceptability studies, examining attitudes of those who assist with or have primary responsibility for implementation of an intervention.

Gresham and Elliott (1989) recently published the *Social Skills Rating System*. It is designed to capture behaviors that affect the teacher-student relationship, academic performance, and peer acceptance. There are three forms, one each for collecting the impressions of teachers, parents, and students. The system was constructed to ensure the social validity of behaviors selected as targets for intervention.

A measure in the social-emotional domain which has improved significantly is the *Revised Children's Manifest Anxiety Scale* (RCMAS) (Reynolds & Richmond, 1985). It has benefited from the efforts of several researchers. Reynolds and Paget (1983) reported the findings of a national standardization effort that included 4,972 children from age 6 to 19. Saigh (1989) and Mattison, Bagnato, and Brubaker (1988) examined and found merit in the use of the RCMAS with the DSM-III. Whereas Reynolds and Richmond's (1978) first attempt at revising the *Children's Manifest Anxiety Scale*, which they called "What I think and feel," would have been criticized for lacking a manual and adequate norms, the RCMAS was endorsed by Gresham (1989)

in the most recent *Mental Measurements Yearbook*. The RCMAS is illustrative of other scales which have profited from an investment of time and effort associated with national standardization samples, and the collective efforts of independent researchers conducting reliability and validity studies.

The *Martin Temperament Assessment Battery for Children* (Martin, 1988) was designed as a measure of temperament variables, activity level, adaptability to new environments, approach/withdrawal tendencies, emotional intensity, distractibility, and persistence. In addition to using the parent and teacher as sources of information, Martin also devised a scale for the clinician to rate a child's temperament during the evaluation process. The standardization sample for the teacher and parent forms was drawn from three regions of the country (Southeast, Northeast, and Rocky Mountain regions). Thus it is not representative of the country as a whole and the scale should be viewed as being in its early phase of development.

Whereas significant developments have occurred in the adaptive-behavior and rating-scale approaches to social-emotional functioning, our view of the progress made in the personality assessment domain, i.e., projective testing, is less favourable.

The *Personality Inventory for Children* (PIC) has been the object of a significant amount of research. Most of the studies have involved the comparison of a clinically identified sample with "normal" peers or other clinically different samples (cf. Clark, Kehle, & Bullock, 1988; Ehly, Reimers, & Keith, 1986; Kelly, 1988). Despite a growing research base on the PIC, its popularity among clinicians, and the publication of revised test manuals (Lachar, 1982; Wirt, Lachar, Klinedinst, & Seat, 1984), the use of the PIC is inappropriate due to the absence of a recent and nationally representative standardization sample. Reviewers for the *Mental Measurements Yearbooks* have harshly criticized the publishers of the PIC for failing to mount a national norming effort (Knoff, 1989; Reynolds, 1985).

Projective drawing techniques, although widely used, continue to be plagued by basic validity problems. Too often the research that is conducted on children's drawings compares the drawing of a clinically identified group with normal controls. The findings of these studies are often contradictory (Cummings, 1986). (For a review of thematic approaches the reader is referred to Obrzut and Cummings [1983].)

Conclusions and Recommendations for the Future

Trying to capture the essence of recent research in the area of psychoeducational assessment is a difficult proposition. It is akin to the fabled doctoral qualifying question of defining the universe and listing three ex-

amples. It would be an understatement to say that researchers are pursuing many different paths. This paper provides coverage of a handful of these directions. Many others should have been included in the review, but had to be omitted due to the space limitations. Among those researchers whose work has considerable merit but has been omitted are: Naglieri (1985), *Matrix Analogies Test*; Achenbach and McConaughy (1987), *Multiaxial Empirically Based Assessment*; and Ysseldyke and Christenson (1987), *The Instructional Environments Scale*.

From our review of a large number of research articles, test reviews, and test manuals we offer the following impressions and recommendations. A substantial amount of research is being conducted under the guidance of test publishers. This has both positive and negative facets. On the positive side, the research conducted by test publishers has meant that much is known about the psychometric properties of a measure at the time of publication. More complete technical manuals mean that consumers are better able to judge the adequacy of the test. There is no question that the vast quantities of applied assessment research have resulted in measures that are better developed and standardized. However, a negative aspect is that much of this research is being approved, directed, and often interpreted by the test author and/or publisher. Clearly, due to the financial investment of the publisher, there is potential for bias. As a consequence, researchers need to have access to publishers' data, in addition to pursuing comprehensive independent investigations.

Keith (1987) makes a good point when he suggests that researchers conducting investigations with tests need to approach the enterprise with a hypothesis-testing perspective. All too often, researchers merely correlate two measures, "simple minded correlations of one test with another for no apparent reason beyond publication" (Keith, 1987, p. 276). There are signs that the sophistication of assessment researchers is improving. The increased use of theory is encouraging. The role that theory has played in the development/revision of the three most recently published cognitive measures (*Kaufman Assessment Battery for Children*, *Stanford-Binet IV*, *Woodcock-Johnson Tests of Cognitive Ability-Revised*) indicates that progress is being made. Theory derived from advances in cognitive science should inform our future hypothesis testing efforts to develop and interpret psychoeducational measures.

The question of why research is frequently ignored by practitioners is a genuine enigma. Empirical investigations have consistently failed to support the use of projective drawings, yet the use of drawings in psychological

batteries is commonplace. Likewise, misinterpretation of WISC-R scatter is rampant and is part of the clinical folklore of too many multidisciplinary teams, yet Kaufman's (1976) analysis of the standardization data and "normal" scatter was published more than a decade and a half ago.

The potential for the use of computers as assessment tools within educational environments has only partially been realized. The availability of powerful microcomputers in combination with item-response theory means that sophisticated adaptive testing procedures have the potential of revolutionizing the way assessment is conducted. Weiss (1982, 1983; Weiss & Kingsbury, 1984) demonstrated that computerized adaptive testing procedures have reliabilities and validities equal to or better than those of comparable conventional measures. By incorporating the use of item-response theory, the computer estimates trait levels and administers items based on the individual's successful or unsuccessful responses to previous items. Thus the number of items administered can be reduced by as much as 50 percent by tailoring the test to the individual. Interactive video enhances the potential of assessment by adding real-time audio and video images. Previously videotaped instructions or situations are stored on a videodisc and may be recalled almost instantaneously by the press of a key.

In conclusion, our understanding of the constructs we attempt to measure is clearer and test developers/publishers are beginning to recognize the importance of psychometric standards. Our challenge is to conduct research that allows us to view the individual we assess from a more holistic perspective, not one fragmented by a litany of separate tests.

REFERENCES

- Achenbach, T.M., & McConaughy, S.H. (1987). *Empirically-based assessment of child and adolescent psychopathology: Practical applications*. Newbury Park, CA: Sage.
- Alley, G., & Foster, C. (1978). Nondiscriminatory testing of minority and exceptional children. *Focus of Exceptional Children*, 9, 1-14.
- Anastasi, A. (1982). *Psychological Testing* (Fifth Ed.). New York: Macmillan.
- Anastasi, A. (1984). The K-ABC in historical and contemporary perspective. *Journal of Special Education*, 18, 357-366.
- Braaten, S., Kauffman, J.M., Braaten, B., Polsgrove, L., & Nelson, C.M. (1988). The regular education initiative: Patent medicine for behavior disorders. *Exceptional Children*, 55, 21-27.

- Bruininks, R.H., & McGrew, K. (1987). *Exploring the structure of adaptive behavior (Report Number 87-1)*. Minneapolis: University of Minnesota, Department of Educational Psychology.
- Bruininks, R.H., Woodcock, R.W., Weatherman, R.F., & Hill, B.K. (1984). *Scales of Independent Behavior: Woodcock-Johnson Psycho-Educational Battery - Part Four*. Allen, TX: DLM Teaching Resources.
- Cattell, R.B. (1983). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54*, 1-22.
- Clark, E., Kehle, T.J., & Bullock, D.S. (1988). Personality inventory for children: Profiles for learning disabled, emotionally disabled, and intellectually handicapped children. *School Psychology International, 9*, 43-49.
- Cole, N.S. (1981). Bias in testing. *American Psychologist, 36*, 1067-1077.
- Cummings, J.A. (1987). Academic assessment. In C.R. Reynolds & L. Mann (Eds.), *Encyclopedia of Special Education*, Vol. 1 (pp. 20-24). New York: John Wiley.
- Cummings, J.A. (1986). Projective drawings. In H.M. Knoff (Ed.), *The psychological assessment of child and adolescent personality* (pp. 199-244). New York: Guilford Press.
- Cummings, J.A., & Simon, M.S. (1988). Review of the Scales of Independent Behavior. *Journal of Psychoeducational Assessment, 6*, 315-320.
- Curtis, M.E., & Glaser, R. (1984). Intelligence testing, cognition and instruction. *International Journal of Psychology, 19*, 475-497.
- Das, J.P., Kirby, J.R., & Jarman, R.F. (1975). Simultaneous and successive syntheses: An alternative model for cognitive abilities. *Psychological Bulletin, 82*, 87-103.
- Das, J.P., Kirby, J.R., & Jarman, R.F. (1979). *Simultaneous and Successive Cognitive Processes*. New York: Academic Press.
- Davis, W.E. (1988). The regular education initiative debate: Its promises and problems. *Exceptional Children, 55*, 40-446.
- Deno, S.L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52*, 219-232.
- Deno, S.L. (1986). Formative evaluation of individual student programs: A new role for school psychologists. *School Psychology Review, 15*, 358-374.
- Ehly, S., Reimers, T.M., & Keith, T.Z. (1986). Discriminant validity of the Personality Inventory for Children: Can it identify learning disabled children? *Learning Disabilities Research, 2*, 26-31.
- Elliott, S.N., Sheridan, S.M., & Gresham, F.M. (1989). Assessing and treating social skills deficits: A case study for the scientist-practitioner. *Journal of School Psychology, 27*, 197-222.
- Gerber, M.M. (1988). Tolerance and technology of instruction: Implications for special education reform. *Exceptional Children, 54*, 309-314.
- Good, R.H., & Salvia, J. (1988). Curriculum bias in published, norm referenced reading tests: Demonstrable effects. *School Psychology Review, 17*, 51-60.
- Gresham, F.M. (1988). Review of the Revised Children's Manifest Anxiety Scale. In J.C. Conoley & J.J. Kramer (Eds.), *The Tenth Mental Measurements Yearbook* (pp. 625-630). Lincoln, NE: University of Nebraska Press.

- Gresham, F.M., & Elliott, S.N. (1984). Assessment and classification of children's social skills: A review of methods and issues. *School Psychology Review, 13*, 292-301.
- Gresham, F.M., & Elliott, S.N. (1989). *Social Skills Rating System*. Circle Pines, MN: American Guidance Service.
- Gresham, F.M., & Reschly, D.J. (1988). Issues in the conceptualization, classification and assessment of social skills in the mildly handicapped. In T.R. Kratochwill (Ed.), *Advances in School Psychology*, Vol. 6 (pp. 203-247). Hillsdale, NJ: Lawrence Erlbaum.
- Guilford, J.P. (1967). *The nature of human intelligence*. New York: McGraw-Hill.
- Harrison, P.L. (1989). Adaptive behavior: Research to practice. *Journal of School Psychology, 27*, 301-317.
- Harrison, P.L., & Kamphaus, R.W. (1984, April). *Comparison between the K-ABC and Vineland Adaptive Behavior Scales*. Paper presented at the meeting of the National Association of School Psychologists, Philadelphia.
- Heath, C.P., & Obrzut, J.E. (1986). Comparison of three measures of adaptive behavior. *American Journal of Mental Deficiency, 89*, 205-208.
- Hilliard, A.G. (1979). Standardization and cultural bias as impediments to the scientific study and validation of "intelligence." *Journal of Research and Development in Education, 58*, 47-58.
- Horn, J.L. (1985). Remodeling old models of intelligence. In B.B. Wolman (Ed.), *Handbook of Intelligence* (pp. 267-300). New York: Wiley.
- Horn, J.L. (1986). Some thoughts about intelligence. In R.J. Sternberg & D.K. Detterman (Eds.), *What is intelligence? Contemporary viewpoints on its nature and definition* (pp. 91-96). Norwood, NJ: Ablex.
- Humphreys, L.G. (1985). Race differences and the Spearman hypothesis. *Intelligence, 9*, 275-283.
- Jackson, G.D. (1975). Another psychological view from the Association of Black Psychologists. *American Psychologist, 30*, 88-93.
- Jenkins, J.R., Deno, S.L., Mirkin, P.K. (1979). Measuring pupil progress toward the least restrictive environment. *Learning Disability Quarterly, 2*, 81-92.
- Jenkins, J.R., & Pany, D. (1978). Curriculum biases in reading achievement tests. *Journal of Reading Behavior, 10*, 345-357.
- Jensen, A.R. (1984). The black-white difference on the K-ABC: Implications for future tests. *Journal of Special Education, 18*, 377-408.
- Jensen, A.R. (1985). The nature of the black-white difference on various psychometric tests: Spearman's hypothesis. *Behavioral and Brain Sciences, 8*, 193-263.
- Jensen, A.R. (1980). *Bias in mental testing*. New York: Free Press.
- Kaplan, R.M. (1985). The controversy related to the use of psychological tests. In B. Wolman (Ed.), *Handbook of Intelligence* (pp. 465-504). New York: Wiley.
- Kaufman, A.S. (1976). A new approach to the interpretation of test scatter on the WISC-R. *Journal of Learning Disabilities, 9*, 160-168.
- Kaufman, A.S. (1979). *Intelligent testing with the WISC-R*. New York: Wiley.
- Kaufman, A.S., Kamphaus, R.W., & Kaufman, N.L. (1985). New directions in intelligence testing: The Kaufman Assessment Battery for Children (K-ABC). In B. Wolman (Ed.), *Handbook of Intelligence* (pp. 663-698). New York: Wiley.

- Kaufman, A.S., & Kaufman, N.L. (1977). *Clinical evaluation of young children with the McCarthy Scales*. New York: Grune & Stratton.
- Kaufman, A. S., & Kaufman, N.L.(1983). *K-ABC: Kaufman Assessment Battery for Children: Interpretive manual*. Circle Plaines, MN: American Guidance Service.
- Keith, T.Z. (1985). Questioning the K-ABC: What does it measure? *School Psychology Review*, 14, 9-20.
- Keith, T.Z. (1986). Factor structure of the K-ABC for referred school children. *Psychology in the Schools*, 23, 241-246.
- Keith, T.Z. (1987). Assessment research: An assessment and recommended interventions. *School Psychology Review*, 16, 276-289.
- Keith, T.Z., & Dumbor, S.B. (1984). Hierarchical factor analysis of the K-ABC: Testing alternate models. *The Journal of Special Education*, 18, 367-375.
- Keith, T.A., Fehrmann, P.G., Harrison, P.L., & Pottebaum, S.M. (1987). The relationship between adaptive behavior and intelligence: Testing alternative explanations. *Journal of School Psychology*, 25, 31-43.
- Kelly, E.J. (1988). Personality Inventory for Children: Selected scales in differentiating conduct-disordered and emotionally disturbed children. *Psychological Reports*, 63, 395-401.
- Keogh, B.K. (1988). Improving services for problem learners: Rethinking and restructuring. *Journal of Learning Disabilities*, 21, 19-22.
- Knoff, H.M. (1989). Review of the personality inventory for children. In J.C. Conoley & J.J. Kramer (Eds.), *The Tenth Mental Measurements yearbook* (pp. 625-630). Lincoln, NE: University of Nebraska Press.
- Lachar, D. (1982). *Personality Inventory for Children (PIC): Revised format manual supplement*. Los Angeles: Western Psychological Services.
- Larry P. et al. v. Wilson Riles et al., C712270 (United States District Court for the Northern District of California, October 1979, slip opinion).
- Leinhardt, G., & Seewald, A.M. (1981). Overlap: What's tested, what's taught? *Journal of Educational Measurement*, 18, 85-96.
- Lieberman, L.M. (1985). Special education and regular education: A merge made in heaven. *Exceptional Children*, 51, 513-516.
- Marston, D., Lowry, L., Deno, S.L., & Mirkin, P.K. (1981). *An analysis of learning trends in simple measures of reading, spelling, and written expression: A longitudinal study* (Research Report No. 49). Minneapolis: University of Minnesota Institute for Research on Learning Disabilities.
- Martin, R.P. (1988). *Temperament Assessment Battery for Children: Manual*. Brandon, VT: Clinical Psychology.
- Mattison, R.E., Bagnato, S.J., & Brubaker, B.H. (1988). Diagnostic utility of the Revised Children's Manifest Anxiety Scale in children with DSM-III anxiety disorders. *Journal of Anxiety Disorders*, 2, 147-155.
- Mayr, E. (1982). *The growth of biological thought*. Cambridge, MA: Bknap Press.
- Mealor, D.J., & Richmond, B.O. (1980). Adaptive behavior: Teachers and parents disagree. *Exceptional Children*, 46, 386-389.
- Mehrens, W.A. (1984). National tests and local curriculum: Match or mismatch? *Educational Measurement: Issues and Practices*, 3, 9-15.
- Mehrens, W.A., & Phillips, S.E. (1986). Detecting impacts of curricular differences in achievement test data. *Journal of Educational Measurement*, 23, 185-196.

- Mercer, J.R., & Lewis, J.F. (1977). *System of Multicultural Pluralistic Assessment: Parent interview manual*. New York: Psychological Corp.
- Naglieri, J.A. (1985). *Matrix Analogies Test: Expanded Form*. Columbus, OH: Charles E. Merrill.
- Naglieri, J.A., & Jensen, A.R. (1987). Comparison of black-white differences on the WISC-R and the K-ABC: Spearman's hypothesis. *Intelligence, 11*, 21-43.
- Obrzut, J.E., & Cummings, J.A. (1983). The projective approach to personality assessment: An analysis of thematic picture techniques. *School Psychology Review, 12*, 414-420.
- PASE: Parents in action on special education et al. v. Hannon et al. No. 74 C3586 (United States District Court for the Northern District of Illinois, Eastern Division, July 1980, slip opinion).
- Reynolds, C.R. (1982). The problem of bias in psychological assessment. In C.R. Reynolds, & T.B. Gutkin (Eds.), *The handbook of school psychology* (pp. 178-208). New York: Plenum Press.
- Reynolds, C.R., & Kaufman, A.S. (1985). Clinical assessment of children's intelligence with the Wechsler Scales. In B. Wolman (Ed.), *Handbook of Intelligence* (pp. 601-661). New York: Wiley.
- Reynolds, C.R., & Paget, K.D. (1983). National normative and reliability data for the Revised Children's Manifest Anxiety Scale. *School Psychology Review, 12*, 324-336.
- Reynolds, C.R., & Richmond, B.O. (1976). What I think and feel: A revised measure of children's manifest anxiety. *Journal of Abnormal Child Psychology, 6*, 271-280.
- Reynolds, C.R., & Richmond, B.O. (1985). *Revised Children's Manifest Anxiety Scale*. Los Angeles: Western Psychological Services.
- Reynolds, M.C., Wang, C., & Walberg, H.J. (1987). The necessary restructuring of special and regular education. *Exceptional Children, 53*, 391-398.
- Saigh, P.A. (1989). The validity of the DSM-III posttraumatic stress disorder classification as applied to children. *Journal of Abnormal Psychology, 98*, 189-192.
- Salvia, J., & Ysseldyke, J.E. (1988). *Assessment in special and remedial education* (4th ed.). Dallas: Houghton Mifflin.
- Sattler, J.M. (1988). *Assessment of children* (3th ed.). San Diego, CA: Jerome M. Sattler.
- Schmidt, W.H. (1983). Content biases in achievement tests. *Journal of Educational Measurement, 20*, 165-178.
- Shinn, M.R., & Marston, D. (1985). Differentiating mildly handicapped, low-achieving and regular education students: A curriculum-based approach. *Remedial and Special Education, 6*, 31-45.
- Sparrow, S.S., Balla, D.A., & Cicchetti, D.V. (1984). *Manual for the Vineland Adaptive Behavior Scales, Interview Edition, Survey Form*. Circle Pines, MN: American Guidance Service.
- Spearman, C. (1927). *The abilities of man*. New York: MacMillan.
- Stainback, S., & Stainback, W. (1984). A rationale for the merger of special and regular education. *Exceptional Children, 51*, 102-111.
- Sternberg, R.J. (1984). The Kaufman Assessment Battery for Children: An information-processing analysis and critique. *Journal of Special Education, 18*, 269-279.

- Sternberg, R.J. (1985a). *Beyond IQ*. New York: Cambridge University Press.
- Sternberg, R.J. (1985b). Cognitive approaches to intelligence. In B. Wolman (Ed.), *Handbook of Intelligence* (pp. 59-118). New York: John Wiley.
- Sternberg, R.J. (1986, Fall). The future of intelligence testing. *Educational Measurement: Issues and Practices*, 10-22.
- Sternberg, R.J., & Salter, W. (1982). Conceptions of intelligence. In R.J. Sternberg (Ed.), *Handbook of human intelligence* (pp. 3-28). New York: Cambridge University Press.
- Thorndike, R.L., Hagen, E.P., & Sattler, J.M. (1986). *Technical manual: Stanford Binet Intelligence Scale*, 4th Ed.. Chicago: Riverside.
- Wall, S.M., & Paradise, L.V. (1981). A comparison of parent and teacher reports of selected adaptive behaviors of children. *Journal of School Psychology*, 19, 73-77.
- Wechsler, D. (1944). *The measurement of adult intelligence*. Baltimore: Williams & Wilkins.
- Weinberg, R.A. (1989). Intelligence and IQ: Landmark issues and great debates. *American Psychologist*, 44, 98-104.
- Weiss, D.J. (1982). Improving measurement quality and efficiency with adaptive testing. *Applied Psychological Measurement*, 6, 473-492.
- Weiss, D.J., (Ed.). (1983). *New horizons in testing: Latent trait test theory and computerized adaptive testing*. New York: Academic Press.
- Weiss, D.J., & Kingsbury, G.G. (1984). Application of computerized adaptive testing to educational problems. *Journal of Educational Measurement*, 21, 361-375.
- Will, M.C. (1986). Educating children with learning problems: A shared responsibility. *Exceptional Children*, 52, 411-415.
- Williams, R.L. (1974). From dehumanization to black intellectual genocide: A rejoinder. In G.J. Williams & S. Gordon (Eds.), *Clinical child psychology: Current practices and future perspectives*. New York: Behavioral Publications.
- Wirt, R.D., Lachar, D., Klinedinst, J.K., & Seat, P.D. (1984). *Multidimensional description of child personality: A manual for the personality inventory for children* (Revised 1984 by D. Lachar). Los Angeles: Western Psychological Services.
- Witt, J., & Elliott, S. (1986). Acceptability of classroom interventions. In T.R. Kratochwill (Ed.). *Advances in school psychology*, Vol. 4. Hillsdale, NJ: Lawrence Erlbaum.
- Woodcock, R.W. (1989, August). *Factor structure of the tests of cognitive ability from the 1977 and 1989 Woodcock-Johnson*. Paper presented at the ACER seminar on Intelligence. Melbourne, Australia.
- Ysseldyke, J.E., & Christenson, S.L. (1987). *The Instructional Environment Scale: A comprehensive methodology for assessing an individual student's instruction*. Austin, TX: Pro-Ed.

Jack A. Cummings, Associate Professor and Director of the School of Psychology Program at Indiana University (Bloomington), has conducted research on variables associated with psychologists' decisions to place children in special education classes, as well as criterion validity investigations of cognitive ability measures. The development and evaluation of interactive video programs for preservice psychologists to refine their interviewing skills have been the focus of his efforts for the past two years.

Marc C. Laquerre is a doctoral student in school psychology at Indiana University (Bloomington). His former studies were at College of the Holy Cross, Worcester, Mass. Mr. Laquerre's research interests include visual-motor assessment, adolescent substance abuse, and therapeutic interventions with families and children.

Jack A. Cummings, professeur agrégé et directeur du programme de psychologie scolaire à l'université d'Indiana (Bloomington), a effectué des recherches sur les variables qui entrent dans les décisions prises par les psychologues de placer certains enfants dans des programmes d'enseignement spécial et sur la validité des critères de mesure des facultés cognitives. Depuis deux ans, il travaille à la conception et à l'évaluation de programmes vidéo interactifs en vue de permettre aux psychologues de peaufiner leurs techniques d'entrevue avant l'emploi.

Marc C. Laquerre est étudiant de doctorat en psychologie scolaire à l'Université d'Indiana (Bloomington). Il a fait ses études au College of the Holy Cross, à Worcester (Mass.). Ses recherches portent sur l'évaluation visuelle motrice, l'abus de substance par les adolescents et les interventions thérapeutiques auprès des familles et des enfants.



Robert L. Thorndike
Teachers College, Columbia University

International Research on School-Based Assessment

The five articles in this issue provide a dramatic illustration of the wide diversity of conditions under which psychologists concerned with the assessment of children undertake research and provide service. One is amazed that Dr. Saigh found it possible to gather and analyze data in strife-torn Lebanon, and his report provides testimony to the devotion with which he pursued his endeavours under most difficult conditions. In the People's Republic of China, also, political strife and turmoil have hardly provided a fertile soil for assessment research. By contrast, the United States is swarming with school psychologists, some providing service in the schools and some who are situated in universities where research and development is encouraged and supported.

Though the authors of the articles in this issue tend quite generally to affirm that assessment should be comprehensive, synthesizing information from a range of different sources, it is clear that the central concern of assessment in the educational context for most countries, and especially of assessment research, has been and will continue to be appraisal of the general level of the pupil's cognitive functioning. Assessment research on instruments or procedures can be directed at any one of three levels: (1) development of an instrument or procedure and internal analysis of its properties, (2) studies of the effectiveness of the instrument or procedure as a predictor of significant educational outcomes, or (3) studies of the usefulness of the instrument or procedure in guiding interventions to improve learning. We may consider each of these in turn, first as they apply to psychometric type instruments that yield one or more scores and then as they relate to less structured approaches such as observation and interview procedures.

The prototype of the psychometric score-yielding approach is the test of general cognitive ability, or intelligence if you will. Tests of this sort are

reported for each of the countries represented in this issue, but such instruments are probably most common in the United States. The internal analyses embrace those that are typically included in an adequate technical manual for the instrument. They include provision of adequate norms based on a representative sample of cases, evidence on the reliability of the resulting scores, and evidence to clarify the nature of the attribute or attributes that the instrument appraises. This last often involves correlational analysis of the several subscores that the instrument provides to determine to what extent all are expressions of a common underlying ability and to what extent distinct abilities are involved, and to find out how well any distinct factors confirm the theoretical constructs assumed by the authors of the instrument. It also commonly involves relating scores on the new instrument to other measures thought to appraise the same or similar attributes. Much of the literature of ability testing is of this sort.

When, as is reported by three of the contributors, a test from one national setting is adopted for use in another, questions arise as to the transferability of the test (or other instrument) to the new national setting. Where the language is the same and the cultures are as similar as is the case with Canada and the United States, the problems seem minimal, so a sensitive review by Canadian editors sufficed to convert the *Cognitive Abilities Test* which had been developed in the United States into the *Canadian Cognitive Abilities Test*. New norms are needed for the adopting country, but most of the other evidence from the country of origin on the properties of the instrument seems directly transferable. In the case of a measure of academic achievement, in addition, a careful review of the appropriateness of the content of the test to the curriculum of the receiving country seems called for.

With larger cultural differences, and especially where translation to a different language is required, a complete re-evaluation of the instrument in the adopting country would seem essential. This is brought out by LaVoie's discussion of the use of the *Wechsler Intelligence Test for Children-Revised* (WISC-R) in China, where subtest scores varied dramatically from those obtained in the United States and certain subtests appeared to have markedly low relationships to total IQ, though the test as a whole still functioned as a good predictor of academic achievement. By contrast, in Israel the complete evaluation of the Israeli adaptation of the WISC-R appears to have confirmed the appropriateness of the original test with only minimal changes.

The reasonably satisfactory effectiveness of general aptitude tests as predictors of school achievement has been repeatedly demonstrated over the years. It is confirmed by each of the authors in the present issue. Questions arise, however, as to whether a given level on the aptitude measure signifies

the same level of expected achievement in all segments of a particular society. This is the ever-present issue of test “fairness” or test “bias.” Does a particular IQ in an 8-year-old, for example, forecast the same level of mathematics achievement at age 10 for black and for white children? For children from upper socioeconomic levels and those from low socioeconomic levels? For the child of recent immigrants from East Asia as for a native-born child? These are all groups that show differences in mean score on the aptitude tests, and these differences between groups describe a situation that does exist. The question is whether there are corresponding differences in other aspects of performance so that the score provides a fair forecast of the expectations for the individual’s future. This is a problem that Canada is aware of in relation to its aboriginal population and to its immigrants from many parts of the world. It is one that Israel faces as immigrants from Asia and Africa are compared with those from Europe. It is one that the United States faces with its black, Hispanic, and native American groups as well as with its recent flood of immigrants from East Asia.

There is certainly no universal answer to the question of “fairness.” The answer will differ for different groups and for different circumstances. Where prediction is limited to academic achievement, the evidence tends to indicate that there is no underprediction of the later academic performance of black pupils in the United States. However, for non-English-speaking groups the situation may be different. And if optimal interventions are introduced, especially at a very early age, the original expectations may need to be significantly modified. Zeidner reports that in Israel the substantial differential between Eastern and European immigrant groups is diminished in second generation Israelis, and this suggests that in time the difference might largely disappear. It is unfortunate when the existence of group differences in mean test scores leads to wrangling about the use of the tests, rather than to a constructive search for interventions that might reduce the differences, or serve to disconfirm the predictions that the test scores currently support for both the majority and minority groups.

This leads to our third type of research – research seeking to determine in what way the measures of aptitudes can guide interventions that will enhance pupil achievement. This field of research has sometimes been spoken of as “trait-treatment interaction,” modifying the educational treatment to fit it to the appraised standing on the predicting trait. As tests of more specific abilities – visual, spatial, verbal, quantitative – have multiplied over the past 50 years, the hope has been expressed that particular forms and organization of instruction might be found to prove especially effective for persons with particular patterns of abilities. However, in their review of the research literature, Cronbach and Snow could find little evidence that this was in fact

the case. What evidence of interaction they did find was largely between level of general cognitive ability and form of instruction. There is some indication that higher aptitude children progress better with discovery approaches to learning while low aptitude children make better progress with relatively highly structured presentations. In addition, some research indicates that low aptitude children lack strategies for learning, and benefit from systematic instruction and practice in the use of learning strategies.

From the beginning days of testing, grouping by ability, or "streaming," was undertaken as a form of adaptation to differences in ability level, but the accumulated evidence, going back almost 70 years now, shows little indication that children learned more in homogeneous than in heterogeneous ability groups. Possibly this was because little perceptive adaptation of instruction was made to fit the characteristics of the more homogeneous group. In the final analysis, it is in instructional adaptations to individual children that result in greater learning that the justification for testing must be found, and research effort for the future should be directed to identifying such adaptations in a form that teachers can and will use.

The same three types of queries can be addressed to inventories, questionnaires, and other types of procedures designed to assess personality characteristics. What attribute or attributes does the procedure appraise and how well does it appraise them? What outcomes of concern to society can the procedure predict? And what guidance can it provide for constructive educational interventions?

Whenever an instrument or procedure yields one or more scores that are conceived to provide an index of some describable and definable attribute, the psychometric properties of the score or scores can be studied in essentially the same way that has been applied to the familiar instruments for measuring abilities. A good deal of the research reported here from the different countries is of this sort, dealing with evaluation of internal consistency, stability over time, and concurrent correlates of the several scores derived from various inventories and questionnaires. Reference is also made to a good deal of work on establishing adequate norms to which scores of individuals may be referred. As academic correlates are sought for the various measures of attitude and adjustment, the bulk of the work appears to deal with concurrent relationships. One could wish for a clearer picture of the extent to which these instruments permit a prediction of further educational progress and provide a guide to constructive intervention.

To take one specific example, reference is made to measures of "test anxiety," and some moderate negative correlations are reported between

reported anxiety and academic performance. Leaving aside for now the question of whether pupils do poorly because they are anxious or are anxious because they are doing poorly, one would like to see studies that use the anxiety score as the basis for some describable modification of the instructional situation and which follow up to determine whether measurable improvement results.

Scores and ratings of personal characteristics are presumably gathered in order to introduce constructive interventions in the education of the child. In a few instances (e.g., Gresham and Reschly as reported by Cummings and Laquerre) explicit attempts are made to relate types of adaptive behaviour deficits to the appropriate treatment. It is not clear, however, that a systematic research design has been set up to apply the recommended intervention and to follow up in order to determine how effective it has been.

It must be admitted that it would be extremely difficult to even begin to achieve rigor in research on a problem such as this. Cases presenting a certain personality deficit, for whom a particular intervention appeared to be indicated, would be likely to be relatively few in number and scattered. Providing adequate monitoring of the manner in which the specified intervention had been introduced, defining the nature of the changes in academic or personal progress that were to be looked for, and setting up procedures for appraising the extent to which the sought-for progress had been achieved is indeed a formidable undertaking. It is hardly surprising that it has seldom, if ever, been carried out.

In several of the papers in the series there is an affirmation of the view that assessment should be comprehensive and should include, and in fact rely fairly heavily on, nontest, nonquestionnaire types of procedures – such approaches as direct observations of pupil behaviour and/or interviews with parents and teachers. These more fluid approaches are felt to add important information that is not well obtained by more formal and structured approaches. The output from these procedures may often be a discursive, ideographic account of the assessor's impressions as derived from all sources of information about the child.

In the case in which observation or interview leads to some type of standard datum, as when an observer records the proportion of "time on task," or the frequency of certain defined categories of aggressive actions toward other children, or when an interviewer, or the reader of an interview protocol, rates the home for one or more attributes such as "degree of parental support of learning," that datum is in theory susceptible to analysis in much the same way as any other score. Evidence of reliability can be sought by taking

successive behaviour samples or by getting a protocol scored by more than one independent rater. Generalizability can be tested by getting behaviour samples in different settings. And the current correlations of the score with academic or other aspects of school performance can be determined. Though time-consuming to obtain and unwieldy to work with, these data can be dealt with in the same way as test scores or any other quantitative data.

But it may well be contended that the effort to generate scores from the individual and discursive synthesis that constitutes the essence of informal assessment procedures loses what is vital to the procedures and is a travesty on true clinical assessment. It may be asserted that the assessment must be dealt with directly in its complex, discursive qualitative form. To the extent that this is accepted, any attempt to do research on the procedures faces the most formidable obstacles. What is the independent variable whose impact is to be determined? Is it simply the global fact that an assessment has been carried out compared with no assessment, and that some sort of intervention was introduced? Would it be technically feasible and at the same time ethically defensible to set up a comparable control group from whom assessment had been withheld in spite of their presenting problems similar to those of the children sent for assessment, and then to compare the two groups on some index or indices of educational or personal progress? It is not to be wondered that there is a dearth of such research.

The formidable problems of conducting research on these discursive, ideographic approaches to assessment make it seem likely that there will always remain an unbridgeable gap between assessment research and assessment practice. Research will be directed primarily at those procedures that generate some type of score or scores that can be obtained for each member of a group and can be related to other facets about that individual – to other aspects of input or to aspects of outcome. Practice will, in part at least, be concerned with a nonquantitative synthesis that is unique to the particular individual and that gets its vindication from the fact that it appears to provide one or another type of benefit to the pupil or comfort to the teacher. It appears that this disparity must be recognized and accepted.

Robert L. Thorndike, Professor Emeritus, Teachers College, Columbia University, received his higher education at Wesleyan University (1931) and Columbia University (1935). He was Professor of Psychology at Teachers College, Columbia University from 1935 to 1976. He served in the U.S. Army Air Corps Aviation Psychology Program (1942-46), retiring with the rank of Major. He has served as president of the Psychometric Society, Educational Research Association, American Psychological Association (Divisions 5 and 19). Among his many publications are: *Personnel Selection* (1949); *Concepts of Over- and Underachievement* (1964); *Applied Psychometrics* (1962); co-author of *Evaluation and Measurement in Psychology and Education* (with Elizabeth Hagen), 1977; *10,000 Careers* (1959); and *The Cognitive Ability Tests (Form 4)* (1985).

Professeur émérite au Teachers College de l'Université Columbia, **Robert L. Thorndike** a fait ses études supérieures à l'Université Wesleyan (1931) et à Columbia (1935). Il a été professeur de psychologie au Teachers College de l'Université Columbia de 1935 à 1976. Il a servi dans l'Armée de l'air américaine dans le cadre du programme de psychologie (1934-1946) et en est sorti avec le rang de major. Il a été président de la *Psychometric Society*, de l'*Educational Research Association*, de l'*American Psychological Association* (divisions 5 et 19). Parmi ses nombreuses publications, signalons *Personnel Selection* (1949), *Concepts of Over- and Underachievement* (1964), *Applied Psychometrics* (1962); il est co-auteur de *Evaluation and Measurement in Psychology and Education* (avec Elizabeth Hagen), (1977), de *10,000 Careers* (1959) et de *The Cognitive Ability Tests (formulaire 4)* (1985).



Jerome M. Sattler
San Diego State University

Philip A. Saigh
*Graduate Center,
City University of New York*

School-Based Assessment Research in Five Nations

This special issue of the *McGill Journal of Education* is devoted to school-based assessment research in Canada, People's Republic of China, Israel, Lebanon, and the United States. The focus of each paper is primarily on psychoeducational assessment issues that involve standardized intelligence, achievement, and personality/temperament tests used to evaluate children having difficulties in school. The nations surveyed are located in North America, the Middle East, and China. Readers, therefore, will be able to gain some understanding of current trends in psychoeducational assessment research in diverse regions of the world.

School-based assessment practices have been shaped by the efforts of investigators from such countries as Great Britain (e.g., Sir Francis Galton, Karl Pearson, Charles Spearman, and Sir Cyril Burt), France (e.g., Alfred Binet, Victor Henry, Edward Seguin, and Theodore Simon), Germany (e.g., Emil Kraepelin, Herman Ebbinghaus, and William Stern), and the United States (e.g., James Cattell, Henry Goddard, Louis Terman, Edward Thorndike, Leon Thurstone, David Wechsler, and Robert Thorndike). Although in recent years the United States has been at the forefront of psychoeducational research, other countries continue to contribute to the development of psychoeducational and clinical assessment procedures.

There are notable variations in the scope and quality of school-based assessment research in the five nations surveyed. The countries considered herein vary with respect to legislation involving mandatory education, assessment of students with special needs, and accountability for student achievement. Moreover, considerable variation exists with respect to the availability of measurement personnel, financial incentives, university systems that reinforce research productivity, national security, and public attitudes toward testing.

Despite these variations, the influence of tests developed in the United States is quite apparent in Canada, China, Israel, and Lebanon. But these four nations also are striving to develop their own indigenous products and have been successful in some areas. Simple translations of English-language tests, standardized in the United States or Canada or England, for example, may not be appropriate in other countries without significant modifications. Test translations are hazardous without careful research into the meaning of the translated items. Psychometric issues – such as item difficulty levels, reliability, and validity – must also be considered. Still, tests developed in the United States are used for psychoeducational assessment purposes in translated form in many countries throughout the world.

It is of interest to observe that Chinese and Israeli investigators have been able to cross-validate a number of individually administered intelligence tests that were developed in the United States. Canada, Lebanon, and Israel also have developed their own group intelligence tests, such as the *Canadian Cognitive Ability Test*, *Lebanese General Ability Scale*, and *Milita Group IQ Test*, respectively. Countries need to be encouraged to continue to develop their own tests in order to serve their own specific purposes. Given that group tests are generally easier to construct, administer, and validate than individually administered tests, investigators in other countries may wish to concentrate their resources on developing group measures and on cross-validating group ability tests developed in the United States. However, for psychoeducational (and clinical) activities, individually administered tests are especially useful.

The issue of test bias is an important concern, not only in Canada and the United States, but in the other nations surveyed in this issue and likely in other nations as well. Test bias is of special importance in nations that have a multiethnic and multilingual society. How to deal with potential test bias, as we know from years of litigation and research in the United States, is by no means easy. Voices in Israel, as Moshe Zeidner points out, are as vociferous and demanding as voices in the United States. These voices are saying that all tests must be abolished because they are culturally unfair to segments of the population. Yet, as D.H. Saklofske and H.L. Janzen seem to be saying, Canada has more moderate voices. Psychologists and educators in Canada recognize that there are difficulties in using tests with diverse ethnic groups, but if used wisely, tests do contribute to the assessment process.

Who should assume the primary responsibility for carrying out the assessment of special children? This issue is by no means resolved in the five nations surveyed. Psychologists as well as teachers are responsible for carry-

ing out psychoeducational assessments in the United States, Canada, and Israel; in Lebanon and China teachers appear to assume more of this responsibility because there are few specially trained personnel to perform the psychoeducational assessment. Will these trends continue in the future?

J. LaVoie's review of school-based assessment research in the People's Republic of China brings out clearly the issue of values and goals associated with assessment activities. In the People's Republic of China the focus is on placement, with teachers trying to help special children work up to the level of their classmates. How often do the teachers reach their goals? Do children who are not performing adequately in classrooms in the People's Republic of China improve their performance after they have been given special attention by their teachers? How do the success rates of the People's Republic of China compare with those in other nations where the focus is on understanding the individual child? This is an important issue for future cross-cultural research.

In a nation like Lebanon, war-ravaged and on the brink of destruction, it is a wonder that any research could have occurred. Children and their families in such circumstances need to place all their energies on survival. It is a testimony to the human spirit that education, research, and scholarship continue under such horrible and inhumane conditions. Given the current situation, it is understandable that research productivity is at a standstill. Although it is difficult to make predictions about the Middle East, it is thought that P.A. Saigh's more favourable prospectus would lead to a rekindling of scientific inquiry.

J.A. Cummings and M.C. Laquerre note that in the United States research in the area of psychoeducational assessment is progressing on many fronts, including intellectual functioning, academic functioning, adaptive behaviour, and social-emotional functioning. The very need for special education is being questioned, but what are the alternatives? Cummings and Laquerre wisely point out the need for researchers not associated with test publishers to independently study the psychometric properties of newly developed assessment instruments.

These five surveys point out that different nations recognize that children who are performing poorly in school need to be identified, evaluated, and given special help. The procedures used in each nation differ depending upon the political, social, and philosophical views of its people and governing bodies. But in spite of the differences among nations, there are similarities as well. Psychologists and educators recognize the continuing need to develop reliable and valid assessment procedures that take into account the language,

thought processes, and culture of children in need of help. This survey attests to the vitality of research on how special children are assessed and identified in different countries around the world.

Jerome M. Sattler is a professor of psychology at San Diego State University where he has taught since 1965. Dr. Sattler is a Diplomat in Clinical Psychology and a Fellow of Division 16 (School Psychology) of the American Psychological Association. He served as an expert witness in the Larry P. v. Wilson Riles case involving the use of intelligence tests for the assessment of black children in California public schools. He has contributed over 90 publications to the fields of clinical and school psychology and psychoeducational journals. Additionally, he has delivered over 140 invited lectures and workshops. In 1972 he was Senior Fulbright lecturer at the University of Kebangsaan in Malaysia. He is the author (and publisher) of *Assessment of Children*, Third Edition. This commentary was written during his stay at University College Cork, Cork, Ireland, where he is a visiting professor.

Philip A. Saigh is a professor and head of the doctoral program in school psychology at the Graduate Center of the City University of New York. He previously served on the faculty of the American University of Beirut from 1977 to 1986. Professor Saigh is a Fellow of the American Psychological Association (Division 16, School Psychology), and his research interests primarily involve the assessment and treatment of posttraumatic stress disorder. Many of his publications have been in that area as well as in the area of school-based assessment.

Jerome M. Sattler est professeur de psychologie à l'Université d'État de San Diego, où il enseigne depuis 1965. Il est diplômé de psychologie clinique et membre de l'American Psychological Association (division 16, psychologie scolaire). Il a témoigné à titre d'expert dans l'affaire Larry P. c. Wilson Riles, à propos de l'évaluation d'enfants noirs au moyen de tests d'intelligence dans les écoles publiques de Californie. Il compte plus de 90 publications dans les domaines de la psychologie clinique, de la psychologie scolaire et de la psycho-pédagogie. Il a en outre été invité à prononcer ou à diriger plus de 140 conférences et ateliers. En 1972, l'Université de Kebangsaan (Malaisie) l'a invité à titre de conférencier principal dans le cadre de son cycle de conférences Fulbright. Il est l'auteur (et l'éditeur) de *Assessment of Children* (3^e édition). Cet article a été écrit pendant le séjour qu'il a fait au University College Cork, en Irlande, à titre de professeur invité.

Philip A. Saigh est professeur et directeur du programme de doctorat en psychologie scolaire au Graduate Center de la City University of New York. Il a enseigné à l'Université américaine de Beyrouth de 1977 à 1986. Il est membre de l'American Psychological Association (division 16, psychologie scolaire). Ses recherches portent principalement sur l'évaluation et le traitement des troubles post-traumatiques. Bon nombre de ses publications portent sur ces questions et sur le problème de l'évaluation en milieu scolaire.

Lewis B. Birch
McGill University

Commentary on the Burt Affair

In 1965, when Sir Cyril Burt was over 80, a book with the Greek title *Stephanos* was edited by Charlotte Banks and P.L. Broadhurst as a tribute to him. There were more than a dozen papers in the book; each was written by a distinguished writer, each covered a different aspect of his wide range of interests, though not covering them all. The first paper in the *Festschrift* was a short biography and appreciations by C.W. Valentine, Burt's oldest friend, their friendship dating back to the time they were both students under Külpe in Würzburg. He begins by asserting that Burt was one of the half-dozen greatest psychologists this century has produced, then he outlines the unique width and depth of Burt's scholarly activities, his pioneer researches, his publications and editorships; he writes too of the eminent people holding important academic positions in many parts of the world who had been his students. Valentine especially emphasizes Burt's quite exceptional generosity in giving his time, effort, and expertise in helping students, colleagues, and public bodies like the BBC and the British Association for the Advancement of Science.

When Burt died in 1971 the position of eminence established in *Stephanos* was confirmed and enhanced by the tributes in obituary notices and in the memorial service at St. Mary's Church, Primrose Hill. Burt had been knighted in 1946, and in 1971 he was presented by the **American Psychological Society** with the Thorndike Award; this honour had never previously been made to a non-American. However, in spite of these and many more honours, Burt, in his later years, received a great deal of hostile criticism particularly related to his views on the nature of intelligence, the nature-nurture controversy, and educational selection. The criticism followed and was related to the general change in thinking about education which developed in the years towards the end of the last war. In the post-war

reconstruction in England, children at the secondary level were to be categorized into types: academic, technical, or practical and to be segregated in three kinds of school. The government could offer little help in defining or selecting children for the three types of school or of the ways which the schools would teach, and, what they would teach. Burt, not surprisingly, objected strongly on psychological grounds (Burt, 1943), but the system remained more or less in place for about 20 years. (The "more or less" refers to the failure of the "technical" schools to materialize in quantity.)

The tripartite system of schooling aroused considerable opposition, specially from left wing political bodies and sociologists like B. Simon, J.E. Floud, A.H. Halsey, and J.W.B. Douglas. They wished to abandon all kinds of educational selection on the grounds that it was a means of perpetuating an elitist, class-stratified society; upper-class children being favoured by the selection procedures, especially IQ tests and such like, used to place the children in the secondary schools. Brian Simon (1953) even went so far as to suggest that children do not have different levels of intellectual ability and hence they could virtually all be educated together. Burt, it is well known, believed that sooner or later selection would have to be employed because children differed so widely that they could not be effectively taught together at the secondary stage. Arguments about the conflicting views became violent; they became political and the press was involved. Support for Burt came mainly from academics opposed to the government's policy. Their chief concern was over maintaining educational standards if selection were not practised; the vehicle for expressing their views was a series of pamphlets called the *Black Papers* (Cox & Dysn, 1969, 1970). Burt contributed long, characteristically scholarly, well researched papers in two of them. The *Black Papers* were well received by many politicians, and by the more serious newspapers and technical journals. Burt was pressed to write more articles on the topics of the *Black Papers*; it is said he wrote fourteen articles in 1969 (Hearnshaw, 1979) in newspapers, weeklies, and technical journals. These reflected the tense atmosphere of the time with disputants loosely connected with left and right political opinions which were strongly held and exaggeratedly expressed. A somewhat similar atmosphere developed about the same time in America with people like S.J. Gould and L.J. Kamin on one side and A. Jansen and R.I. Hernstein on the other.

In the absence of other stimulants it seems likely that interest in Burt would have waned but in 1977 an article appeared in the British *Sunday Times* by that paper's medical correspondent, Oliver Gillie. It started: "The most sensational charge of scientific fraud in this century is being leveled against the late Sir Cyril Burt, father of British educational psychology. Leading scientists are convinced that Burt published false data and invented crucial

facts to support his controversial theory that intelligence is largely inherited." The main specific charges were that Burt had guessed parental IQs, that he had invented data to support his own genetic theories, that his quoted correlations between sets of twins were bogus, and that two of his named collaborators had probably never existed. The *Sunday Times* article was followed by another, in *The Times*; shortly, the truth of the statements was supported by a number of psychologists and was vehemently denounced by many more. Neither side could establish the truth of Gillie's charges, or of the others which cropped up in the spate of letters to the press and in discussions. It was therefore suggested that until L.S. Hearnshaw, who had been invited to write a biography of Burt, had completed his task, no judgement should be made.

Hearnshaw's biography came out in 1979. In it he found that, in the main charges and in some other respects which he had found himself, Burt was guilty. Such was Burt's stature that news of the findings was widely publicized throughout Britain and many other parts of the world including the United States and Canada. Hearnshaw had written so well, his research of the literature was so extensive, and his arguments so convincing that reviewers of the book almost without exception accepted his judgement. Eventually, J. Tizard and A. and A.D.B. Clarke, who had from the beginning supported the view of Oliver Gillie, were joined by other psychologists like P.E. Vernon, S. Chown, and H.J. Eysenck; further, the Council of the British Psychological Society accepted Hearnshaw's verdict. Hence in ten years Burt's reputation had fallen from the peak of eminence to the gutter.

After the publication of the biography the denigrating writings and broadcasts subsided but there were sequels. References to Burt were generally omitted from text books of educational psychology, some referred to him but commented on his disgrace; in reporting on such topics as twin studies they would append "after deleting the fraudulent data of Burt" or some such phrase. A few people who had been specially antagonistic towards Burt had taped cassettes of a CBS program in the series "60 Minutes" featuring Liam Hudson demeaning Burt. In England tapes of a particularly scurrilous BBC production still seem to exist.

From time to time I received news in Canada from English psychologists and educators that enquiries had been made which seemed to show that Burt had been seriously misrepresented. Because I could not trace relevant work in Canada, Professor Hazel Francis, editor of the *British Journal of Educational Psychology* was approached. She was good enough to send to me a copy of an uncorrected proof of a paper by Ronald Fletcher called the "Doubtful Case of Cyril Burt", published in 1987 in *Social Policy and Administration* (Vol. 21), a British journal not usually read by most educa-

tional psychologists. The editor of the *Journal* prefaced the paper by his own comments that it was concerned with "...investigating the truth of one of the most dramatic controversies in the social sciences in British universities in recent years." He continues: "Here Fletcher raises serious doubts about the validity of the attack on Burt and questions the quality of the evidence produced by his denigrators."

It would seem that from the start Fletcher had been uneasy about the testimony of the detractors. "...[T]he accusers seemed...to be waging a war rather than pursuing an argument; but it was the BBC's television dramatization in 1984 which finally persuaded me that something was seriously at fault." In particular he felt that the attack on Burt was an organized campaign by a small cohesive group determined to destroy Burt as a scientist and to smear him as a person. He noticed that in the film people who knew Burt well were not called upon to offset the evidence of people like Tizard, the Clarkes, and L.J. Kamin. Even A.R. Jensen and R.B. Cattell, old friends of Burt, who appeared in the film in short clips, seemed to be supporting the notion that Burt was a psychopathic fraud. So, Fletcher wrote to them both and both replied protesting that they had been grossly misrepresented by the BBC reporters who went to Berkeley and to Hawaii to interview them, and by the editors of the film. Both psychologists had been assured that the program was to be a balanced, non-judgemental documentary. Both were very angry at the result and Cattell, who only had second-hand reports of the broadcast, complained that he had failed to get a copy of the cassette even after writing twice to the BBC. The comments of Cattell and Jensen provoked Fletcher to investigate further the allegations made against Burt and the people who made them and who reported them.

First Fletcher gave brief but effective outline of Burt's main contributions in the area of the controversy, particularly stressing his major purposes, his methods of work, the sophisticated mathematical tools he devised to identify individual differences through which he could provide appropriate opportunities for all children. One of Burt's main conclusions was that inequality of incomes seemed to be largely, though not entirely, an indirect effect of the wide inequality in innate intelligence. This hypothesis was reached as long ago as the last war; by the end of his life it was opposed by many, and particularly violently by many of the people who were his main accusers. In pursuing his investigation of these accusers Fletcher uses a Socratic method, posing searching, clearly annunciated questions like "Why...did Hearnshaw pay such little attention to the evidence and representations of Burt's supporters?" He then follows by comments and presentation for evidence of any conclusions. With the case of the Clarkes and the two "articles" of theirs which Burt was said to have written with serious adverse

criticisms of Eysenck, Fletcher comments: "Astonishingly, no such articles existed!" They were, in fact, not articles but thesis abstracts which, under the written instructions of the editor, were submitted through the department where the thesis was presented. Further, they do not contain a criticism of Eysenck.

In total, there seems little doubt that Fletcher has shown that certain psychologists and the media must explain why it appears they used very disreputable methods in blackening Burt's repute; why they made serious charges without checking the authenticity of the evidence. He showed that some of the charges were untrue, others not proven. He pointed to the need for further detailed enquiry and study of the allegations made against Burt. In two years this need was largely satisfied.

In 1989, Robert B. Joynson, a reputable historian from Nottingham University who also has qualifications in psychology, published a book, *The Burt Affair*. The author explains that he had not known Burt and was not interested in the psychology of individual differences or educational problems; he came to read Hearnshaw's biography because he was enquiring into the origins of intelligence testing. He found that Hearnshaw held that one of Burt's early deceptions concerned the history of the relationship of his own work to that of Charles Spearman. Joynson meticulously checked the early publications germane to the claims of Burt and found that in all four of the items he examined Burt's accounts were correct. He writes: "When I became convinced that Burt had not been guilty of any deliberate historical falsification, it was unthinkable that I should not go on to examine the further allegations against Him" (p.viii). Even at this stage he was of the opinion that "a grave injustice has been done."

Joynson decided to look further for signs of Burt's alleged distortions by reading his *Young Delinquent* (1925) for the Clarkes and Kamin had claimed that Burt had always been a devious, dishonest scientist. Nothing dishonest or devious could be found; it was perhaps old-fashioned, but in the light of the expected readers and the facilities available in 1925 it was altogether admirable. Next he looked at Hearnshaw's assertion that Burt had distorted the history of the development of factor analysis and the identification of general intelligence. It was investigated in minute detail and Joynson concluded that Hearnshaw's "historical accusations were poorly established" and his account of what had happened was in error "not occasionally and incidentally but repeatedly and crucially;" in fact, "Hearnshaw has misunderstood Burt's account."

Burt was quite viciously attacked over his twin studies in which he compared, among other things, the IQ of twins brought up with their own parents and others reared apart. Kamin and others noted that a number of

anomalous values appeared in the tables of correlation coefficients published in many different articles. The number of monozygotic twins reared apart, for example, increased from 15 to 53 but the coefficient remained the same. Joynson tediously examined all the relationships tracing the correlations in dozens of tables by methods too complex to describe here. In the end he concludes: "The invariant correlations do not provide evidence of deliberate manipulation of data on Burt's part. It seems quite likely that the strange values are often copying errors and the enquiry provides fresh grounds for supposing that Burt's most important data are genuine. It should be recalled too that Burt was over eighty years old when he wrote the main papers on twins in 1966.

One of the alleged frauds which was much referred to by the general public after reading the Gillie article, since it added a little spice to a rather mundane story, concerned the case of the two missing assistants. Burt refers to these two ladies who helped him for many years, but now they cannot be found or identified. The Clarkes hold that the women never existed; that both they and the twins data were fabricated by Burt. Joynson points out that they would have been quite old and probably died before Gillie began looking for them in 1976. In any case they were not of any great moment, except as making a good story for a reporter.

There are two chapters of Joynson's book which look at another kind of attack on Burt. Hearnshaw and others had hypothesized that Burt's fall from grace was due to a "psychological disturbance," a marginal "paranoia." He then goes to great lengths to try to show how the condition developed. Joynson argues that the explanation is "little more than an invention and selection of causes and consequences which, if true, would be highly convenient for psychology." But, many people who knew him well did not notice any disturbance; minor eccentricities perhaps, but no more. Similarly, Hearnshaw built up a picture of Burt's dual character; on the one hand cantankerous, belligerent, devious, unscrupulous; on the other courteous, helpful, considerate, honest. Joynson thinks that Hearnshaw used selective reporting to build the image of Burt's character. This could be done by asking the opinion of his friends on the one hand and his detractors on the other – but would the picture mostly reflect the respondents or Burt?

One important aspect of the Burt affair which Joynson discusses is the reaction of the psychological world to it. He wonders why the accusations were received so easily by the President and Council of the **British Psychological Society** and, reluctantly, by such admirers of Burt as Vernon and Jensen. This makes fascinating but not wholly convincing reading. It does not

deal adequately with the other side of the matter. The rehabilitation of Burt inevitably involves the laying of blame for various kinds of misdeeds on the part of the attackers and those who did not find the time, inclinations, or competence to undertake serious study of the evidence of Burt's alleged frauds. I wonder, had I still been a member of the Council of the **British Psychological Society**, whether I should have gone along with the President and accepted the verdict of the biography or whether I should have had the fortitude to do the kind of thing which Fletcher and Joynson did.

The Burt affair evokes theses which are highly selective in the material that is chosen to establish the point of view of the writer. This is probably the crux of the whole Burt affair. All who write about Burt's integrity must display some degree of bias, therefore it is only fair that I should declare my own position.

I became one of Valentine's students just after the last war; I became his assistant editor in 1952 and ultimately for six years, till 1967, I was editor of the *British Journal of Educational Psychology*. Through Valentine I met Burt quite frequently, visiting him quite often, including the time when he was in hospital with a broken leg. I admired him enormously and was very grateful to him for his ever available help in running the Journal. He gave me quite a few books, always graciously inscribed. My last gift has two inscriptions; one says

"A true friend may be reckoned the masterpiece of nature."
Of Friendship, R.W. Emerson

Robert B. Joynson.
THE BURT AFFAIR.
London & New York: Routledge, 1989.
pp. xiii + 347. £25.

REFERENCES

- Banks, Charlotte, & Broadhurst, P.L. (Editors). (1965). *Stephanos*. Studies in Psychology, presented to Cyril Burt. London: University of London Press.
- Burt, Cyril. (1925). *The young delinquent*. London: University of London Press.
- Burt, Cyril. (1943). Ability and income. *British Journal of Educational Psychology*, 13, 83.
- Clarke, Ann M., & Clarke, A.D.B. (1977). Correspondence, *Bulletin of the British Psychological Society*, 30, 83.

- Cox, C.B., & Dyson, A.E. (Editors). (1969a, 1969b, 1970). *Black papers, I, II, III*. London: The Critical Quarterly Society.
- Douglas, J.W.B. (1964). *The home and the school*. London: MacGibbon and Kee.
- Fletcher, Ronald. (1987). The doubtful case of Cyril Burt. *Social Policy and Administration, 21*.
- Floud, J.E., Halsey, A.H., & Martin, J.M. (1956). *Educational opportunity*. London: Heinemann.
- Gillie, O. (1976, Oct. 24). Crucial data was faked by eminent psychologist. London: *Sunday Times*.
- Hearnshaw, L.S. (1979). *Cyril Burt psychologist*. London: Hodder & Stoughton.
- Joyson, R.B. (1989). *The Burt affair*. London & New York: Routledge.
- Kamin, L.J. (1977). Burt's IQ data. *Science, CXCIV*, 246.
- Vernon, P.E. (1980). Review of Hearnshaw (1979). *Bulletin of the British Psychological Society, 33*, 325.

Lewis B. Birch was Professor in the McGill Faculty of Education for 22 years. Before coming to McGill, he was editor fo the *British Journal of Educational Psychology* (1955-1962). Previously he served as associate editor. During those years, Professor Birch was a close friend of Sir Cyril Burt. In addition, Burt was a continuous advisor to Professor Birch on many professional matters.

Lewis B. Birch était professeur à la faculté des Sciences de l'éducation pendant plus de 22 ans. Avant son entrée à McGill, il était éditeur de la revue *British Journal of Educational Psychology* (1955-1962). Avant cette période, Professeur Birch assumait la position d'éditeur associé. Durant toutes ces années, Professeur Birch a été un proche ami de Sir Cyril Burt. En outre, Burt a été continuellement son conseiller dans divers domaines professionnels.

Chronique de test

Épreuve individuelle d'habileté mentale

Depuis quelques années déjà, les corporations de psychologues, de conseillers et conseillères d'orientation comme les spécialistes eux-mêmes demandaient la révision des échelles d'intelligence existantes. Nous avons alors décidé d'entreprendre une recherche scientifique qui puisse nous conduire à la création d'un nouveau test individuel d'intelligence. Il était important de conserver dans notre démarche une **continuité**; aussi notre Épreuve a-t-elle été normalisée dans le milieu francophone québécois. Elle se situe dans la lignée des échelles établies par de nombreux chercheurs francophones et anglophones. La structure et l'organisation des divers sous-tests ressemblent à la série des épreuves de Wechsler, de l'Épreuve individuelle d'intelligence générale de l'Institut de Psychologie de l'Université de Montréal et de l'Échelle d'intelligence Ottawa-Wechsler.

L'échantillon, structuré dans les diverses régions administratives, tient compte de la population scolaire totale des élèves de la quatrième année à l'université, du relevé de la population active et de la population inactive, des données obtenues de plusieurs ministères, dont le Ministère de l'Éducation du Québec, le Ministère de l'Enseignement supérieur, de la Science et de la Technologie, du Bureau de la statistique du Québec ainsi que du recensement de Statistique Canada.

Cette entreprise d'envergure de testing individuel, sur des sujets des deux sexes âgés de 10 à 24 ans, couvre une période de quatre ans à compter de la conception et de la planification de l'Épreuve. Nous avons effectué le travail, avec le concours de 150 personnes-ressources, coordonnateurs/trices et examinateurs/trices, dans 42 organismes et commissions scolaires et 92 écoles, sur un échantillon global de 1014 sujets répartis comme suit: 14 en pré-expérimentation; 200 pour l'analyse et le choix des items; 600 sujets de 10 à 15 ans et 200 sujets de 16 à 24 ans pour la normalisation. En outre, 38 élèves ont participé à une séance individuelle de retest.

Choix et analyse des items

Nous avons regroupé un choix d'items dans un ensemble de onze sous-tests qui mesurent une habileté cognitive répartie parmi les différents secteurs suivants: Connaissances, Jugement, Mémoire des chiffres, Arithmétique, Similitudes, Vocabulaire, Histoires en images, Images à compléter (ou Lacunes), Dessins avec blocs (ou cubes), Substitution, Assemblage d'objets.

Les critères utilisés sont en relation directe avec chacun des sous-tests. Ces items devaient donc constituer une mesure acceptable des habiletés mentales, verbales, non verbales et globales, et être assez souples pour subir certaines corrections qui leur assureraient fiabilité et validité. Ils devaient, hors de tout doute raisonnable, être libres de tout préjugé, de discrimination sexiste ou raciale, et pouvoir être facilement soumis à divers groupes d'âge, dans un environnement diversifié. Tous les items ont été revus en comité et inclus dans la version expérimentale, dite de l'analyse des items.

Après l'étape de la pré-expérimentation, nous avons entrepris cette analyse, une opération qui ne tenait aucun compte de l'évaluation de la personnalité de l'individu, mais qui avait pour but d'établir le niveau de difficulté de chacune des questions et de vérifier l'homogénéité des items.

Comme méthode d'échantillonnage, nous avons utilisé la méthode randomisée contrôlée. Nous avons retenu trois critères: l'âge chronologique, le sexe et la scolarité, pour la préparation de l'échantillon. Ce dernier comprenait, pour l'analyse des items, 200 sujets, 100 de chacun des deux sexes, dont les âges s'échelonnaient de 10 à 20 ans et plus; nous retrouvions ainsi les groupes 10 ans, 11 ans, 12 ans, 13 ans, 14 ans, 15 ans, 16-17 ans, 18-19 ans, 20 ans et plus.

Pour le critère scolarité, nous avons classé les sujets en trois groupes: a+: les sujets talentueux; b=: les sujets "ordinaires"; c-: les sujets médiocres. Cette façon de procéder valait pour tous les candidats.

Les sujets étaient francophones et baignaient dans notre milieu culturel depuis au moins 5 ans.

Population étudiée lors de la normalisation

Nous avons rassemblé les régions administratives inscrites au Répertoire des municipalités en cinq (5) regroupements de régions et avons administré l'Épreuve à 600 sujets de 10 à 15 ans. Le Tableau VIII illustre l'échan-

tillon de ces sujets dans chaque regroupement de régions du Québec. Il s'agit là d'un travail d'envergure qui a requis une coordination minutieuse et des contacts permanents avec les responsables partout dans la province. . .

Tableau VIII

Population de l'échantillon selon les objectifs visé et réalisé dans les regroupements de régions, pour les sujets des deux sexes, de 10 à 15 ans

Regroupements de régions	Objectif visé		Objectif réalisé	
	N	% arrondi	N	% arrondi
I Côte-Nord, Nouveau-Québec, Bas-Saint-Laurent, Saguenay — Lac-Saint-Jean	60	10	60	10
II Québec	96	16	96	16
III Montréal	342	57	329	54,8
IV Trois-Rivières, Estrie	66	11	84	14
V Outaouais, Abitibi-Témiscamingue	36	6	31	5,2
TOTAL	600	100	600	100,00

Pour le choix des sujets de 16 à 24 ans, nous avons utilisé une recherche systématique effectuée par Statistique Canada selon nos instructions et financée par l'Institut, en plus de nous servir de la Classification canadienne descriptive des professions (CCDP). Ainsi, notre population est très représentative de toute la population francophone du Québec et tient compte de la population aux études, au travail et inactive.

Études statistiques

Après avoir transformé les cotes brutes en cotes pondérées (étape importante et essentielle qui permet de comparer les résultats obtenus à différents sous-tests de nature dissemblable et souvent de longueur différente) et ce, pour l'échantillon des 600 sujets de 10 à 15 ans et celui des 200 sujets de 16 à 24 ans, nous avons calculé les moyennes, les écarts-types et les coefficients de variation.

Nous avons procédé de la même manière pour les quotients intellectuels, ce qui nous a donné, pour tous les âges, un QI moyen de 100 et un écart-type de 15.

Propriétés de l'Épreuve

1. Fidélité

Lorsque nous parlons de la fidélité ou de la constance interne d'un test, nous nous demandons si un sujet peut obtenir le même rendement au même test à des intervalles différents, soit 2 mois, 6 mois, un an. . . Si tel est le cas, on dira que le test donne des résultats auxquels un examinateur peut se fier comme étant valables, que les facteurs environnementaux ou personnels ont été vérifiés et que le sujet a donné sa pleine mesure. Cette caractéristique est importante; en effet, si un test n'est pas consistant, on ne peut le considérer comme valide.

a) La méthode du test-retest

Pour évaluer cette qualité d'un test, nous pouvons donc redonner le test à certains intervalles ou utiliser des méthodes statistiques. La première, dite de retest, peut être utilisée surtout pour mesurer les tests de vitesse, comme Mémoire des chiffres, Arithmétique, Substitution, Histoires en images, Dessins avec blocs, Assemblage. Nous avons fait le test-retest pour 38 sujets et la corrélation obtenue est élevée, indiquant ainsi un haut degré de fidélité ($r = 0,855$).

Nous avons comparé nos résultats à ceux de l'Épreuve individuelle d'intelligence générale de l'Institut de Psychologie de l'Université de Montréal et ils figurent dans le même ordre de grandeur.

b) La méthode de bissection

Cette dernière nous a donné les résultats illustrés au Tableaux XXXI.

Tableau XXXI

Comparaison, selon la méthode de bissection, entre les corrélations obtenues à l'épreuve individuelle d'intelligence générale (Barbeau-Pinard) et l'épreuve individuelle d'habileté mentale (Chevrier)

ÉPREUVE	N	ÂGE	r⁴
Barbeau et Pinard	35	12-18 ans	0,933 ± 0,014
Chevrier	600	10-15 ans	0,946 ± 0,004

c) La formule de Kuder-Richardson 20

Le Tableau XXXIII montre le haut degré de fidélité obtenu par cette formule.

Tableau XXXIII

Corrélations indiquant un haut degré de fidélité trouvées à l'aide de la formule KR-20

N	Âge	SOUS-TESTS	r	EP
600	10-15	Connaissances	0,885	± 0,006
		Jugement	0,985	± 0,001
		Similitudes	0,993	± 0,000
		Vocabulaire	0,987	± 0,001
		Images à compléter	0,769	± 0,011

2. Validité

Garrett suggère d'utiliser certaines méthodes indirectes en vue de déterminer la validité d'un test. L'une de ces méthodes consiste à calculer la corrélation de chacun des sous-tests avec l'ensemble du test. Nous avons établi ces corrélations et cette méthode a confirmé la validité de l'Épreuve, car les résultats obtenus sont suffisamment élevés et constants.

Par ailleurs, les niveaux intellectuels augmentent selon l'importance de la profession. Nous sommes donc à même de vérifier l'observation de Anne Roe et W.V.D. Bingham.

3. Quotients intellectuels et cotes Z

À l'aide de tableaux différents de cotes pondérées pour les deux groupes d'âges (de 10 à 15 ans et de 16 à 24 ans), nous avons créé les tableaux de quotients intellectuels.

En outre, afin de faciliter l'interprétation du rendement individuel de chaque sujet, nous avons conçu un profil et un psychogramme différent pour chaque groupe d'âges. Ce profil et psychogramme se complète avec les cotes Z pour chacun des sous-tests, ainsi que pour les échelles verbale, non verbale et globale et ce, pour chacun des groupes d'âges différents.

Enfin, nous avons dessiné une courbe des probabilités qui met en parallèle, avec le QI de l'Épreuve, d'autres échelles telles que le centile, le stanine, le sten, les cotes Z, T et z, l'écart-type (ou sigma).

Conclusion

Nous offrons cette Épreuve scientifique à ceux et celles qui oeuvrent dans les milieux scolaires, gouvernementaux, judiciaires, industriels, hospitaliers, universitaires, dans le monde du travail et dans les cabinets de psychologues et de conseillers/ères d'orientation.

Nous espérons qu'elle leur sera utile et qu'elle servira, en dernière analyse, à tous ceux et celles qui auront à la subir.

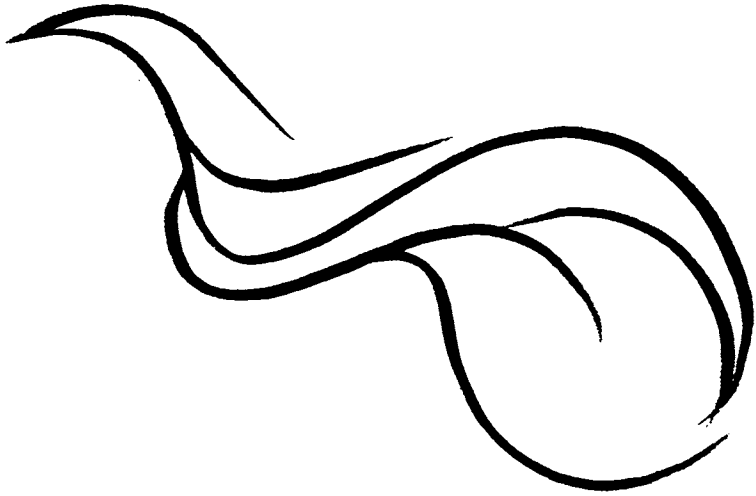
Enfin, nous n'oublions pas que tout travail est perfectible: nous comptons ajouter de nouvelles normes et pousser plus loin notre recherche en oeuvrant dans l'analyse des résultats pour créer des profils dans les domaines de la santé et du travail.

BIBLIOGRAPHIE

- Barbeau, G., Pinard, A. (1951-1963), *Épreuve individuelle d'intelligence générale*. Montréal: Le Centre de Psychologie et de Pédagogie. Réédité par l'Institut de Recherches psychologiques, inc.
- Chagnon, M. (1953). *Manuel et normes de l'Échelle d'intelligence Ottawa-Wechsler*. Ottawa: Éditions de l'Université d'Ottawa.
- Chevrier, J.-M. (1989). *Échelle d'intelligence Stanford-Binet: quatrième édition*. Montréal: Institut de Recherches psychologiques, inc.
- Lindquist, E.F. (1940). *Statistical analysis in educational research*. Cambridge: Houghton Mifflin.
- Sattler, J.M. (1988). *Assessment of children, third edition*. San Diego: Jerome M. Sattler Publisher.
- Simon, T. (1929). *Écoliers anormaux*. Montréal: Comité d'hygiène mentale de la Province de Québec.
- Simon, T. (1929). *La mesure du développement de l'intelligence par l'échelle B.S. avec démonstration*. Montréal: Comité d'hygiène mentale de la Province de Québec.
- Wechsler, D. (1944). *Measurement of adult intelligence*. Baltimore: The Williams and Wilkins Company.
- Wechsler, D. (1974). *Wechsler intelligence scale for children – revised*. New York: The Psychological Corporation.
- Weinberg, R.A. (1989). Intelligence and I.Q., Landmark issues and great debates. *American Psychologist*, 44(2). American Psychological Association.

Jean-Marc Chevrier a obtenu son Ph.D. en psychologie à l'Université de Montréal et effectué des stages post-doctoraux dans plusieurs universités américaines (Iowa, Californie à Berkeley, et Minnesota). Il est actuellement président-directeur général de l'Institut de Recherches psychologiques, Inc. et directeur général des Presses JMC Ltée à Montréal. M. Chevrier est l'auteur de nombreux articles scientifiques, manuels et textes scolaires dans le domaine de la psychologie; il a reçu de nombreux prix et distinctions dans les domaines de la psychologie et des arts graphiques.

Jean-Marc Chevrier received his Ph.D. in Psychology from the University of Montreal, and completed post-doctoral studies at several American universities (Iowa, California at Berkeley, and Minnesota). He is currently President and Director General of the Institut de Recherches psychologiques, Inc., and Director General of Les Presses JMC Ltée., in Montreal. Dr. Chevrier has authored numerous scientific articles, school manuals, and texts, in the field of psychology; he has received many awards in the fields of psychology and graphic arts.



Book Reviews

Carol Schneider Lidz, Editor.
DYNAMIC ASSESSMENT:
AN INTERACTIONAL APPROACH TO
EVALUATING LEARNING POTENTIAL.
New York: Guilford Press, 1987.
511 pp. \$53.50.

Approaches to the assessment of aptitude and learning potential have traditionally been based on the assumption that an individual's intelligence and learning capability were inborn and fixed. Historically, intelligence tests were erroneously thought to contain items that were culture-free and, thus, it was believed that they provided valid scores across cultures and population groups. Psychologists in the last two decades have examined the structure of intelligence tests and have concluded that most standardized psychometric instruments emphasize achievement factors, or "what one has learned." Furthermore, the tests tend to measure crystallized, rather than fluid, ability. In response to criticisms of this traditional testing paradigm psychologists and educators have moved to a new model, that of dynamic testing.

Dynamic Assessment has been available in Canada for only the last year. It contains articles by the most influential writers in the field, providing a comprehensive examination of the history, methodologies, and contributions of this new testing paradigm. The first article offers a historical overview of the early developments in the area, leading to Feuerstein's seminal work in the late 1970s. The socio-psychological roots of qualitative assessment are examined and Vygotsky's concept of the "zone of proximal development" is outlined. Contemporary applications of Vygotsky's theory are then analyzed in recent empirical studies. The focus of these three studies is Feuerstein's *Learning Potential Assessment Device* (LPAD).

Of particular interest is the article by Campione and Brown which links dynamic assessment with school achievement. The practical examples offered by the authors explain and clarify their work started over 10 years ago. Two articles by Budoff, one of the early pioneers of the test-train-test method are noteworthy in their overview of learning potential. They provide the link between early approaches to dynamic testing and the more contemporary Feuerstein approach.

The LPAD receives the most concentrated examination available in the current literature. Several articles delineate the potential use of Feuerstein's instruments in different situations: groups, primary-age children, and intellectually handicapped children. The article by Jensen and Feuerstein further illuminates the philosophy and structure of dynamic testing methods. It also explores differential applications of the LPAD.

The issue that has received the most criticism from traditional assessment devotees is the problem of transfer or generalizability (i.e., Does the mediation process described in the dynamic assessment procedure carry over to real-life situations and thus, give this "process" predictive validity?). The Campione and Brown and the Embertson and Budoff articles review this thorny issue. The conclusion of these authors is that the extent to which transfer occurs is to some extent determined by intellectual factors rather than the quality of the mediation process. This contentious issue will need future consideration in the literature as this book does not provide a definitive solution.

Dynamic testing models enable psychologists and educators to discover intervention strategies designed to improve learning and performance. The authors of the concluding chapter assess the state of affairs for intervention and suggest that future research will minimize any weakness inherent in this new model. Dynamic assessment seeks not a product but rather accessible characteristics for meaningful change. In other words, this approach emphasizes the **process** of testing. This shift in paradigm is becoming the state-of-the-art thinking in psychometry as we move toward the next century.

This book fills a void in the contemporary assessment literature. It is appropriate for psychologists, educators, and researchers. In one volume it examines the area of dynamic testing with primary source material. As testing models shift their focus from quantitative to qualitative aspects of measurement, researchers will need to explore the contents of this book as it is bound to be a milestone in the dynamic assessment literature.

John Lewis
McGill University

John Salvia & Charles Hughes.
CURRICULUM-BASED ASSESSMENT:
TESTING WHAT IS TAUGHT.
New York: MacMillan Publishing Company, 1990.
338 pp.

Traditionally, teachers have not been particularly concerned with the construction of their classroom tests. With an increasing emphasis on the accurate assessment of learning, though, teachers are developing the awareness that their tests may not be measuring what has been taught. They now know that an optimal match between what has been taught and what is being tested is necessary to ensure that course objectives are satisfied. Additionally, teachers now recognize that different test-item types (e.g., matching, true-false, essay) measure different levels of learning in terms of Bloom's taxonomy.

This cognizance has important implications for books in the area of criterion-referenced assessment. The contemporary trend is to focus specifically on the technical construction of test items with little or no attention paid to their relevance to the classroom curriculum. In addition to this, texts in the area of evaluation in teaching tend to lump criterion-referenced and norm-referenced testing into the same volume. Although combining the two concepts allows teachers a more extensive picture of educational assessment, the technical jargon used in norm-referenced assessment may frighten them away from any form of testing. Another disadvantage of the combination is that teachers very seldom use standardized assessment to fulfil curricular objectives, making its inclusion unnecessary.

Salvia and Hughes have attempted to meet current educational needs by making a solid effort to integrate the essential concepts and procedures of curriculum-based assessment into one volume. What makes this text stand out in a comparative sense is that they have done it by focusing solely on criterion-referenced testing without using technical jargon or a great deal of statistical facts. Each chapter of this "user friendly" text contains an advance organizer to assist the reader in focusing on particularly important concepts. So as not to encumber the introductory reader, but yet enable those interested to obtain additional information, all technical supplementary information is dealt with through the use of footnotes. Keeping these facts in mind it is easy to see how this book is designed specifically for teachers and education students.

The primary focus of this book is the measurement of classroom learning. In chapter one the importance of assessment in education is discussed, then the advantages and disadvantages of published achievement tests

and teacher-made tests are outlined. The last segment of this chapter contains an 8-step model for curriculum-based assessment. Subsequently, each of these points is expanded upon in the next four chapters. An explanation of how this model can be applied to each of the major areas of classroom instruction is then presented in the six succeeding chapters: reading, mathematics, written language, adaptive and social behaviour, learning strategies, and preschool education. Practical examples are furnished at the conclusion of each of these six chapters. In these examples Salvia and Hughes have attempted to include the various types of assessment decisions made in schools, the diverse ages of students assessed, the different levels of severity of a student's problem, and many educational settings. Finally, chapter twelve offers suggestions to the teacher for developing a more effective and efficient classroom.

From the perspective of someone who was previously a classroom teacher this book has long been awaited. The presentation of classroom assessment without technical jargon and elaborated statistics is a major feat. Salvia and Hughes are to be commended for their practical and thoughtful approach to curriculum-based assessment. This book will be a welcome addition to any classroom teacher's library.

Scarlett Alex
McGill University

Joseph M. Williams.

STYLE: TEN LESSONS IN CLARITY AND GRACE.

Third Edition.

Glenview, IL: Scott, Foresman and Co.

241 pp.

This 241-page book teaches how to write well in English. And it does this superbly. There are few people who would not be able, upon reading this book, to learn some additional skills for use in their preferred literary genre. The book purports to teach how to write clearly, cohesively, concisely, even elegantly and has the double merit of succeeding in this, on the one hand, and exhibiting these virtues itself, on the other. Aside from its purely pedagogic elements, the book is fun to read. It is splashed with colourful, historical, entertaining, and often amusing examples of good writing. And the quotations from renown writers are generally instructive, relevant, even memorable.

and teacher-made tests are outlined. The last segment of this chapter contains an 8-step model for curriculum-based assessment. Subsequently, each of these points is expanded upon in the next four chapters. An explanation of how this model can be applied to each of the major areas of classroom instruction is then presented in the six succeeding chapters: reading, mathematics, written language, adaptive and social behaviour, learning strategies, and preschool education. Practical examples are furnished at the conclusion of each of these six chapters. In these examples Salvia and Hughes have attempted to include the various types of assessment decisions made in schools, the diverse ages of students assessed, the different levels of severity of a student's problem, and many educational settings. Finally, chapter twelve offers suggestions to the teacher for developing a more effective and efficient classroom.

From the perspective of someone who was previously a classroom teacher this book has long been awaited. The presentation of classroom assessment without technical jargon and elaborated statistics is a major feat. Salvia and Hughes are to be commended for their practical and thoughtful approach to curriculum-based assessment. This book will be a welcome addition to any classroom teacher's library.

Scarlett Alex
McGill University

Joseph M. Williams.

STYLE: TEN LESSONS IN CLARITY AND GRACE.

Third Edition.

Glenview, IL: Scott, Foresman and Co.

241 pp.

This 241-page book teaches how to write well in English. And it does this superbly. There are few people who would not be able, upon reading this book, to learn some additional skills for use in their preferred literary genre. The book purports to teach how to write clearly, cohesively, concisely, even elegantly and has the double merit of succeeding in this, on the one hand, and exhibiting these virtues itself, on the other. Aside from its purely pedagogic elements, the book is fun to read. It is splashed with colourful, historical, entertaining, and often amusing examples of good writing. And the quotations from renown writers are generally instructive, relevant, even memorable.

The lessons of this book are well designed. The author introduces a topic, concisely delineates its features, states sources of difficulty, gives poorly written examples (then rewritten examples) bearing on the topic being explicated, makes summary statements of principle, and then presents exercises. As an example of how he proceeds, let us take Lesson Five, "The Grammar of Concision." Following a statement of the problem of wordiness, the author delves into its sources: redundant pairs, redundant modifiers, redundant categories, meaningless modifiers, pompous diction, belabouring the obvious, excessive detail, a phrase for a word. In treating these topics, he liberally admixes examples and exercises.

This section is followed by another entitled, "Talking to the Reader: Metadiscourse." Following a brief treatment of the nature of metadiscourse, that is, text which orients and contextualizes the ideas that the writer is communicating, the author presents types and examples along with advice on judicious use of this kind of discourse. The types focused upon are (1) hedges and emphatics, (2) sequencers and topicalizers, and (3) attributors and narrators. As in his other statements of principle, there is a lack of dogmatism in regard to how much and what kind of metadiscourse is appropriate.

The last section on concision treats the problem of the excessive use of indirect negatives. An example of indirect negatives that bloat as well as obscure text is the following (which I take from his summary on cutting fat). "There is no reason not to believe that engineering malfunctions in nuclear energy systems cannot always be anticipated." If we rephrase this: "One can assume that malfunctions in nuclear energy systems will take us by surprise," we benefit from increase in clarity as well as concision. Combining these negative constructions with excessively abstract, passive, complex terms produces the turgid, opaque text that fills academic journals. Even a cursory reading of this section cannot help but have a salubrious effect on offenders. (Or if you prefer . . . will certainly benefit offending readers.)

The lessons of this book run a gamut of useful topics which it will be useful for some to know in advance. They are the following, each of which the author treats in turn in separate lessons: clarity, cohesion, emphasis, concision, control of sprawl, management of long sentences, elegance, style and usage, and style and punctuation. Though each lesson can be read with profit without necessarily having read those that precede it, the author's specialized use of terms makes it desirable to learn such terms as they are introduced.

I would except from this counsel the preface, which, anomalously, has the most obscurantist prose I have read in a long time. Somewhere in the middle of the preface (I cannot tell you exactly where as the author, for some reason, has chosen not to number its pages) it is written

I know that many undergraduates have a problem precisely opposite to that which most of this book addresses – a style characterized by one fifteen-word sentence after another. But that’s not a problem that endures very long. I have worked with many adult writers in government, in the professions, in business; I have met not one whose major writing problem was a style that was immature. I am encouraged in this observation by every other adult writing program I have ever seen: Not one of them takes up the matter of writing longer rather than shorter sentences.

The context doesn’t help in interpreting this passage. I and others I have given it to read find it confusing. If one reads it several times carefully, one may actually become more confused. What, indeed, is the problem in question? Is it a style opposite to one fifteen-word sentence after another? Is the style characterized by the long sentences the one that the book addresses, or is it the opposite one? Perhaps it is the one the students have? Is it the long sentences that are opposite to what the book addresses? Do many students write sentences that are too long? If they do, do many others write sentences that are too short? Is this the problem that doesn’t endure very long? The author tells us that he has worked with writers in the professions, business, and government and not one has had an immature style. But is an immature style one that is excessively long, or one that is excessively terse and spare, or both, or something else? The paragraphs that follow did not help us.

Let me continue with some minor but not insignificant quibbles. Excellent as the lessons themselves are, they are not demonstrably flawless. Indeed, they would have been less enjoyable in the reading were that not the case (that is, *pace* Williams, not not flawless). First, there was the customary incantations about the use of Latinate abstraction, yet the author lauds the *Gettysburg Address* as the greatest speech in American history (p. 58) even though it is also recognizably the most Latinate English-language speech any of us ever had to memorize as school children. The implicit inconsistency in the above advice is also apparent in the abandon with which the author indulges in Latinate language. On page 14, every verb listed (in a discourse on nominalization) is of Latin origin: discover, move, resist, react, impede, fail, and refuse. Words have certain qualities that are more important than their etymology. Who would wish to change Macbeth’s “. . . multitudinous seas incarnadine” into “. . . many seas red”? There are qualities of meter and rhythm and overtone that override other considerations.

Similarly with the size (read number of syllables) of words. Comprehensibility is more important than size, as anyone who has perused a *Scrabble*

dictionary will quickly realize. The polysyllabic is not necessarily a correlate of the esoteric or even the pretentious. In any event, that the author considers "deem" and "apprise" to be "big words" (p. 86) and "think" and "inform" their respective simpler near-synonyms is an opinion like so many others that needs to be substantiated.

Lest I seem to be unduly undermining the credibility of a truly good book (that is now in its third edition), I will restrict my criticisms to one final and minor observation. The author rightly emphasizes the need to use verbs to express action rather than to express static conditions (unless of course one is speaking of static or immutable things). He calls changing verbs into subjects nominalizing; the results are nominalizations. Examples are regulate and regulation, establish and establishment, express and expression. But we cannot assume that if a noun and a verb are cognates, the noun is a nominalization of the verb. On the contrary, it may be that the verb is a verbalization of the noun. For example, formula is not a nominalization of "to formulate"; table is not a nominalization of "to tabulate" or to "tabularize" or even "to table"; context is not a nominalization of "to contextualize," and so forth.

I will close this exercise by noting that there are a number of qualities of good writing that are more important than either clarity or grace; one of them is accuracy. I may not need to say that even if one writes with the sparkle and pungency and elegance of Hemingway, it is of less importance than making the fine distinctions and nuances characteristic of scholarly writing. The cost of doing the latter is to make our prose heavier, more tedious, and less pleasurable than would otherwise be the case. *Style: Ten Lessons in Clarity and Grace* will help you minimize those costs.

Florent Dumont
McGill University.

Cynthia Solomon.
COMPUTER ENVIRONMENTS FOR CHILDREN.
Cambridge: MIT Press (1988).
192 pp. \$9.95.

Over the last decade, the area of educational computer applications has been rapidly evolving. While only a few years ago it was difficult to find any books pertaining to the topic of educational computing, today we find a number of top researchers in this ever growing field who have authored books ranging from practical applications of the technology in the classroom to more theoretical, research-oriented manuscripts.

dictionary will quickly realize. The polysyllabic is not necessarily a correlate of the esoteric or even the pretentious. In any event, that the author considers "deem" and "apprise" to be "big words" (p. 86) and "think" and "inform" their respective simpler near-synonyms is an opinion like so many others that needs to be substantiated.

Lest I seem to be unduly undermining the credibility of a truly good book (that is now in its third edition), I will restrict my criticisms to one final and minor observation. The author rightly emphasizes the need to use verbs to express action rather than to express static conditions (unless of course one is speaking of static or immutable things). He calls changing verbs into subjects nominalizing; the results are nominalizations. Examples are regulate and regulation, establish and establishment, express and expression. But we cannot assume that if a noun and a verb are cognates, the noun is a nominalization of the verb. On the contrary, it may be that the verb is a verbalization of the noun. For example, formula is not a nominalization of "to formulate"; table is not a nominalization of "to tabulate" or to "tabularize" or even "to table"; context is not a nominalization of "to contextualize," and so forth.

I will close this exercise by noting that there are a number of qualities of good writing that are more important than either clarity or grace; one of them is accuracy. I may not need to say that even if one writes with the sparkle and pungency and elegance of Hemingway, it is of less importance than making the fine distinctions and nuances characteristic of scholarly writing. The cost of doing the latter is to make our prose heavier, more tedious, and less pleasurable than would otherwise be the case. *Style: Ten Lessons in Clarity and Grace* will help you minimize those costs.

Florent Dumont
McGill University.

Cynthia Solomon.
COMPUTER ENVIRONMENTS FOR CHILDREN.
Cambridge: MIT Press (1988).
192 pp. \$9.95.

Over the last decade, the area of educational computer applications has been rapidly evolving. While only a few years ago it was difficult to find any books pertaining to the topic of educational computing, today we find a number of top researchers in this ever growing field who have authored books ranging from practical applications of the technology in the classroom to more theoretical, research-oriented manuscripts.

Computer Environments for Children by Cynthia Solomon (1988) offers a unique view of computers in education. From an historical perspective, Solomon presents an excellent overview of some of the pioneering work in educational computer research and applications. Solomon devotes a chapter to each of four prominent researchers in the field. The first of these is Patrick Suppes who was involved in the development of drill and practice programs in the 1960s. She then discusses the Plato system and the paradigm teaching strategy of Robert B. Davis. The next two chapters deal with programming. Solomon first discusses the work of Tom Dwyer and his use of BASIC programming to give students mastery over the computer. Last, but certainly not least, she devotes a chapter to a discussion of Seymour Papert's work on LOGO.

While the book does present this historical overview, that is not its primary focus. Rather, it serves as a framework for the discussion of these four researchers' theories of learning, in particular, their philosophy of elementary mathematics education and their vision of computers in education. If her idea is not original, her format is, and she presents some good critiques of their work.

Solomon's LOGO background is quite apparent, and while she does make some very good points about what she calls the "innovative research of the 60s and 70s," her criticisms are more valid for today's use of educational computing than at the time at which those research projects were carried out. She also fails to discuss the contribution of this research to current developments in software.

One criticism of the book is that it is somewhat biased towards LOGO. This is understandable since Solomon's own research was on LOGO and she was a student of Seymour Papert. However, it is unfortunate, because it is not reflective of the current state of educational computer use. LOGO failed to really achieve its goals. An influx of educational software and a shift away from the programming aspect of computers towards word processing and the use of the computer as a tool has left LOGO in limbo in so far as its classroom use.

Solomon does address some of the questions educators have concerning educational computing with suggestions that are more theoretical than practical. Her discussion of the "Computer Educator" is based on a series of what she calls models (reminiscent of Papert).

Unfortunately, it does not contain much information which would be considered of value to the classroom teacher. Sadly, one feels that her view of computers in education has been presented before. This is perhaps the biggest criticism of the book.

This is not to say that she has not made some very good points about educational computing. Solomon is a very good writer, and her extensive background in educational computing is evident. Some of her views on computers as an expressive medium and teacher reactions to learning about computers are right on target. Indeed, the last two chapters in which she discusses her vision of computers in education, based on the framework which she has provided, are probably the best.

Solomon has included an extensive reference list, particularly valuable for anyone interested in reading more about this popular area of research. Although the title of the book would suggest that it might be a practical manual for classroom use, the text is probably better suited as a good theoretical book providing an excellent historical framework of particular interest to researchers as well as educators.

Lorraine Coffin
McGill University

Lorri Neilsen.

LITERACY AND LIVING:

THE LITERATE LIVES OF THREE ADULTS.

Portsmouth, NH: Heinemann, 1989.

143 pp.

“Literacy, like an organism, only manifests itself in action” (p. 139). At a time when parents, communities, and governments are hotly debating the issue of literacy, Lorri Neilsen’s work provides a valuable perspective on the ways in which literacy grows and changes in the daily lives of three adults in her community of Hubbards, Nova Scotia.

Neilsen, a writer, researcher, and literacy consultant, provides us with detailed descriptions of the lives of her three participants and their community. Emerging as it does from her doctoral dissertation, this book is a research report; it is also a story of the village of Hubbards (population 750) on the Atlantic coast, thirty miles from Halifax.

As a participant observer (this is her community, too), she reveals the evolution of her own learning “to read, write, and see Hubbards in the ways the adults who worked with me do” (p. 11). The impulse for the research grew as she witnessed shifts in theory from a product to a process orientation, as

This is not to say that she has not made some very good points about educational computing. Solomon is a very good writer, and her extensive background in educational computing is evident. Some of her views on computers as an expressive medium and teacher reactions to learning about computers are right on target. Indeed, the last two chapters in which she discusses her vision of computers in education, based on the framework which she has provided, are probably the best.

Solomon has included an extensive reference list, particularly valuable for anyone interested in reading more about this popular area of research. Although the title of the book would suggest that it might be a practical manual for classroom use, the text is probably better suited as a good theoretical book providing an excellent historical framework of particular interest to researchers as well as educators.

Lorraine Coffin
McGill University

Lorri Neilsen.

**LITERACY AND LIVING:
THE LITERATE LIVES OF THREE ADULTS.**

Portsmouth, NH: Heinemann, 1989.

143 pp.

"Literacy, like an organism, only manifests itself in action" (p. 139). At a time when parents, communities, and governments are hotly debating the issue of literacy, Lorri Neilsen's work provides a valuable perspective on the ways in which literacy grows and changes in the daily lives of three adults in her community of Hubbards, Nova Scotia.

Neilsen, a writer, researcher, and literacy consultant, provides us with detailed descriptions of the lives of her three participants and their community. Emerging as it does from her doctoral dissertation, this book is a research report; it is also a story of the village of Hubbards (population 750) on the Atlantic coast, thirty miles from Halifax.

As a participant observer (this is her community, too), she reveals the evolution of her own learning "to read, write, and see Hubbards in the ways the adults who worked with me do" (p. 11). The impulse for the research grew as she witnessed shifts in theory from a product to a process orientation, as

she worked with students in school settings and adults in industry, and as she read, wrote, discussed, and observed during her graduate studies. Transformations in Neilsen's thinking about reading and writing motivated her to study literacy "not as a means for describing what people read and write, but as a means of seeing ways they behave in the world" (p. 7). Metamorphosis is a central theme of *Literacy and living*.

Neilsen's views on literacy and learning are, as she says, most influenced by the fields of semiotics and adult development. Thus her observations about Judy, Jim, Elizabeth, and the village of Hubbards are formed in part by her beliefs that human growth and development are lifelong processes and that "as human beings with intention, we receive and create signs that embody meaning from the world around us" (p. 7). Toward the end of the book, Neilsen proposes that her three participants primarily use linguistic, social, mathematical, and spatial signs to negotiate their lives. These are also the signs by which Neilsen creates her reading of Judy, Jim, and Elizabeth in Hubbards. She describes Hubbards in terms of its speech and conversations, its physical layout, and its institutions. Judy, Jim, and Elizabeth are described through Neilsen's conversations with them; the books they read; the lists, diaries, letters, and reports they write details about their family structures; and their social interactions at work, at home, and in the community enrich these stories.

Neilsen's engaging style draws us into Hubbards and the ordinary lives of her three subjects. On a morning run with Neilsen through the village, we learn the names of some of its inhabitants and the daily routines that shape their lives. We see their houses from a distance as we run past the school, over the bridge, and around the cove until finally, we stop to pick up the mail at the post office. Neilsen takes us into the offices and the homes of her participants – "A home reveals priorities, expresses who we are" (p. 97). And thus we become participants and observers in the community.

The first case study is of Judy, a mother of two children, a substitute teacher, and a community volunteer who "lives the traditional vows espoused by the community: the good daughter, mother, and wife" (p. 53). Having lived in Hubbards all her life, Judy "reads the lives of people here like a book, and as an active member of the community, helps to write their lives as well" (p. 51).

Jim, formerly the controller for a software company, is now the vice-president in charge of finance of a small gold exploration company near Halifax. He has three sons and is also active in community affairs. Neilsen

takes us to Jim's home and to his office where we learn how he deals with reading and writing on the job.

Elizabeth is the general manager of the Halifax branch of a data processing systems company. We visit Elizabeth in her office and discover how she reads and replies to her electronic mail. Conversations with Elizabeth reveal the evolution of her self-concept and her management philosophy. She spends fourteen hours a day at work in the city, and unlike Judy and Jim, is not deeply involved in community life. Elizabeth actively pursues success through reading, reflection, and interaction and "extends the terms reading and writing to include understanding and creating signs in human behaviour, such as speech and body language" (p. 118).

For Judy, Jim, and Elizabeth writing and reading are an integral part of their lives, but as Neilsen's research indicates "each is a literacy of one" (p. 123). Although each has a unique literacy, Neilsen states that they share a common process for becoming literate in context. The process begins by learning to read the signs in their environments and then evolves as they take an active part in shaping meaning in those contexts. In each of the case studies, Neilsen uncovers a "Transformative Experience," a literacy event that changes her participants' views of reality. Judy's literary experiences influenced her decision to become a teacher. Jim's career change was aided by the autobiography he wrote for a career counsellor. Through a re-reading of her adolescent diaries and the discovery of Jungian psychology, Elizabeth found herself and a way out of a life without purpose.

Neilsen's conclusions about the ways literacy shapes lives ground her recommendation for literacy education: children need complex knowledge of many contexts to know and shape their own worlds. Neilsen's final chapter explains the implications of her study for education. Supporting John Dewey's claim that the "ideal aim of education is the power of self-control," she suggests that this aim can be met by 1) making the child, not the subject, the centre of the curriculum, 2) recognizing that schooling is only a part of literacy and must be connected to literacies outside school, and 3) helping the child to understand self in society through literacy. She urges educators to think of reading and writing "not as forms of knowledge, but as means of forming knowledge" (p. 140).

As an ethnographic study, Neilsen's book is a rich and detailed description of three adults' literacy. One might quibble that these participants were not representative of the community; they all had university degrees and were engaged in the very literate activities of teaching and high-level man-

agement. What are the literacies of individuals who have no university degrees, and who are managed? Neilsen frequently includes her own life experiences in the case histories of her participants. Although these experiences seem to interrupt, they clearly remind us of Neilsen's role not only as observer, but participant as well.

Literacy and Living is a book for all those interested in the teaching of reading and writing. For researchers, it is a provocative invitation to a way of coming to know. Finally, it is a delightful inside look at the people and ways of Hubbards, Nova Scotia.

Jane Ledwell-Brown
McGill University

Review Board/Comité consultatif

Mark Abley
Pointe-Claire, Quebec
Glen S. Alkenhead
University of Saskatchewan
Anthony Adams
Cambridge University
Isa Aron
Hebrew Union College, Los Angeles
Rodney Baln
University of Western Ontario
Robert Baker
North Shore Community College
(Lynn, MA)
John Baugh
University of Texas at Austin
Garth Boomer
Education Dept. of South Australia
André Brassard
Université de Montréal
N. Peggy Burke
University of Iowa
Pierre Calvé
University of Ottawa
Richard Coe
Simon Fraser University
Donald G. Doehring
McGill University
James Draper
Ontario Institute
for Studies in Education
Eleanor Duckworth
Harvard University
Glen Eskedal
Suffolk University
Avigdor Farine
Université de Montréal
Mona Farrell
Concordia University
Françoys Gagné
Université du Québec à Montréal
Léonard Goguen
Université du Moncton
Gerald Grace
University of Durham (England)
Janet Hansche
Tulane University
Nancy E. Hicks
University of Regina
Paul Hirst
Cambridge University
Ken Hoeltzel
State University of New York
at Plattsburgh
Wayne K. Hoy
Rutgers University

Roselmina Indrisano
Boston University
Ishu Ishiyama
University of British Columbia
J.D. Jefferis
Bishop's University (Retired)
Henning Johannson
Luled University (Sweden)
Vivian F. Kinsey-Talley
St. Charles Parish Schools, LA
Edward W. Keyserlingk
McGill University
Jacques Lamontagne
Université de Montréal
Winfred Lehmann
University of Texas at Austin
Donald M. Little
Acadia University
Barry Lucas
University of Saskatchewan
Nell McDonald
University of Manitoba
Charles P. McFadden
University of New Brunswick
Robert Mulcahy
University of Alberta
Hugh Munby
Queen's University
Allan Neilsen
Mount St. Vincent University
Carole Oglesby
Temple University
Richard Pearson
Syracuse University
Lydia Raubenheimer
National Institute for
Personnel Research, RSA
Douglas Ray
University of Western Ontario
Joseph Reimer
Brandeis University
Bruce Roald
Dalhousie University
Douglas A. Roberts
University of Calgary
D.H. Saklofske
University of Saskatchewan
Ronald Smith
Concordia University
Claude Trotter
Université Laval
Karen Watson-Gegeo
University of Hawaii
Norman Watts
Acadia University

Le McGill Journal of Education paraît trois fois par an, en hiver, au printemps et à l'automne.

Tarif d'abonnement, post payé: 18\$ pour une année.

Prix du numéro: 6.50\$. Les numéros specials: 10.00\$

L'abonnement est payable au *McGill Journal of Education* et toute demande doit être adressée au 3700, rue McTavish, Montréal (Québec) Canada H3A 1Y2. Téléphone: (514) 398-4246.

Le *MJE* est indexé par le *Canadian Education Index* et par le Centre d'information sur les ressources pédagogiques (CIRP). Il figure dans l'*International Periodicals Directory* d'Ulrich, il est résumé dans *Sociology of Education Abstracts* et dans *Canadian Social Science Abstracts* et il existe en microfilms chez University Microfilms, Inc., à Ann Arbor dans le Michigan. Les anciens numéros sont disponibles sous microforme auprès de Micromedia Ltd., 158 rue Pearl, Toronto (Ontario) Canada M5H 1L3. Numéro de série international: CN

ISSN 0024-9033.

Intending authors should be aware that the *Journal's* readership extends well beyond the ranks of academics and professionals in education to which they are likely to belong. While the content of an article is expected to stand up under the scrutiny of specialists in its field—during the process of review before its acceptance for publication—its language and tone should not follow the conventions of scholarly writing in such a way as to shut out the non-specialist. (Unless they are of exceptional clarity, the *Journal* does not normally accept articles intended for specialist audiences alone.) The terminology employed should be accessible to intelligent lay readers; and notes, tables, and other apparatus of the formal paper should be omitted or else kept to the minimum necessary for the purposes of an interested reader. In short, the *Journal* asks for the lucidity that is a hallmark of the expert.

The *Journal* accepts English or French articles in the form of essays, interviews, descriptive reports of research, and critical reviews of books. Lighter pieces, humorous material, poetry and graphics of quality are also welcome. All written material should be furnished (following the style outlined in the American Psychological Association Publication Manual, 3rd edition) double spaced, in quadruplicate. Each copy should have a separate title page containing the author's name, which should not appear on the manuscript itself. A desirable length of article is between 10 and 15 pages. Articles (except book reviews) must be accompanied by a 50-75 word abstract (if possible, in the other language), and a short biographical note about the author. It is assumed that articles submitted for consideration have not been simultaneously submitted to any other publication. Authors whose articles are accepted for publication will be requested to assign all rights to copyright to the *Journal* by means of a standard form.

All submissions should be addressed to the Editor, McGill Journal of Education, 3700 McTavish Street, Montreal, Quebec, Canada H3A 1Y2.

Les futurs auteurs doivent comprendre que les lecteurs du *Journal* dépassent largement le groupe de spécialistes et de professionnels de l'enseignement dont font partie les auteurs. Bien que le contenu d'un article doive être en mesure de supporter l'examen des spécialistes dans le domaine – au cours de l'étap de révision précédent l'acceptation du texte destiné à la publication – la langue et le ton ne devraient pas suivre les règles de la rédaction savante au point d'être hermétique aux non spécialistes. (A moins qu'ils ne soient d'une limpidité exceptionnelle, le *Journal* n'accepte normalement pas d'articles destinés à un auditoire composé uniquement de spécialistes.) La terminologie utilisée doit être accessible aux lecteurs intelligents mais profanes; les notes, tableaux et autres éléments en retrait du texte doivent être omis ou du moins réduits au minimum pour répondre à l'intérêt du lecteur. En d'autres termes, le *Journal* demande aux auteurs de faire preuve de la lucidité qui est un trait propre aux spécialistes.

Le *Journal* recherche des articles rédigés en français ou en anglais, sous forme d'exposés, d'entrevues, de rapports descriptifs de recherche, ou d'études critiques d'ouvrages. Il accueille aussi des articles de style plus léger, des éléments humoristiques, des poèmes et des illustrations graphiques de qualité. Tout article doit être soumis (selon les indications de style données par le manuel des publications de la American Psychological Association, 3^e édition), à double interligne, en quatre copies. Chaque copie doit comprendre une page de garde portant le titre de l'article et le nom de l'auteur, mais celui-ci ne doit pas apparaître sur le manuscrit même. La longueur ordinaire d'un article est de 10 à 15 pages. Les articles (sauf les comptes rendus de livres) doivent être accompagnés d'un résumé de 50 à 75 mots (si possible dans une autre langue) et d'une courte biographie de l'auteur. Les articles soumis sont censés ne pas avoir été présentés simultanément à une autre publication. Les auteurs dont on accepte de publier les articles seront tenus de céder tous les droits d'auteur au *Journal* au moyen d'un formulaire standard.

Tout manuscrit doit être adressé comme suit: Le Rédacteur en chef du McGill Journal of Education, 3700 rue McTavish, Montréal (Québec), Canada H3A 1Y2.

Table of Contents

Editorial A Special Issue on School-Based Assessment	WMT	2
Guest Editorial International Perspectives on School-Based Assessment	Philip A. Saigh	3
School-Based Assessment Research in Canada	D.H. Saklofske H.L. Janzen	5
School-Based Assessment Research in the People's Republic of China	Joseph C. LaVoie	25
School-Based Assessment Research in Israel: Current and future directions	Moshe Zeidner	37
School-Based Assessment Research in Lebanon	Philip A. Saigh	65
Psychoeducational Assessment Research in the United States: From sociocultural context to extrapolations into the future	Jack A. Cummings Marc C. Laquerre	81
International Research on School-Based Assessment	Robert L. Thorndike	101
School-Based Assessment Research in Five Nations	Jerome M. Sattler Philip A. Saigh	109
Commentary on the Burt Affair	L.B. Birch	113
Chronique de test Épreuve individuelle d'habileté mentale	Jean-Marc Chevrier	121
Book Reviews		129
Graphics	Clifford Papke	Passim

McGill Journal of Education
Faculty of Education
3700 McTavish St.
Montreal, P.Q., Canada
H3A 1Y2

