1986

4. Future Directions for Educational Achievement and Ability Testing

Nancy S. Cole

University of Pittsburgh

Follow this and additional works at: http://digitalcommons.unl.edu/burosfuturetesting

http://digitalcommons.unl.edu/burosfuturetesting/5

This Article is brought to you for free and open access by the Buros-Nebraska Series on Measurement and Testing at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in The Future of Testing by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
Future Directions for Educational Achievement and Ability Testing

Nancy S. Cole
University of Pittsburgh

Forecasting future directions is at best a risky business. Few can claim to see with confidence the shape of things to come. In addition, the issues related to the future directions of educational achievement and ability testing are broad and numerous. They range from such important, but difficult to forecast, areas as the national mood toward education and educational accountability, legislation and the political arena, and court rulings to theoretical and technological advances. However, instead of providing a survey of possible futures resulting from these diverse and numerous potential influences, this chapter focuses on two areas of overriding importance in considering these future directions. The two areas to be addressed are: (a) the evolving conceptual understanding of the nature of the achievement and ability constructs, and (b) the opportunities afforded by advances in computer technology.

The chapter begins with a brief review of the current status of such testing to establish a context for the consideration of these two central issues and their implications for the future. It then turns, first, to the conceptual directions likely to evolve from changes in the familiar ability and achievement constructs and, second, to some of the implications of computer technology for testing in the future.

CURRENT STATUS OF ACHIEVEMENT AND ABILITY TESTING

A Recent Period of Criticism

To examine the current status of educational testing, it is necessary to recognize that the period from the late 1960s through the 1970s was a time of intense criticism of standardized testing. During this period of increased social awareness, test score differences between groups were attributed quickly to bias in
tests, particularly when the major alternative explanation under debate was hereditary differences in intelligence between groups (Jensen, 1969). Access to education and appropriate employment for minority applicants was one major objective of the civil rights movement and often tests were viewed as tools which stood in the way of such access. The courts were asked to judge the appropriateness of tests and test use and tests became central in several court rulings. When the concern with citizen rights made its way into the market place and the consumer movement began in earnest, the result was calls for “truth in testing,” which was implemented in some states basically as a requirement for periodic public disclosure of the content of major tests.

As this time of criticism blossomed, testing specialists were in the forefront both criticizing and defending tests in the public and professional literature and seeking ways to improve them. Measurement specialists began to examine with a new seriousness the possibility of test bias. The professional literature contains hundreds of studies conducted during this period (e.g., see Cole, 1981, 1983) which address technical approaches and substantive issues of bias. It became common for manuals of educational tests to routinely address the issue of bias. More measurement professionals began to disassociate themselves from arguments that test score differences necessarily reflect hereditary differences and in so doing emphasized with new vigor the effects of educationally related experience on educational ability tests. In addition, many test developers began to adopt a more critical view of their tests in preparation for disclosure and other forms of public scrutiny—to view them as an interested public might—and in so doing discovered a number of small ways of improving test content.

This period of criticism perhaps ended about the time of the 1982 National Academy of Sciences’s Committee on the Ability Testing report (Wigdor & Garner, 1982), which provided a broad survey of the social context of testing and concluded with both criticisms and endorsements of ability test use. In addition to a number of specific recommendations, the Committee noted several broad limitations of tests:

Although a well-developed test can be a reasonable good predictor of the performance of people in the aggregate, it may be a poor predictor of the performance of any particular individual.

Ability tests do not measure many things that are important to performance in school and at work.

The relative immaturity of theories of cognition places significant limits on the explanation of abilities that can be derived from test results (p. 237).

The Committee described its general view as a “call for balance”:

By emphasizing the limitations of tests we mean to counteract the widespread tendency to look to ability tests as a panacea for deep-seated social ills; and by discussing testing in the context of social developments that far transcend it in
importance or effect, we hope to counter the equally prevalent tendency to use tests as a scapegoat for society’s ills (p. 236).

Renewed Demands for Tests

Long before the period of criticism had run its course, there were trends in counter directions. “Back to the basics” became a rallying cry, particularly for elementary education, both as the focus of federally financed compensatory programs and in response to a growing public concern about educational quality. The more test scores declined, the more demand there was for testing as a method of educational accountability. The move to evaluate programs encouraged the development of tests designed for such purposes, and there arose a special concern to have tests directed more specifically to the particular curriculum in question. Criterion referenced tests (CRTs) became popular basically as tests referenced to a set of specific educational objectives. The minimum competency testing movement gained force with several states adopting statewide tests prior to high school graduation; although among testing professionals the movement had its critics (e.g., Madaus & McDonagh, 1979) as well as its defenders (e.g., Lerner, 1981). Test use in the schools continued at a high level apparently little affected by the period of criticism noted above. Houts (1975) estimated that students received from six to twelve full batteries of achievement tests during their school years. Anderson (1982) noted that such a high level of testing continues and is only a part of the total school testing program of most school districts.

Current Practice

Although educational achievement and ability tests appear to be on the brink of major types of change, current practice is notably similar to testing in previous decades in terms of the types of items and the methods of administration. By far the largest amount of commercially prepared standardized educational testing is in a paper-and-pencil, multiple choice, group-administered form. School achievement batteries continue to cover the same general areas (reading, mathematics, language, social studies, science) and report norm-referenced scores. A major alteration from past decades is the addition of score reporting in terms of item clusters tied to particular educational objectives.

School group ability tests continue to focus on verbal and quantitative areas, featuring changes in name in an effort to discourage the hereditary interpretations often associated with terms such as intelligence and aptitude. Individual intelligence tests are being updated after being challenged in the courts for out-of-date and biased content but focus on the same features of language, reasoning, and general knowledge as before.

Admissions testing continues to feature verbal and quantitative skills in the Scholastic Aptitude Test (SAT) of the College Board, the Graduate Record
Examination (GRE), the Law School Admissions Test (LSAT), and the Graduate Management Admissions Test (GMAT). The American College Test (ACT) which is organized instead around high school subject (i.e., English, mathematics, social studies, and natural science) remains an exception as do several attempts to expand and redefine in directions of analytical reasoning (GRE, LSAT) and more realistic contexts (Medical College Admissions Test).

Two areas currently receiving renewed attention are writing and reasoning/problem-solving. Concern among the lay public with the writing skill of students has arisen anew. Professional study of writing and how to teach it is active. Development of writing assessment procedures has been spurred by work on the National Assessment of Educational Progress and reports from the results of NAEP assessments has further stimulated concern. An optional writing sample has been returned to the SAT; many states mention writing skills much more directly as educational goals with some requiring written products from students in statewide assessments.

Renewed concern with reasoning and problem solving has also become prevalent. College, graduate, and professional school faculty frequently name reasoning skill as an area of deficiency in students and are sympathetic to attempts to include such skills in admissions assessment. Cognitive psychologists have been centrally interested in the cognitive processes of reasoning and problem solving and their attention has furthered this area as well.

Thus, although technological advances and concerns about the nature of the constructs being measured are both increasing, current practice continues primarily as in the past. The economy and efficiency of multiple-choice group testing with machine scoreable answer sheets makes it quite resistant to change. To replace it, any new assessment procedure must not only be educationally useful but also be practical, economic, and efficient. The time may be near when new computer technology will make possible those practical, economic, and efficient replacement methods. Although the tendency may be to transfer old constructs to the new methodology, this period of pending change begs as well for a reexamination of the familiar constructs. The remainder of this paper is devoted to a discussion of future conceptual and technical directions and some of the aspects that need to be reexamined.

CONCEPTUAL DIRECTIONS—THE NATURE OF ACHIEVEMENT AND ABILITY

The Traditional Achievement Construct

Traditionally the notion of achievement has been linked closely to the school curriculum. Achievement is expected in areas in which there has been instruction and, commonly, the content of instruction is used to determine areas of achievement to be assessed. Also, the names of areas of study are used to name the
achievement tests. Teacher-made tests in schools are prototypical achievement tests and practically all standardized achievement tests justify their content on the basis of addressing common elements in a school curriculum.

Historically, achievement testing has been the concern primarily of measurement specialists with backgrounds in education, as opposed to psychology, and the definition of achievement has been closely linked, as noted, to educational goals rather than to psychological constructs. E. F. Lindquist and his colleagues and students, such as Robert Ebel, advanced one line of educational thought in which achievement of understanding of content areas and the application of that understanding to new contexts and to the solution of problems was the primary educational goal as well as the primary target of educational measurement (Ebel, 1974; Lindquist, 1951). Similar lines, still from the educational perspective, produced schemas such as Bloom’s (1956) taxonomy of educational objectives as a blueprint for both educational goals and educational tests.

When achievement assessment is tied so closely to educational goals, its nature is subject to change as society’s educational goals change. With the back-to-the-basics philosophy of the 1970s and with public attention primarily at the elementary school level, emphasis on shorter-term and often behavioral objectives in the classroom and in achievement testing became prominent. Criterion-referenced testing has been the popular term associated with such testing. During this period both the educational goals and the assessment became more specific and more directly tied to what was being taught in relatively short instructional periods. However, both the broader view of achievement with concern for understanding and application of knowledge and the more recent narrow one in terms of very specific instructional objectives have been viewed as extensions of educational goals (Haertel & Calfee, 1983) and have been largely unencumbered by elaborate theory.

The Traditional Ability Construct

In contrast to achievement, the notion of ability has a more complex history involving psychologists more than educators and closely tied to the concept of intelligence. The popular early notion of intelligence was as an innate, general cognitive ability. The development of the construct has centered around two primary issues: (1) the issue of intelligence as one general cognitive ability versus several general abilities versus many highly specific abilities, and (2) the issue of the extent to which intelligence is inherited.

Modern theories of intelligence began in the late-19th century with British writers who tended to view intelligence as predominantly unitary—a single or a very few general abilities.Later, theorists elaborated separate factors. Thurstone (1938) identified seven primary mental abilities: verbal comprehension, word fluency, number, reasoning, spatial visualization, perception speed, and memory. Guilford (1967), perhaps the most extreme of the multiple factor theorists,
identified 120 separate ability factors in his structure of the intellect by crossing five operations by six products by four contents. Horn (1968) and Cattell (1971) proposed subdividing the general construct into crystallized intelligence (information-based and developed through environmental influences) and fluid intelligence (abstract and general and thought to be innate).

From a more educational perspective came definitions such as Thorndike’s (1911) definition of intelligence as the ability to learn. Early stimulus-response (S-R) theorists saw intelligence as resulting from a build up of S-R bonds. From this trend Gagne (1970) proposed a hierarchical theory of eight kinds of learning from simple Pavlovian conditioning to rule learning and problem solving.

Many recent theories of intelligence have arisen from an information processing perspective. Sternberg’s (1977) componential theory of intelligence is one example. A component, in this conception, is an elementary information process that may be classified by its function and level of generality. Higher order processing similar to general factor notions of intelligence are referred to as metacomponents in Sternberg’s system. Other authors approaching the concept of intelligence from an information processing view include Hunt (1978), Pellegrino and Glaser (1979), and Snow (1979).

The issue of the extent to which intelligence is inherited began as well with early British theorists (Galton, 1883) who strongly supported the view that intelligence represented an innate ability. Relying heavily on the evidence from twin studies, the work of Burt (1940) and others supported a strong inherited component. More recently, some writers (e.g., Anastasi, 1976) have pointed to difficulties in twin studies and the many environmental influences on cognitive performance have received more emphasis.

Early notions of intelligence as largely perceptual and sensory (e.g., Galton, 1883) led to tests, measuring perceptual skills (e.g., Cattell, 1890). However, the more successful efforts to assess intelligence often arose in more practical than theoretical contexts. For example, Binet (Binet & Simon, 1905) tried to assess skills that would identify children who would have difficulty in a regular French school context. His efforts and other later assessments were often successful in the sense of identifying performances that related to other similar performances and predicted important subsequent school behaviors. Analyses of the results of most educationally related ability tests produced empirical evidence for general factors across types of performances consistent with many theories of intelligence. However, the evidence did not provide any resolution to the theoretical debates and, in fact, there was often little connection between the theoretical concerns and the development of intelligence and ability tests designed for practical use. An exception is the recent Kaufman Assessment Battery for Children (Kaufman & Kaufman, 1983) in which the theoretical notions of fluid and crystallized intelligence are applied.

Today we view the results of intelligence tests as showing what a person can now do based on learning and experience (i.e., a developed ability) without the
implication of permanence or heredity it once held (Anastasi, 1981). We know such scores relate to many significant events from educational to job performance. We know that what is effective, competent functioning differs in some ways from culture to culture and may differ qualitatively at different developmental levels. We know that some theorists still focus on a single, general concept of intelligence but many other consider instead more specific abilities. And we know that measures we now have, though quite efficient and predictive, only approximate the construct we conceptually describe.

**Distinctions Between Achievement and Ability**

These descriptions of achievement and ability, though suggesting the educational content relatedness of the former and the more psychologically theory-based character of the latter, still give only hints of the similarities and differences inherent in measures of the two concepts. Yet there have been reasons to distinguish ability (intelligence, aptitude) from achievement for as long as there have been standardized tests, and conceptually the distinctions go back much farther. The constructs have been distinguished historically in terms of the relationship of each to heredity (i.e., the measurement of innate capacity versus the measurement of learned accomplishment). Currently, however, most writers recognize now that this distinction is largely inappropriate because both types of measure reflect current attainments learned though a variety of experiences. Another distinction often made concerns the use to be made of the information (i.e., to predict future performance versus to assess past learning). Yet another distinction involves the nature of the skills measured (i.e., generalized skills usually involving reasoning versus specific knowledge of information in particular content areas). Finally, distinctions have been made about the length of time required to learn the skill and the site of the learning with ability tests measuring skills learned over longer periods of time from many types of learning experiences with achievement tests measuring skills learned in shorter time periods, primarily in school. Anastasi (1976) and Cronbach (1970) described aptitude and achievement tests as falling at different locations on a continuum with the continuum reflecting the nature of the skills, source of the learning, and time of the learning period.

Many (e.g., Cooley & Lohnes, 1976) have noted that one cannot rely on a label on current tests to distinguish the types of skills, because many current ability tests ask for specific factual knowledge or specific instructed skills whereas many current achievement tests include questions requiring general reasoning skills and application of knowledge to new domains. The distinctions have become increasingly fuzzy as cognitive psychologists have directed attention to knowledge structures in specific content fields such as mathematics, reading, and science as typified by Glaser’s (1981) examples of potential contributions of
cognitive psychology to psychometrics. Thus, although there are many conceptual distinctions that can be made between achievement and ability, there remains considerable confusion, especially in relation to current test content.

Messick (1984) suggested that we begin directly with the constructs not with the possibly contaminated content of current tests. What do we and should we mean by the concept of educational achievement and that of educational ability? According to Messick, (1984),

Educational achievement refers to what an individual knows and can do in a specified subject area. At issue is not merely the amount of knowledge accumulated but its organization or structure as a functional system for productive thinking, problem solving, and creative invention in the subject area as well as for further learning (p. 1).

Messick goes on to develop the idea of the organization or structure of knowledge as follows:

A person’s structure of knowledge in a subject area includes not only declarative knowledge about substance (or information about what), but also procedural knowledge about methods (or information about how), and strategic knowledge about alternatives for goal-setting and planning (or information about which, when, and possibly why).

Achievement is thus viewed as a knowledge structure and contrasts with ability which may be conceptualized as a process structure, a relatively stable constellation of psychological processes (Messick, 1982b) developed over time through learning into “a coherent set of habit skills, knowledge, conceptual developments, and tactical and strategic ‘know how’” (Cattell, 1971, p. 319) or, in information processing terms, as assembly and control processes (Snow, 1980), functioning much like subroutines or prior assemblies in computer terms.

Not all writers agree with this dual formulation of achievement as knowledge structure and ability as process structure. Ebel (1969, 1974, 1982) described the knowledge structure similarly to Messick’s definition, but did not distinguish the ability or process structures from the knowledge structure. In fact, Ebel seemed to argue that we have tended to call too many of the higher level achievements abilities when they are, in fact, a crucial part of achievement. He was especially concerned that we not leave to the achievement construct only lower level factual acquisition. Anastasi (1976, 1980, 1981) treated achievement within the notion of developed abilities and described achievement and ability within a continuum from recent, more factual, school-based performance (achievement) to longer term, high level, more generalized performance (ability).

In spite of the complexities in distinguishing abilities and achievements in practice, at this point it seems useful to accept the viability of aspects of both
constructs with several reservations: (a) that we not relegate achievement to acquisition of facts and leave the complex accomplishments to the ability domain; (b) that we retain the learned nature of each; and (c) that we remain willing to mix the two constructs in our measurement efforts because we cannot measure achievement without ability being reflected nor ability without achievement in one or many domains.

Ability and achievement can be conceptualized as a matrix, with the knowledge structure in a content area (i.e., achievement) crossed with more general process structures (i.e., abilities). Such a matrix structure illustrates the conceptual distinction but also their interrelated nature when measured. Measurement occurs then as tasks are presented involving one or more than one cell in the matrix. When we assess factual recall we are at an information level on the achievement dimension and perhaps the memory level on the ability dimension. For high levels of achievement a person is presented with tasks requiring the higher levels of the knowledge structure called achievement and using advanced processes called abilities. Any educational performance can then be viewed as a combination of exercising a portion of a knowledge structure (i.e., achievement) and certain portions of a process structure (i.e., ability).

Research is occurring in many content areas directed toward defining the knowledge structure in each area. Although Bloom’s (1956) taxonomy provides one type of general terminology that could be used to define the knowledge structure in a variety of subjects, most recent work seems quite specialized for the particular content domain and suggests a unique knowledge structure definition for each subject. For example, structures have been proposed for several subject areas (Glaser, 1981).

The ability structure is the focal interest of many current cognitive and information processing theorists. It may be defined by elemental to higher order information processes or components or in traditional terms such as memory, reasoning, etc. The ability dimension of the matrix is less content specific and one definition of the ability structure is likely to be useful in a number of cognitive content areas.

Even with such a cursory look at this conceptually complex area several features are clear:

1. Both achievement and ability as described here have tremendous educational significance and can even be characterized as the ultimate goals of education in terms of developing knowledge structures (achievement) and process structures (abilities) in students.
2. Both structures as described include high level cognitive activities and build upon lower level ones.
3. Both arise through learning often over a substantial period of time.
4. The two structures are intricately interrelated and each relates to the accomplishment of the other.
Implications of the Constructs for Education. A first implication in the consideration of these constructs for education is the importance of high level cognitive accomplishments in an adequate description of educational attainment. Although educators have long endorsed the importance of higher levels of learning such as reflected in Bloom’s taxonomy, there have sometimes been demands to teach and test at lower levels of information and skill acquisition. With requirements for evaluation, people have tried to measure those things that can be taught successfully and to demonstrate that success in a short time by focusing on precisely definable and immediately observable instructional objectives.

There are indications of renewed concern for higher levels of learning. For example, public attention is now directed to concerns with excellence at the secondary education level (National Commission on Excellence in Education, 1983; Boyer, 1983). With concern for excellence at the high school level, it seems likely that concern with higher cognitive structures will occur.

A desirable direction for education is to move toward greater concern with the instruction and the assessment of higher level educational goals whether we call them comprehension, analysis, and evaluation as Bloom did, or knowledge structures with terms specific to each of the content domains. With such higher level goals, it will be necessary to address process structures as well as knowledge structures. It would likely be possible to address the development of process structures more directly educationally with such a focus. At these higher levels there may be diagnostic educational purposes for which we will wish to distinguish achievement from ability in order to assist student development or to predict future behaviors in which more general process structure accomplishments will be useful.

Finally, we would profit educationally from a better understanding of ability and achievement constructs, how they are distinguished, and how they work together. As one examines Messick’s (1973, 1982, 1984) theoretical notions of knowledge structures and process structures, the interrelationship of the two in the learning process and their intermixing on many present day tests becomes easier to understand. However, the distinction is likely not an essential one for all time and all places. In education, particularly, as we better understand knowledge structures in various subject areas and study such things as word decoding, semantic access, sentence processing, and discourse analysis in reading or error patterns in mathematics and common misunderstandings of physical principles (Glaser, 1981), we may find achievement and abilities mixed together without loss in the mixture. The proper goal here is not clear and complete separation—likely an impossible goal—but better understanding of the interrelation and how tests reflect both. This understanding may help us concentrate more explicitly in the educational process on both the high level knowledge structures indicative of important achievement in a field and on the process structures that are developing through learning experiences as well.
4. FUTURE DIRECTIONS FOR EDUCATIONAL TESTING

TECHNOLOGICAL DIRECTIONS—USING COMPUTER TECHNOLOGY

In this section the focus is on one major technical development that holds the promise of a revolution in educational testing—computerized testing. Of course, the notion of computerized testing is not new. What makes it potentially revolutionary today is that it has become practically feasible with small computers linked to TV screens in essentially every school in the nation.

In this section, three major areas in which computerized testing may revolutionize testing are addressed. The revolution seems more certain to occur in the first of the three areas and seems desirable, if not certain, in the latter two. The three areas are:

1. Use of computers to handle clerical functions of administering tests (e.g., presenting stimuli, recording answers), scoring tests, and reporting results back immediately to students and teachers.
2. Use of computers for adaptive testing so that each individual is tested with different questions depending on previous responses.
3. Use of computers to develop entirely new types of questions involving more complex stimuli and responses which affect additional stimuli.

Computers As Clerks

Clerical functions have always been one of the strengths of the computer. Today, with computers in nearly every classroom, it seems virtually certain that clerical functions of managing instruction, keeping track of student progress, presenting and scoring practice and drill, and presenting and scoring tests will become commonplace within 10 to 20 years. Of course, there are complexities in accomplishing this goal (e.g., compatibility of different machines in the same school, sales and ownership of computer-based instruction and testing) but the complexities seem small compared to the opportunities for more efficient classroom operation in a form likely to be accepted by schools.

Adaptive Testing Via Computer

Throughout its history, standardized testing has been locked into fixed tests in which all individuals took all items regardless of their performance. Because one gets more information about an individual level of performance by testing near that person’s performance level it was necessary for tests to have a wide spread of item difficulty when used with groups with broad spreads of performance levels. Hence, the better performers had to answer many questions too easy to provide much information on them and the poorer performers had to answer
many questions too hard to help much in locating their performance level. Thus, in a 100 item test, maybe only 30 to 50 items were giving much information with respect to each individual.

Two things were needed to give each test taker the right 30 to 50 items: (a) a statistical procedure to guide the selection of which items to give each individual and to produce a final score on a scale comparable to the scores others taking different items would receive, and (b) hardware to record the individual’s response to items, perform the analyses for the next item choice, and present the next item to the test taker. Item response theory has provided a needed statistical procedure and today’s minicomputers and microcomputers provide the hardware.

Adaptive testing provides the opportunity not only to assess more efficiently the global achievement and ability accomplishments we now assess but to probe adaptively into the misconceptions underlying errors made to previous questions. The opportunities of fitting in such information directly into instruction are exciting indeed.

New Item Types by Computer

In this area we have only begun to tap the surface of the possibilities for developing entirely new item types. However, some possibilities can already be envisioned in the types of complex stimuli that are being presented in the arcade and home video games and in the possibilities for studying problem solving strategies or perceptual processes in learning.

A VIEW OF THE FUTURE

One of the rewards of preparing a paper on future directions is the license it gives for sharing one’s own images of a future time. I am taking that license here to describe to you what I think educational achievement and ability testing might be like by about the beginning of the 21st century. What I do here, however, is to combine those events which I feel confident will occur with some about which I have considerably less confidence. Although to some this may be an uncomfortable vision, I view it positively with many opportunities for better education.

We begin this fantasy by entering an elementary school classroom. What strikes us first is that on every child’s desk is a small computer and screen and at several locations in the room are large screens visible to the whole class. The teacher is talking to a group of 10 youngsters about social studies and uses the computer at hand to call up a short film strip on one of the large screens to illustrate a point. As the lesson ends, each child activates his or her own computer and is given questions over the material covered that day. When a mistake is made, a hint is given and the child tries again and each student may receive a
different set of questions depending on earlier answers. The student also types in short answers to some questions. While this group of students moves on to lessons on science or art, the teacher calls up a display of the results of the children’s performance on the questions asked, which includes an error analyses identifying particular children with incorrect conceptions or misunderstandings of the content. The results are summarized by various levels of knowledge from acquisition of facts to higher levels of learning with categories appropriate to the content area—the type of learning we used to call comprehension and application. Today’s lesson was near the end of a unit and had more questions in the comprehension and application categories although several weeks ago the emphasis had been more on facts. For students who perform poorly on these higher levels, several special activities were suggested. The teacher notes which topics or activities to review the next day. Tomorrow the students will first review today’s material, then hear from the teacher, then answer some questions over the entire unit.

While this lesson was going on, three students sat with headphones on, operating computers at their desks. These students were drilling on multiplication facts and since each differed in the way they learned these facts from other children (i.e., perviously they were called learning disabled), they were receiving auditory as well as visual stimuli as well as hints when they made mistakes and words of encouragement for correct answers. These three students had been identified as having difficulty learning the math facts in the usual way and had been given a special diagnostic procedure to identify process structure difficulties. Problems in visual memory had been identified and these students were receiving training to use their more effective auditory memory systems to compensate for the visual memory problems. After 10 minutes of this drill, the teacher called up the results for the three students separately. The computer report indicated how many facts they attempted, answered correctly, and the prominent errors.

That evening at home, each child would contact the school computer to get a report on that day’s activities for the parents. These reports indicated each student’s progress and areas needing special work and described the child’s own individual homework assignment based on the day’s performance. The parents could receive a special message from the teacher or leave one for the teacher for the next day.

Meanwhile the principal, with several classes at this grade, wishes to check on the children’s progress in science in each classroom and calls up a summary report by classroom. Several times during the year the principal must report to the superintendent on the students’ progress in all major content areas for sharing with the school board. Those reports will be due in 2 weeks and this check is to insure the students are on track. That report to the school board will include norm-referenced information on the students’ standing in relation to students nationally but will require no special testing. Instead it uses particular questions
from the student’s regular daily and end-of-unit review questions administered by the computer. This report is organized as well by categories within the knowledge structure showing both information acquisition and higher levels as well. In addition, however, the test report from the computer notes that certain higher levels are not covered well by the computerized procedures and refers the principal to projects and reports evaluated by the teacher as a supplement to the computerized segment.

A mile away in the high school building, two juniors were preparing to begin their college admission tests. The school principal connected one to Princeton for the College Boards and one to Iowa City for the ACT using a voice identification procedure to identify the caller and begin the process. Both students took adaptive tests in the major areas of high school achievement required for college as defined slightly differently by two major commissions. Both would have been called achievement tests in the 1890s although there was a mix of process structures in the content areas of concern. Because of their adaptive nature, the basic survey tests required on 1½ hours to complete compared to 3 hours in years past. The remaining time was spent by the student preparing short written responses to questions requiring high levels of cognitive functioning in the content areas. Immediately, the survey scores would be transmitted along with the short answers to the colleges of the students’ choice.

CONCLUDING REMARKS

Fantasy aside, it is an exciting time for educational achievement and ability testing with tremendous opportunity through technology to become not an add-on process done exclusively for evaluation or accountability reasons but an integral part of the instructional process much as homework assignments and ditto sheets now are. Once computerized we can then select the relevant information about students’ performance for those evaluation and accountability purposes too, choosing the part of the information best suited to those uses. But even those functions become an ongoing monitoring process that can support the instructional enterprise, not a separate once a year event.

Even with these advances all our problems will not, however, be solved. It seems likely that classrooms of the future will involve primarily the familiar multiple choice format or open-ended versions which produce machine scorable answers even though we would wish to see many new item types as well. There will continue to be difficulty assessing many important educational goals. Perhaps the ultimate advantage of the technology is that, if correctly programmed, it can remind us of those areas not measured well in the system and requiring special teacher instructional attention and evaluation through other means.

In conclusion, it should be noted what the successful implementation of such a system will involve. As an integral part of the instructional system, its pro-
gramming will require as much educational theory as psychometric theory. We will have to spend as much or more time learning how to ask questions to help children learn as in assessing their learning. The largest focus of our attention will be on achievement testing even though we may explicitly include under that label those ability processes involved in learning and accomplishing content domains. Educational testing with a predominant ability focus will occur on a smaller scale in a diagnostic mode and will result in educational prescriptions. Teachers will study testing, not as a separate subject, but as part of instructional methods. Teachers will be trained to spend their time considerably less on classroom management and drill and considerably more on activities related to higher levels of learning.

Let’s hope we’re up to the tasks ahead.

REFERENCES


