TECHNICAL REPORT

Assessing Mathematical Knowledge of Adult Learners: Are We Looking at What Counts?

NCAL Technical Report TR98-05
(May 1998)

National Center on Adult Literacy
Graduate School of Education
University of Pennsylvania

NATIONAL CENTER ON ADULT LITERACY

The National Center on Adult Literacy (NCAL), part of the University of Pennsylvania's Graduate School of Education, was established in 1990 with a major grant from the U.S. Department of Education. Its mission is to enhance the quality of literacy work by pursuing three primary goals: (a) to improve understanding of adult learners and their learning, (b) to foster innovation and increase effectiveness in adult basic education and literacy work, and (c) to expand access to information and build capacity for adult literacy service provision. The Center is currently supported by federal, state, and local agencies as well as private foundations and corporations. NCAL is located, along with the UNESCO-cosponsored International Literacy Institute, in its own building on the Penn campus.

This document is based in part on work sponsored by the Education Research and Development Center Program (Grant No. R117Q00003) as administered by the Office of Educational Research and Improvement, U.S. Department of Education, in cooperation with the Departments of Labor and Health and Human Services. The findings and opinions expressed here do not necessarily reflect the position or policies of the National Center on Adult Literacy the Office of Educational Research and Improvement, or the U.S. Department of Education.

Assessing Mathematical Knowledge of Adult Learners: Are We Looking at What Counts?

Joy Cumming

Centre for Literacy Education Research
Griffith University, Australia

Iddo Gal

University of Haifa, Israel

Lynda Ginsburg

National Center on Adult Literacy
University of Pennsylvania

Author Biographies

Joy Cumming is a Senior Lecturer at the Faculty of Education, Griffith University, Brisbane, Queensland, Australia. Her work focuses on adult and mathematics education, and on the policy implications of educational assessment and evaluation.

Iddo Gal is a Lecturer at the Department of Human Services, University of Haifa. His research interests focus on adult numeracy and statistical literacy, individual and organizational aspects of empowerment, and acquisition of functional skills. Direct correspondence to: Dr. Iddo Gal, Dept. of Human Services, Eshkol Tower, University of Haifa, Haifa 31905, ISRAEL. (E-mail: iddo@research.haifa.ac.il)

Lynda Ginsburg is a Senior Researcher at the National Center on Adult Literacy at the University of Pennsylvania and directs the Center's participation in a number of projects. Her research focuses on the processes of adult learning and instruction particularly in the areas of technology and numeracy. She develops and teaches courses and workshops for educators on enriching instructional practice and has extensive experience teaching adult learners in a variety of settings.

Abstract

Adult students' numeracy skills are typically assessed at various points during instruction for a variety of purposes. These include student placement, informing instructional decisions, and measuring student learning. However, the assessment instruments available may not be adequate for providing interpretable and useful information to instructors, program directors, funders, or learners. The authors of this report advance a set of principles that reflect psychometric concerns and current research policies. These principles can be used to evaluate existing assessment practices and guide the development of new assessment models. Commonly used numeracy assessment tools are examined in light of these principles and are found wanting. The authors suggest alternative items and strategies. They further discuss emerging trends in numeracy assessment. Implications for policymakers and practitioners are also indicated.

TABLE OF CONTENTS

INTRODUCTION

The increased public attention paid in recent years around the world to adult literacy and numeracy issues has been based in part on economic rationalism—a more productive or competitive workforce requires people with higher skill levels—as well as on a growing acknowledgment that education (and literacy) is a basic right of all individuals. This attention has been reflected in considerable increases in funding by governments and other organizations for literacy and numeracy programs that aim to help adult learners achieve personal goals and prepare for the labor market.

While attention to literacy and numeracy has been growing, assessment of the literacy or numeracy skills that people (students) have, can apply, or need to further develop, has become a major challenge for those involved in adult education. This challenge is caused by the joint influence of several interrelated processes:

- Theoretical developments over the last two decades have increased our understanding of the multiplicity of contexts and pathways in which people's literacy and numeracy skills and dispositions can develop and be practiced. (Venezky, Wagner, & Ciliberti, 1990)
- New perspectives on the goals of mathematics education (e.g., NCTM, 1989) and workplace preparation in schools (SCANS, 1991) outline an expanded set of skills and dispositions as the targets of educational interventions designed to prepare students for real-world functioning.
- Advances in the theory and practice of educational assessment are resulting in the demand for more elaborate and authentic assessments that encompass the full range of skills, knowledge, and dispositions included in real-world literacy and numeracy.
- There is a growing emphasis on educational accountability and demands from funders that adult education programs provide credible and informative data about changes in students' skills.
- These processes present dilemmas for students, teachers, program administrators, and policymakers. In particular, teachers and programs face a problem if they want to address a wide range of issues in their instruction and expand their assessments within a limited time frame, while also collecting data about student performance that will satisfy the information needs of policymakers and funders.

WHAT IS NUMERACY?

The term *numeracy* describes an aggregate of skills, knowledge, beliefs, and habits of mind, as well as general communicative and problem-solving skills, that individuals need in order to effectively handle real-world situations or to interpret *mathematical* or *quantifiable* elements embedded in tasks. The diversity of life contexts in which learners may need to use numeracy skills implies that numeracy is relative and dynamic, rather than a fixed, static set of knowledge and skills. (In contrast, the traditional mathematics school curriculum attempts to impart to all learners a complete, ordered set of computational skills and factual knowledge.) One can have knowledge of a mathematical fact, concept, procedure, or strategy, but it is only in the appropriate selection, use, or adaptation of that mathematical knowledge *in purposeful contexts* that one demonstrates that one is numerate (Gal, 1994).

Adult numerate performance involves the confluence of many components, including domain-specific knowledge and strategies as well as general cognitive skills and world knowledge that may have been acquired inside and/or outside the classroom (Perkins & Salomon, 1989; Sticht & McDonald, 1993). These skills and knowledge often develop from common experiences and can form the foundations of mathematical reasoning skills although they may not always be identified by the adults as mathematical. For example, in a study of 160 people, including school children from ages 5 to 15 years and adult numeracy students, only two participants reported that they did not play any form of games and therefore could not explain rules of a game (such as scoring) or any strategies they might use to win in a game (Cumming, 1995). In addition, virtually all adults shop, plan events, and use and manage money. These types of everyday experiences, together with many others, provide contexts and constraints within which adults develop and use numeracy skills. Improving numerate performance requires developing additional skills and strategies that enhance understanding of the exigencies of significant contexts, promote flexible thinking, and widen the scope of participation in activities involving numbers.

PURPOSES OF NUMERACY ASSESSMENT

Adults who are in need of education attend programs that attempt to support adult learning through the use of effective curricula and instructional processes. To inform and support this educational endeavor, assessment is undertaken for many purposes related to the learner, the curriculum, and the program itself. Unfortunately, assessment practices are often not reflective of these divergent purposes. There is an overall tendency to use only one or two styles of information gathering for multiple purposes; this impacts on the type of information gained from the assessment and its ultimate usefulness.

Assessing Individual Learners' Skills

Assessments of individual students are conducted to gain information on the status of the students' mathematical development. Typically, assessment activities are scheduled at various points in time during the students' involvement in a program and may be used for different purposes to support and enhance student learning.

Initial Assessment to Inform Placement

Diagnostic assessment for an adult learner should encompass the computations the learner can or cannot do, the mathematical experiences and informal mathematics knowledge the learner brings, and the learner's problem-solving and reasoning strategies. This provides a broader basis for designing educational experiences than would a single summary score from an assessment of limited scope.

In Australia and the United Kingdom, where the adult literacy and numeracy sectors have arisen mostly from community-based, holistic philosophies of provision, many teachers are resistant to standardized forms of assessment. For initial ' placement' purposes, they usually use a locally developed *sheet of 'sums'* (i.e., simple computational exercises for which the student has to generate answers), as well as qualitative observations. In the United States, the opposite trend seems to hold; almost three fourths of all U.S. programs rely on or are required by federal and local funding agencies to use a common standardized test such as the Test of Adult Basic Education (TABE; this assessment tool is discussed below), though some programs also use additional measures of a more qualitative nature (Gal & Schuh, 1994).

In practice, both approaches tend to emphasize computational activities similar to primary school arithmetic. The use of a written computational test, whether standardized or not, will assess only a *subset* of the many skills and dispositions the student may possess. While computational skills are important, they are not the only aspect of *numeracy* and thus should not be (and usually, are not) the sole focus or emphasis of an adult education program. Indeed, it is important to place students' initial experience with any formal testing in a broader educational context and prevent an unwarranted perception that computations are the sole focus of mathematics learning for adults.

Initial diagnosis in a workplace context requires special consideration, and should be based on an examination of the actual vocational setting and on an audit of expected performance and needed skills. Identification of skills requires not only a survey of employers as to what mathematical skills they believe the job requires, but also a task analysis involving direct observation of performance and interviews with workers. Information from these two perspectives is often quite different and may lead to the assessment of different skills and different curricular content. (See Lloyd & Mikulecky, in press.)

When considering the degree of 'precision' required of the initial assessment, recent findings by Venezky, Bristow, and Sabatini (1993) suggest that ABE/GED programs may be over-assessing for placement purposes when relying on standardized tests such as the TABE. Venezky et al. show that students' scores on the TABE vocabulary locator test (a brief pretest with items representative of the full version of the vocabulary test) are at least as predictive as the more time-consuming full TABE vocabulary and reading tests or the TALS document or quantitative scales for purposes of student placement. Although this study addressed broad placement issues within ABE and GED classes and did not focus on numeracy issues, the implications for assessment of numeracy skills are clear. The time and resources that can be saved by using shorter standardized tests for initial placement if they are to be used at all (Venezky et al. suggest exploring the efficacy of self-evaluation for placement) could be used instead to assess the informal or practical mathematical knowledge of learners via alternative assessment methods.

Initial assessments are also opportunities for students to identify their own numeracy goals and priorities and make these known to instructors. In programs that emphasize improving competence in functional tasks rather than focusing on specific skill acquisition for a GED test or workplace task, student priorities can have a major impact on curricular design. Implementation of curricula designed with student priorities in mind can lead to high student engagement and satisfaction since instruction is perceived to be useful and applicable.

Ideally, diagnostic assessment in adult numeracy should encompass a broad definition of numeracy and maintain the following goals:

- Establishing the learner's own goals so they can be addressed within the curriculum;
- Determining what knowledge, strategies, and reasoning processes the adult learner already possesses and is able to use;
- Determining what needs to be learned to meet the learner's goals and/or those of the learning setting (e.g., a workplace, certification); and
- Providing guidance on which strategies should be used for effective instruction (including strategies proven effective for individuals with cultural differences or learning disabilities).

Formative Assessment to Track Progress and Inform Instruction

Formative assessment should be undertaken on an ongoing basis to provide feedback both to the teacher and to the student on how the student is progressing, to identify areas that need redressing, and to inform decisions on directions in which instruction might proceed. This type of assessment is continually used informally by adult numeracy teachers, although many teachers are not aware that they are doing this. Formative assessment can involve a mix of formal and informal methods, including but not limited to the teacher talking with or looking over the shoulder of students, the use of extended group projects, and brief written tests. By inquiring about and exploring the strategies and reasoning processes a student uses when solving (or attempting to solve) numerical problems, an interviewer observes, for instance, if the student has a limited, but well understood, repertoire of strategies; has patchy knowledge that can be applied in some contexts but is limited in generalizability; applies 'advanced' memorized algorithms in inappropriate ways; or exhibits other patterns. In addition, while elaborating on responses, the student demonstrates comfort levels and competence with concrete and/or graphical representations and verbal explanations. Since different learners may have different strengths and preferences in how they find meaning in mathematical notions, a variety of assessment strategies may be necessary to communicate effectively with a diverse group of students.

Formative assessments can also provide information to an instructor on the efficacy of alternative instructional strategies and activities. For example, sometimes teachers experiment with alternative ways of teaching a topic to students in a class. By examining the impact of the alternative practices on learners' understanding and performance as demonstrated in informal assessment activities, teachers have an opportunity to evaluate their own practices and improve their ability to help students progress. In such a context, assessment can affect curriculum design, without a direct need to provide complete information as feedback to students.

As with diagnostic assessment, the meaning and effectiveness of formative assessments are dependent on the *definition of numeracy* held by the teacher and the breadth of mathematical activities in the program from computation to problem solving and strategy development. Assessments that focus only on a narrow definition of numeracy (e.g., computational skills such as division of fractions when numbers are presented without context) may not be particularly informative about the student's depth of understanding or ability to use skills in a variety of contexts. For formative assessments, discussion of numeracy strategies and reasoning while engaging in context-laden, problem-solving activities are probably most effective for evaluating gains in mathematical knowledge, understanding, and application.

Summative Assessment to Display Learning and Exit Status

Adult education programs are frequently required to report to outside agencies on students' level of achievement at the end of a course of study. The particular assessment tools used for this purpose may depend on the external agency's constraints. Many programs use assessments that have been chosen by the programs or by funding agencies because they are easy to administer and provide a simple summary score, regardless of whether the skills assessed reflect the content of instruction or the actual goals of the teacher, the program, or the student. Some commonly used *summary score* assessment tools include the Tests of Adult Basic Education (TABE), Adult Basic Learning Examination (ABLE), and Comprehensive Adult Student Assessment System (CASAS). Certified programs often require evidence of progress towards a specific documented goal (e.g., passing a GED test or exceeding a minimum score on a qualifying examination). Other programs, such as community programs and some work preparation programs, may report individual progress towards idiosyncratic goals. Clearly, numerous forms of summative assessment now exist and are accepted as appropriate in different education sectors but the form and content of all of these assessments are quite similar.

The need to provide data to external authorities to satisfy accountability demands has unfortunately distracted many adult numeracy practitioners from principles of good assessment practice. Assessments used for reporting on student achievement should be seen as extensions of assessments used for diagnostic and formative purposes. In actuality, however, summative assessments often examine only a small subset of what adults should and perhaps do learn in numeracy classes; many assessments focus on decontextualized computations whether or not those particular computational skills are important in real contexts or can be applied appropriately. In examining reporting systems in workplace education programs, Mikulecky and d'Adamo-Weinstein (1991) found that few programs had structured assessments, and that most relied instead on (problematic) anecdotal information. The most effective programs, however, did use highly structured assessments, but used ones specifically designed for measuring change in skills identified as relevant in the particular workplace rather than generic tests. Thus, instructors in all settings should undertake systematic, documented, and appropriate assessment of their students' overall progress, after reexamining their definition of numeracy.

Program Evaluation

Program evaluation underwent considerable theoretical exposition in the 1970s and early 80s, whereby the concepts of assessment and evaluation were differentiated. In the 1990s there appears to be a growing convergence or lack of clarity in terminology, with *assessment* used not only for documenting student performance but also for evaluation of a program's effectiveness as a whole.

Program evaluation should consider not only what the students gain from their programs but also the match between these gains and the intentions and expectations of the program. More broadly, evaluation should identify aspects of a program that were successful, as well as aspects that need to be modified, such as instructional methods, student selection, or curriculum orientation. Padak and Padak (1991, p. 376) provide a simple list of six guidelines for adult literacy program evaluation that are equally applicable to adult numeracy contexts:

- Base evaluation on the program's stated goals;
- Make the evaluation comprehensive;
- Make the evaluation systematic (not anecdotal);
- Use a variety of forms of assessment, both qualitative and quantitative;
- Review the evaluation results in terms of the three categories of program effectiveness: personal factors, programmatic factors, external factors; and
- Use evaluation data to identify parts of the program that need strengthening.

Presently, many evaluations of program effectiveness look at "gain scores" (the difference between students' scores before instruction begins and their scores on the same assessment tool after a period of instruction or program completion). Frequently, programs or evaluators report the average gain across all participating students and thus mask different progress profiles. In light of the six guidelines listed above, individual students' change scores on standardized basic skills assessments do not seem to provide very much meaningful data that could be of use in program evaluation. Indeed, even the information they do provide may be suspect because, as Venezky (1992) notes, (a) initial scores tend to be artificially low due to students' poor test-taking skills and lack of recent skill practice so therefore, gain scores may be artificially high; and (b) test content often does not reflect actual curricular content.

CONSIDERATIONS FOR MEANINGFUL NUMERACY ASSESSMENT

This section explores a number of factors that must be taken into account when designing or evaluating numeracy assessment tools so that resulting assessments can be used in meaningful and important ways to inform decisions about individuals, instruction, or programs.

Validity and Reliability Concerns

Training of teachers (in colleges) has traditionally included courses on topics such as "measurement and evaluation" with an emphasis on psychometrics and standardized tests (Linn, 1990). It is thus not surprising that many teachers in all education sectors, as well as policymakers, turn to standardized tests; such tests appear to satisfy rules of psychometric robustness, are readily available, and are easy to administer and score. However, we need to question the extent that the principles of reliability and validity and their underlying assumptions are meaningful and appropriate for adult numeracy assessment purposes.

In the past, tests have been designed with the assumption that a test is a method for measuring latent traits or abilities that are not in themselves directly observable but that are measurable by inference. For example, items on intelligence tests are intended to measure the intangible construct of intelligence by ways of specific manifestations or representations of this construct, such as recognition of patterns, richness of vocabulary, and so forth. Likewise, tests of mathematical skills claim to examine students' *mathematical ability* through their performance on samples of items that presumably require the activation of specific, significant skills.

The quality of assessment instruments and practices has traditionally been evaluated in terms of their psychometric properties, which are in part determined through statistical studies. Psychometric definitions have emphasized the reliability and validity of such tests, defining *reliability* as the consistency with which a test measures a domain and is free of error. For example, a test administered on repeated occasions, say six weeks apart with no instruction or expected development occurring in between, should produce the same result. Also, different judges evaluating the same open-ended response or performance should in general reach similar conclusions.

Validity has been defined as the degree to which an instrument is representative of the (content of the) domain that it is testing. Most reference books talk of four types of validity: face, content, predictive (criterion-related), and construct. However, it is important to keep in mind that

> Validity is an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the *adequacy* and *appropriateness of inferences* and *actions* based on test scores or other modes of assessment...hence what is to be validated is not the test or observation device as such but the inferences derived from test scores or other indicator. (Messick, 1989, p. 5)

The *interpretation of scores* that people obtain on psychometrically developed tests has been traditionally done through reference to group norms (i.e., what others, in a given age/grade level, can do). However, such tests make a number of major assumptions about the distribution of performance in a population, the characteristics of the population itself, and the transportability of such performances across contexts. In addition, defining group norms for adult performance in terms of children's grade levels (as many tests do) may suggest inappropriate interpretations that do not take account of the differences between the breadth of adults' versus children's experience. For example, an adult who scores 5.6 on a standardized test may demonstrate quite different performance patterns than those of 11-year-old children.

When considering numeracy assessment, all assumptions should be examined to see if they are being met or are appropriate: (a) What underlying mathematical ability are we measuring? (b) Would such an ability be distributed throughout the population in particular ways? (c) Is what we are measuring important or meaningful to measure? and, (d) What non-mathematical skills (such as literacy skills) may account for performance and need to be taken into account in interpreting and inferring from assessment results? Interpretation of standardized test results may lead to erroneous inferences when the scores in question come from adult students with diverse characteristics and unique life histories.

Reliability and validity are necessary, but not sufficient, criteria as they do not embrace the full scope of issues involved in constructing and using valuable assessments in the context of adult mathematics education. Issues that are more conceptual rather than technical should also be considered in the design and use of good adult numeracy assessments.

General Principles of Good Assessment and Their Implications

Good assessment is that which is tied to particular contexts and purposes and can inform students', teachers' and program activities. Assessment should also reflect (for teachers, students, and policymakers) what is valued in a student's performance and learning. The recent report by the Mathematical Sciences Educational Board (MSEB, 1994) has identified three clear principles of good assessment in mathematics:

- *The Content Principle*: Assessment should reflect the mathematics that is most important for students to learn.

- *The Learning Principle*: Assessment should enhance mathematics learning and support good instructional practices.
- *The Equity Principle*: Assessment should support every student's opportunity to learn important mathematics.

These principles are also at the core of the *Assessment Standards for School Mathematics*, a key document released by the National Council of Teachers of Mathematics (NCTM, 1995) that lists extensive guidelines and provides many illustrative examples for alternative assessments. Overall, these principles reflect changes occurring in formal mathematics education in schools (Lesh & Lamon, 1992) and they reinforce the often-stated axiom that assessment practices need to reflect good instructional theory (Shepherd, 1991). The MSEB and NCTM perspectives on assessment were developed in the context of K-12 education in the United States but the ideas can be extended to inform adult education assessment in numeracy.

The Content Principle

For an assessment to be meaningful, the tasks should reflect the content we consider important and want to emphasize in our adult education system. In our modern world, low-level, procedural mathematical skills (such as long, repetitive computations) are increasingly being done by calculators or computers. In addition, many "mathematical" situations do not require computations but rather the ability to look at a body of information or data and identify and interpret that which is either problematic or meaningful. It therefore has become essential for individuals to develop problem-solving skills and a conceptual understanding of mathematical ideas so that they will be able to use available tools when necessary to solve daily and workplace problems. For example, Scribner and Stevens (1989) describe a work scenario in a dairy produce factory wherein workers invented and deployed mathematics procedures that were accurate, fast, and flexible, but not based on the standard number base '10' or heuristics developed in school mathematics. These workers were mentally and visually calculating in several number bases simultaneously because of the need to deal with specific computational problems created by the work demands. Consideration of the real performance of these workers illustrates that teachers need to develop assessment instruments that demonstrate the breadth of things that students "can do" and likely "will do" rather than "can't do" and "would never do."

"Doing mathematics," whether during instruction or in an assessment, should emphasize or at least enable the appearance of "higher level" strategies and processes involved in the solving of meaningful, multifaceted problems. The lower level computational procedures are the tools that can be called upon when needed but should not be the sole focus of either instruction or assessment. Focusing assessment on these lower level skills is tantamount to focusing assessment of reading skill on letter recognition rather than on comprehension of meaningful text. If the goal of instruction for a particular program is to develop everyday or workplace numeracy skills, then the content of assessments (i.e., tasks) should give rise to those functional skills students are expected to acquire.

In light of these content issues, a number of related concerns should be addressed. The use of calculators during numeracy assessments should be thoughtfully evaluated. Since calculators are readily available and their use is expected in virtually all out-of-school contexts, their use as a tool for all numeracy education purposes seems reasonable. Similarly, if the ultimate goal of the educational endeavor is to enable students to become independent learners and problem solvers, other resources (such as reference books, even "times tables") should also be available. If access to "outside" resources must be limited during assessments, then questions on the assessment might include "What information would you have to find out in order to decide…?" or "How would you collect information to explain…?" Finally, time constraints during assessment activities may only be providing information on how a person copes with a time constraint, not on the extent or depth of mathematical understanding. In most situations in which adults actually use numeracy skills such as decision-making, shopping, or problem solving, time limitations are often self-imposed and generally somewhat flexible (taking a few more minutes to rethink a problem or recalculate an amount is considered reasonable and responsible). Increasing student anxiety by creating time pressure usually does not serve the purposes of the assessment well.

The Learning Principle

Students learn from all that they do, including their assessment tasks. When students complete an assessment, they comprehend that the form and content of that assessment is considered by the instructor or program to be of prime importance. If a student's numeracy status and progress are measured by assessments that emphasize swift computational accuracy, then the student feels there is little point in spending time trying to understand the meaning and implications of what he or she is doing (and in trying to integrate the mathematical concepts within activities away from the classroom). The phrase “I don’t care why, I just want to know how to get the answer” becomes a common refrain. This type of learning was unproductive for many people when they were in the K-12 system and there is evidence that it limits ability to transfer learning across contexts (Nunes, Schliemann, & Carraher, 1993; Perry, 1991). The messages about mathematics learning and the scope of numeracy conveyed to adult students are of particular importance since, as adults, they will and should monitor their own learning and prioritize learning-related activities.

Since adults’ educational time is often limited by demands from work, family, and other sources, the time spent in learning environments should be used as effectively as possible. Thus, time for assessments should also double as time for content learning. Ideally assessment tasks should foster learning by encouraging students to make and reinforce connections between ideas or between ideas and experiences or to construct meaning. These tasks would necessarily be broad, require a variety of problem-solving strategies, perhaps require production of a verbal explanation or communication, and probably have more than one reasonable solution path. They would not be limited to responses that require only memorization of a procedure that could be performed upon specific request.

Assessments can support instruction by providing important information about students’ mathematical skills, experiences, and levels of understanding. Therefore for maximum information, assessments should provide a multidimensional picture of a student’s numeracy status. Limited assessments, those that explore only one or a few areas of mathematical knowledge, provide sparse, one-dimensional information.

The Equity Principle

All students should have the opportunity to learn and to display that learning. Because students come to adult education from different backgrounds and may have had different experiences, perhaps it is unrealistic to expect them all to display growth in numeracy in the same ways and in response to the same assessment stimuli. Certain problem contexts may not be familiar or relevant to all students. If assessments include a number of different types of tasks (i.e., some verbal, some visual, others tactile, some to be done in collaboration with others and some alone), many more students may display areas of strength and indicate foundations upon which to build further success.

Many assessments may be flawed with respect to cultural differences. Adults who immigrate to another country (say from Europe to the United States or from Korea to Australia) are often given a written placement test upon first coming to an adult education class. Superficial characteristics of the test could have a large impact on students’ results (and on their ease in functioning in their new country) if teachers do not know much about students’ backgrounds, including the nature of schooling in mathematics and the conventions of the numeration system in the country of origin. The number “five thousand,” can be written in different cultures as

5,000.00 5 000.00 5 000,00.

The disappearing (or moving) comma may well create confusion for some adults being tested. Often teachers are unaware of this and similar phenomena among immigrant students and do not identify sources of difficulties with reading, communicating about, and computing with numbers.

The ways different languages enumerate quantities and describe mathematical symbols and operations can also impact on conceptual understanding and learning; for example, the concepts represented by number names in English and Chinese are quite different, as shown in Figure 1. How confusing it must be for a Chinese person now trying to learn and prepare for work in English!

Obviously, test design and the interpretation of test scores need to be done with caution in many cases. To heighten the awareness of learners (and educators) to the impact of cultural diversity on learning and assessment, teachers and students should discuss cultural differences in, for example, representations of numbers and number words, conventions for writing symbols and operations (such as for denoting long division or multiplication), and culture-specific beliefs about broader aspects of mathematics learning (such as the importance of memorization vs. understanding).

ASSESSING NUMERACY SKILLS WITH STANDARDIZED SHORT-ANSWER ASSESSMENT TOOLS

A recent large-scale national survey showed that over 70% of all adult literacy programs in the United States use some standardized short-answer assessment instrument, with the Tests of Adult Basic Education (TABE) used in over 48% of the programs (Gal & Schuh, 1994). In this

Figure 1: English and Chinese Number Nomenclature (using phonetic translation of Chinese symbols)

Number	English	Chinese
1	one	gi
10	ten	shi
100	hundred	bai
1,000	thousand	tsin
10,000	ten thousand	wan
100,000	hundred thousand	shi wan (literally: ten "wan")
1,000,000	million	bai wan (literally: hundred "wan")
10,000,000	ten million	tsin wan (literally: thousand "wan")
100,000,000	hundred million	yi
1,000,000,000	thousand million (Australia) billion (U.S.) milliard (Britain)	shi yi (literally: ten "yi")

section we use the TABE as a case study to examine the extent to which it reflects the principles of good assessment as defined above.

Evaluating One Commonly Used Numeracy Assessment Tool: The TABE

The TABE is published commercially by CTB/McGraw Hill for the purpose of assessing achievement in reading, math, language, and spelling skills. It includes four tests that the test manual describes as assessing overlapping skill levels roughly corresponding to grade levels 2.6-4.9 (Easy), 4.6-6.9 (Medium), 6.6-8.9 (Difficult), and 8.6-12.9 (Advanced). A brief Locator Test is used to determine the approximate skill level of the individual and help decide which of the four tests should be administered to obtain a detailed analysis of the person's skills.

Each of the four tests includes seven subtests: vocabulary, reading comprehension, language mechanics, language expression, spelling, mathematical computations, and mathematical concepts and applications. The test is timed and a full administration of all subtests takes almost 3.5 hours. All items in all subtests use a multiple-choice format requiring the test taker to choose one of four possible answers. A person's total score is normally translated into a "grade equivalency" in the range 1.0-12.9, based on conversion tables in the manual.

The Mathematical Computation subtest takes 43 minutes and includes 48 multiple-choice items requiring test takers to compute the answer to problems of the type "605 X 256=?" or "3.7 - 2.251=?"; items in this subtest do not use any words. The Mathematical Concepts and Applications subtest takes 37 minutes and includes 40 multiple-choice word problems. Some items briefly describe a situation and ask for a computation; others pose a question about mathematical information presented in a graph, diagram, or coordinate system, usually without any context. (Some examples are analyzed in detail below.)

Before discussing further the content of the TABE, it is important to note that the test manual (Tests of Adult Basic Education, 1987) asserts that the content of the test "…focuses on basic skills that are required to function in society. Because the tests combine the most useful characteristics of norm-referenced and criterion-referenced tests, they…enable teachers and administrators to diagnose, evaluate, and successfully place examinees in adult education programs" (p. 1).

Below we consider the extent to which the TABE mathematical subtests reflect the three key assessment principles discussed above as well as additional issues pertaining to the meaning of the grade levels determined by the test.

Content

The TABE Level E (Easy) Computation subtest focuses mainly on whole number operations taught in the first three years of school, with a few items involving fractions and decimals; the Level D (Difficult) Computation Test is almost entirely focused on decimals and fractions. While percents are usually introduced in the school curriculum in grades 5 or 6, there are no items on percent in Level M (Medium, which corresponds to grades 4.6-6.9). Test D (Difficult, grades 6.6-8.9) contains a total of five percent questions, two in the Computation subtest and three in the Concepts subtest. Only one of the latter three questions pertains to a realistic context. Furthermore, percent questions in Level A (Advanced, grades 8.6-12.9) are mostly restricted to interpretation of percents in pie charts, and do not represent the range of real-life situations in which adults have to use or make sense of percents.

The implication of the above observations is that the range of mathematical topics and skills identified as important in these subtests is quite limited. Percent, for example, is of critical importance in adults' lives; percents are frequently encountered in everyday shopping and financial contexts, and understanding them is imperative for adults to be considered effective consumers or for understanding of trends and statistics reported in the media (Ginsburg, Gal, & Schuh, 1995).

The mathematical content of any test is embedded within contexts although the "context" of the Computation subtest of the TABE is "isolated computations." The Concepts and Applications subtest purports to set mathematical content within "real" contexts, however, only a modest portion of items are about everyday events in the lives of many adults. Sticht (1990), for example, noted that of the 40 items in the Concepts and Applications subtest in Level M (Medium, grades 4.6-6.9), only one item is about calculating the correct change for a given transaction, no items address savings from bulk purchases, and no items are about the total cost of a purchase with an installment plan and finance charges. (Note that such topics are of as much importance in the lives of adults taking Level M as in the lives of those taking more advanced forms.)

When we examined item contexts in the Concepts and Application subtest of all four levels of the TABE tests, the deficiencies noted by Sticht regarding the Computation subtest appear even more glaring. For example, of the 40 items in this subtest in Level D (Difficult), 22 have no context at all (e.g., "Which group of numbers contains common multiples of 2 and 9?"); six items have only a nominal context, usually related to systems of measurement, time, or money (e.g., "How many millimeters are equal to 2 centimeters?"); and seven items have a superficial context (e.g., "It is now 2:30 p.m. What time was it 3 hours and 5 minutes ago?"). Such items appear contrived, as if created for the express purpose of testing computational knowledge, not as a simulation of a realistic problem that motivates mathematical reasoning. Only five items out of the 40 seem to simulate a real-world problem (e.g., "Mr. Rankin made a deposit of $10 on a hotel room that costs $16 a day. How much money did he owe on the room after 7 days?").

Learning

Each of the items on the Concepts and Applications subtest can reveal some aspect of a person's mathematical knowledge and may thus be deemed valuable for some assessment purpose. However, few items in the TABE enable test-takers to demonstrate the real-world mathematical knowledge that they may possess. Thus, when used as a placement test, the test is likely to underestimate the skills and knowledge that learners possess. The over-emphasis on school-like tasks means that test-takers who have been away from pure computational tasks for years may be "rusty" and thus not perform well on decontextual tasks that require memorization of school-based procedures that are seldom used in everyday contexts, such as multiplication and division of fractions. Indeed, Lave, Murtaugh, and de la Rocha (1984) found that adults (who were not involved in adult literacy education) scored an average of 59% on arithmetical problems related to shopping tasks presented in a written test but averaged 98% when confronted by roughly equivalent problems when shopping in a real supermarket.

This focus of the TABE reinforces narrow perspectives on desired mathematical knowledge and numeracy that are no longer accepted in mathematics education as noted earlier. By taking a test that focuses on these narrow definitions, students perceive that the goal of numeracy education is limited to the acquisition of computational skills even if the students cannot apply the skills usefully or reasonably.

In addition, the multiple-choice format used in both the Computation and Concepts and Applications subtests reinforces the notion that all mathematical activity leads to answers that are either correct or incorrect. Good mathematical principles of application, estimation, and appropriateness of an exact or rough answer are not reinforced. Furthermore, the multiple-choice format means that no information is available to the assessor about the strategies being used by the test-taker to solve the questions or about possible sources of error and confusion, all information that could help inform instruction. For example, for the question

Which of these can be evenly divided by 3?

F	81
G	83
H	86
J	89

the response "F" may indicate that respondents worked out the answer by laboriously dividing each two-digit number by 3, or perhaps they used a simple heuristic such as the numerals of numbers divisible by 3 add up to a number also divisible by 3, or maybe they were running out of time and selected “F” at random because a former teacher told students never to leave questions unanswered on a multiple-choice test. Indeed, an incorrect response to this question could reflect (a) an unsuccessful random guess, (b) a careless error using a division algorithm, (c) an inappropriate strategy (i.e., “83 ends in a three so the number must be divisible by 3,”), (d) confusion differentiating the phrase “divided by 3” from “divided into 3” (a distinction that is problematic for many students with limited knowledge of English as well as many native speakers), or (e) incorrect use of a calculator if its use was allowed. Without having answers to these and similar questions, the use of such an item is limited for informing placement decisions or designing meaningful instruction.

An over-emphasis on decontextual problems can also bias the interpretation of scores on pretesting and posttesting in reporting contexts. Typically the TABE is administered to new students before the onset of any instruction. Once instruction begins, students are helped to brush up on their general test-taking skills and on procedural skills. When the TABE is re-administered, many students demonstrate a sizable gain in TABE scores within a few weeks, which looks impressive when programs report “change scores” to funders. However, most of this gain is artificial and does not represent true change in underlying knowledge or conceptual understanding. Such score gains may also mislead students to believe they have learned more than they really have.

Many TABE items also appear to indicate a hierarchy not of conceptual complexity but of computational complexity. With access to calculators that today can handle fractions as well as decimals, are such computations necessarily more complex at all? These computational subtests offer little to the adult numeracy student or teacher in terms of the MSEB’s Principle of enhancing learning and supporting good instructional practices. There is no emphasis in the test on strategies used by students, ability to apply knowledge appropriately, or problem-solving processes.

Finally, when the TABE is used as the entry and departure test for a program, the test itself often becomes the driving force for the instructional program. Computational performance will tend to be emphasized, duplicating the school environment within which many adult numeracy students experienced frustration and failure, while other important numeracy activities will be de-emphasized.

Equity

The equity principle implies that all students should be given the opportunity to learn and to display that learning. Some apparent confusion and “misunderstanding” of numeration system conventions have been linked to cultural differences (as discussed above). This factor as well as other “non-mathematical” issues may contribute to incorrect responses on some TABE tasks but targeted remediation may not occur due to the lack of information about students’ reasoning behind responses. For example, one question on the Concepts and Applications subtest in Level D (Difficult) may be particularly confusing to some students for reasons other than lack of mathematical understanding:

What is another name for 59,600 ?

F 500 + 90 + 6
G 5000 + 900 + 60
H 50,000 + 9000 + 60
J 50,000 + 9000 + 600

Children complete such tasks in primary school to demonstrate their conceptual understanding of place value. Of course, a child can learn to do the tasks without ever having a sense of the purpose of the task or a concept of place value. Conversely, adults may forget how to “do” these sums but may have developed a keen sense of numerical place value from life experiences, such as in the context of money and commercial transactions. Thus, it makes much more sense to assess knowledge of place value by asking adults (or older children) whether they would rather have $100, $10, $1,500, or $1,050.

Other characteristics of this item are also worth addressing. Why is the comma used for the 50,000 but not for the 9000 or 5000 (in the original test booklet)? Would a comma (as in 9,000 or 5,000) give a clue? Is this not only dubious mathematically but also making a false presentation to the student? What information does an item such as this present to the student about what elements in mathematics are important? What instructional activities are such an item likely to lead to?

Living With Existing Standardized Numeracy Assessment Tools

The numerous problems and shortcomings of short-answer standardized tests suggest that, contrary to the claims of test developers, these tests are of limited value as assessment tools to support both diagnostic/placement and summative/reporting functions. It is therefore surprising that such a large proportion of programs in the United States continues to use them and that policymakers continue to require their use.

One particular problematic feature of many standardized assessments is the use of "grade levels" to describe the educational status of the adult learner. The grade level score presumably indicates that the adult's performance on the assessment is similar to that of a child at a particular point in the child's schooling. The use of this terminology is humiliating and demoralizing to adults and certainly does not reflect the complexity of adults' experiences in the variety of settings in which numeracy is embedded. Adults often have patchy mathematical knowledge resulting from their incomplete schooling and self-developed strategies for coping with real world tasks (e.g., understanding and able to compute percentages while shopping but not on a paper and pencil "test" of percents; Ginsburg, Gal, & Schuh, 1995; Lave, Murtaugh, & de la Rocha, 1984). Adults may be able to compute 90% of a number (taught in 6th grade) yet not be able to do long division (taught in 4th grade). Thus, since the patterns of adult learners' numeracy skills and knowledge do not closely resemble those of most children at particular grade levels in K-12 schooling, using K-12 labels to describe adults is misleading and lacks purpose. Grade level scores provide no information on learners' particular strengths, weaknesses, or knowledge gaps.

However, as long as use of the TABE and other similar assessments are required, educators can glean some useful information from them. For example, some teachers have created detailed lists of the mathematical content explored by each item. These lists enable teachers to track the questions answered correctly in each of many mathematical areas, such as multiplication, division, place value, percents, and so forth. This information is then used to go beyond the "grade equivalency" information and identify specific areas of knowledge where remediation is warranted. Some teachers have also mapped test items onto specific sections of available textbooks and workbooks, and point students to resources that they can use on their own. Other teachers have created a diagnostic subtest from the TABE with items from all key mathematical domains covered by the test, and administer it in an interview format to individual students to study their solution processes and reasoning about problems.

Such steps are encouraging and can perhaps alleviate some of the many shortcomings of these assessments. Yet, the lack of context for most items, and the inattention to communicative skills and to reasoning processes under regular administration of the assessments, require that teachers use results from the TABE and similar tests with great caution and supplement their use with additional measures, some of which are discussed below.

EMERGING PERSPECTIVES ON MATHEMATICS ASSESSMENT

For most teachers, memories of mathematics assessment at school include visions of endless arithmetic computations, geometry proofs, and word problems. The emphasis in assessment was on "correctness," although as the problems became more complex, partial marks (scores) could be obtained for being "correct" in some of the reasoning. These types of activities are still valuable for developing skills and logic. However, all of these assessment activities encouraged convergence of mathematical thinking, arriving at the one right answer preferably via the one best solution path. Often, directions stated "show your work" so the instructor could be assured that you used algorithms correctly.

The essence of problem solving in many real-life situations is divergent thinking. Tasks may not be clearly defined, considerations other than simple problem variables may be significant, and many real-life tasks do not have one "correct" answer or one "best" way of solving the problem. Reasonability and correctness may vary according to the context in which a problem is set. For example, a school word problem might purport to simulate real-life problem solving by describing packages of coffee that differ in size and price and asking "Which is the cheaper coffee to buy?" In such a question, the task requires calculating and comparing prices per ounce and identifying the package with the lowest price per ounce of coffee. However in real life, if one does not drink coffee frequently, the coffee might well go stale; purchasing a large quantity at a low price per pound might not ultimately be cheaper if a significant portion of the purchase is discarded (Willis, 1990). Similarly, a question such as "You are having 5 children over for a party and want to serve 3 cookies to each child. How many cookies should you buy?" could (and did) elicit the response "20, because you never know if someone will bring a younger brother or sister" (Ginsburg, 1992). The answer, "20," reflects an effective mathematical strategy for real life, but would be considered wrong in a "school setting" if the question poser were only interested in evaluating responses for computational accuracy. On the other hand, for some vocational numeracy tasks, precise accuracy may be essential. It is therefore important that judgments about accuracy and correctness are made with respect to the demands of context in which the problem is posed, rather than to formal concepts of accuracy (i.e., what is computationally correct). For adults who come to mathematics learning with real world experience and with personal goals in mind that go beyond the school context, meaningful assessments should clarify contexts and parameters rather than excise them for simplification.

New trends in mathematics assessment include tasks and activities that recognize the diversity of mathematical thinking and encourage creativity and divergence of thought. These assessment approaches are aligned with current pedagogical approaches, primarily constructivism (Davis, Maher, & Noddings, 1990), and emphasize development of understanding and reasoning rather than rote learning, thereby focusing as much on process and strategy as on outcome.

Portfolio Assessment

Portfolio assessment, being developed in elementary and secondary education in many countries including the United States, offers possibilities for adult numeracy assessment. A portfolio parallels the notion of an art student's portfolio, being a means of collating examples of different types of work that a student has produced; it can include examples of best work and/or chart progress through a course, with examples of the student's work at various points in time (beginning, weekly, conclusion) during a course. Among the immediate benefits of the gradual collection of work is feedback for students; a portfolio can provide tangible and meaningful evidence of progress.

Portfolios are gaining in popularity as assessment tools (see NCTM, 1995; Stenmark, 1991), although the concept of "portfolio" is being operationalized differently in different settings. Some adult literacy programs have begun experimenting with portfolio assessment but apparently only in the content areas of reading and writing (Fingeret, 1993). In an adult numeracy setting, learners could document their mathematical knowledge and strategies on entry (using a broad definition of numeracy), record their goals, progressively note goals reached and increases in mathematical activity outside of class, and display evidence of their final achievements as noted above. These final achievements could incorporate standardized tests if appropriate.

These approaches should not be perceived as "soft"; they can be highly structured. However, in situations where information on progress is required for summative reporting purposes, issues such as validity and comparability of portfolios are still being investigated. A likely solution appears to be the development of specific criteria for grading portfolios rather than procedures for norm-referencing. To establish comparability of standards and provide checks on reliability and validity, a process known as "moderation" has been used; graders provide peer reviews of each others' evaluations of sample portfolios until consensus is reached.

Competency-Based Assessment

In contrast to the movement toward instruction in and assessment of problem-solving processes and strategies, the vocational education communities in the United Kingdom, Europe, and Australia have moved toward competency-based instruction and assessment. Many countries have also developed lists of core or generic competencies that are considered prerequisites for competent performance in the workforce, not only in vocational education. (For instance, in Australia, core competencies are described in general statements such as "Investigating mathematical ideas and techniques"; Mayer, 1992.) "Competency-based" relates to the identification of specific tasks in work or life and successful completion of these tasks. A list of competencies can be extensive and include overarching competencies and layers of subtasks. Examples of numeracy competencies include 'can calculate the total of a dinner bill for a three-course meal with drinks for two customers' and 'can arrange seating for a meeting of 12 people.'

While lists of core competencies may be gaining acceptance, this area is not without debate particularly as competencies are expressed in terms of minimum performance and assessed as either achieved or not achieved. Issues of standards of performance and the difference between competence and expertise are being raised. An additional criticism of competency-based assessment in mathematics and numeracy is that it does not allow students to do extended analyses, solve open-ended problems, or display a command of complex relationships (Resnick, 1987), a critical element in good instruction and in numeracy assessment.

The identification of core competencies is similar to the identification of job performance characteristics through task analysis and the development of corresponding criterion-referenced assessment activities, a common practice in workplace training contexts (Shope, in press). Wolf (1991), however, has argued that core skill competencies are by definition inseparable from the contexts in which they are developed and displayed, and thus should not be assessed separately. Wolf's position is shared by other recent perspectives on the need for workers to have integrated skills that are more than just an aggregate of individual competencies. (See Carnevale, Gainer, & Meltzer, 1990; Lloyd & Mikulecky, in press; O'Neil, Allred, & Baker, 1992; SCANS, 1991.)

Authentic Assessment

Another recent development in mathematics assessment has been the focus on authentic assessment, tasks that involve "(i) real mathematics, (ii) realistic situations, (iii) questions or issues that might actually occur in a real-life situation, and (iv) realistic tools and resources" (Lesh & Lamon, 1992, p. 18). These activities are different in breadth and scope from traditional school-like word problems; students are expected to bring their own experiences and judgments to bear on problems that are complex, require decisions about the relative importance of pieces of information, and may not have a single, correct response. Supporting the use of authentic assessment, Lesh and Lamon (1992) theorize that people individually develop models that they use to understand situations and interpret information within the contexts of those situations so that they can make decisions. When an individual applies these models, the models are refined and elaborated as additional constraints become understood and additional information is incorporated. Therefore, instruction and assessment should encourage and measure growth in the sophistication, usefulness, applicability, and generalizability of students' models.

Since adults do make decisions every day and use their "models" of how organizations and processes work to structure and inform those decisions, the authentic assessment approach potentially can be informative and useful within adult numeracy educational frameworks. However, at this point, while some good demonstration problems have been created that can be used effectively in instructional settings (Cumming, 1995; Lesh & Lamon, 1992), complete assessment tools based on authentic assessment principles are not currently available.

To reflect the authentic situations in which people use mathematical thinking, assessments do not have to be limited to written formats. Students can construct their own answers and then explain their reasoning through individual and/or group oral presentations. In the United Kingdom, oral presentations are being required even at the university level; such presentations must be mathematically rigorous, but must also emphasize the logic behind an approach and include reflection on the quality and adequacy of processes employed. Since adult students have a range of life and work experiences and often possess mathematical ideas that may not be reflected in multiple-choice testing, and since the development of adults' oral presentation skills is an important component of effective communication, assessments that rely on and support growth in this area will both provide and assess learning experiences.

SUMMARY AND IMPLICATIONS

Appropriate assessment has to be clearly related to and directed by the instructional focus of the class or program, and express what is valued regarding what students are to know, do, or believe (Webb, 1992). The field of adult numeracy needs to reconsider its perspectives on assessment in cognizance of recent reforms in the fields of mathematics education and assessment theory and practice. On the basis of these considerations, there are a number of implications for future assessment practices in adult numeracy and literacy programs:

- Both instruction and assessment of adult numeracy skills should be informed by a broad definition of numeracy, and should encompass the work and mathematical life experiences and strategies adults bring.
- Ideally, assessment should address reasoning processes and (mathematical) problem solving, conceptual knowledge and computation, ability to interpret and critically react to quantitative and statistical information embedded in print or media messages, as well as examine transfer of mathematical problem solving across life and work contexts (at a minimum, students' ability to apply mathematical/numeracy skills in meaningful contexts).
- Assessment content should parallel the instructional focus and goals of the program.
- One type of assessment alone (e.g., use of standardized tests) will not be sufficient to inform all placement and instructional decisions and program evaluation requirements.
- Convenient and apparently simple assessments such as standardized tests may not be appropriate or informative, and may do a disservice to students, teachers, and a program.
- Adult numeracy assessment should encompass the range of assessment

forms being used in other educational settings including oral reports, group activities, and portfolios; the numeracy-related strategies that a person uses will be more apparent in these activities than from a multiple-choice test.

- Adult numeracy assessment should recognize that adult learners may perform at quite different levels in oral mathematical discussions than on written tasks.
- Assessment indicators for workplace programs are most appropriately drawn from a task analysis of work.
- Assessment should inform students in a systematic way of their progress and achievement in a program.
- Only appropriate interpretation and use should be made of assessment information; adult numeracy practitioners need to be aware of cultural differences that can affect performance in assessments and subsequent interpretation.

Implications for Practice

Many teachers may be presently unaware of the promise of new assessment forms or they may have had few chances to try them out in a systematic way for a reasonable duration of time. As continual implementation of alternative assessments will require more intellectual and time investments from teachers, there is little that can be done by external players in the field to force teachers to change their assessment practices in this regard (Ball, 1990). University-based researchers and developers can advocate and present relevant information and suggestions, yet it is up to practitioners to learn about and begin to experiment with new forms of assessments in collaboration with their students, colleagues, and program administrators, in order to discover what types of improvements in teaching, learning, and achievement can be realized.

Several forms of practitioner inquiry processes exist that can support teacher experimentation. Some teachers may decide to try informally one of the above ideas in their classrooms; others may create small teams to experiment with and explore one or two ideas within their program, or create a supportive environment in which informal yet systematic reflection can develop. Depending on local conditions and motivations, it is also possible and desirable to implement a more elaborate and systematic inquiry process in which a larger group of teachers and administrators from several programs is involved over a period of time; participants can aim to document and publish personal accounts of experimentation and change, as well as summaries of the group effort as a whole. Such practitioner inquiry projects or study circles have been implemented in different settings. (See Cochran-Smith & Lytle, 1993; Lytle, Belzer, & Reumann, 1992; Schmitt, 1994.)

Teachers, with a vested interest in providing the most effective and valuable instruction for their students, need to carefully examine new assessment procedures that cover the full range of skills and dispositions encompassed by the term numeracy. They will then be prepared to argue for fair and valid uses of assessment for accountability purposes and will have access to powerful examples to support their arguments.

Implications for Policy

While the ideas discussed may be appealing and reflective of good practice, realistically, many practitioners in adult education have very limited access to preservice training and professional development opportunities, in part due to funding issues; also, many work part time, and lack a strong background in math education (Gal & Schuh, 1994). If hiring and training practices stay unchanged, few resources and little time will be available to train teachers in using more informative, yet also more demanding and costly, alternative assessments. Under such circumstances, teachers and programs may continue to rely heavily on standardized tests, as they are convenient to administer and easy to score, and their results, though not very meaningful, are easily reportable. Funders in turn will expect programs to continue to administer such tests, leaving the situation that was lamented earlier in this chapter largely unchanged.

Many policymakers rely on testing results to hold educational institutions accountable for student progress. Funding agencies typically require programs to report test scores of students and often suggest the use of specific testing instruments. These reporting requirements support the continued use of assessment techniques that do not reflect the overall goals of numeracy education, do not reflect the current pedagogy in mathematics education, and do not provide meaningful information to students, teachers, programs, or agencies. Since continued funding may be contingent on improving scores on standardized tests, teachers may feel direct or indirect pressure to "teach to the test" rather than focusing on the broader aspects of numeracy instruction.

While a change in policies or official reporting schemes would greatly help, a top-down process of change has a limited chance for achieving a measurable impact on a complex field such as mathematics education (Lindquist, 1994); many adult education programs are based on local and independent management, on volunteer operations, or emphasize learner-based approaches, and may thus not respond to or be out of reach of government mandates. Therefore, to improve assessment practices throughout the adult education community, an open, multi-pronged and long-range approach is needed that seeks the involvement of all stakeholders and recognizes the complexity of the task and its solutions. ■

REFERENCES

Australian Committee for Training Curriculum (ACTRAC). (1994). *A national framework of adult English language, literacy and numeracy competence*. Victoria, Australia: ACTRAC Products Ltd.

Ball, D. L. (1990). Reflections and deflections of policy: The case of Carol Turner. *Educational Evaluation and Policy Analysis*, *12*(3), 263-276.

Carnevale, A. P., Gainer, L. J., & Meltzer, A. S. (1990). *Workplace basics: The essential skills employers want*. San Francisco: Jossey-Bass.

Cochran-Smith, M., & Lytle, S. L. (1993). *Inside/outside: Teacher research and knowledge*. New York, NY: Teachers College Press.

Cumming, J. (1995). The adult numeracy student as problem solver. *Good Practice in Australian Adult Literacy and Basic Education*, *27*, 9-10.

Davis, R. B., Maher, C. A., & Noddings, N. (1990). Constructivist views on the teaching and learning of mathematics. *Journal for Research in Mathematics Education Monograph No. 4*. Reston, VA: National Council of Teachers of Mathematics.

Fingeret, H. A. (1993). *It belongs to me: A guide to portfolio assessment in adult education programs*. Durham, NC: Literacy South.

Gal, I. (1994). Reflecting about the goals of adult numeracy education. In I. Gal & M. J. Schmitt (Eds.), *Proceedings: Conference on adult mathematical literacy* (pp. 19-24). Philadelphia: University of Pennsylvania, National Center on Adult Literacy.

Gal, I., & Schuh, A. (1994). *Who counts in adult literacy programs? A national survey of numeracy education* (Technical Report TR94-09). Philadelphia: University of Pennsylvania, National Center on Adult Literacy.

Ginsburg, L. (1992). *Young children's solution strategies for elementary proportion problems*. Unpublished doctoral dissertation. Milwaukee, WI: University of Wisconsin-Milwaukee.

Ginsburg, L., Gal, I., & Schuh, A. (1995). *What does "100% juice" mean? Exploring adult learners' informal knowledge of percent* (Technical Report TR95-6). Philadelphia: University of Pennsylvania, National Center on Adult Literacy.

Kirsch, I. S., Jungeblut. A., Jenkins, L., & Kolstad, A. (1993). *Adult literacy in America*. Princeton, NJ: Educational Testing Service, National Center for Education Statistics.

Lave, J., Murtaugh, M., & de la Rocha, O. (1984). In B. Rogoff & J. Lave (Eds.), *Everyday cognition: Its development in social context* (pp. 67-94). Cambridge, MA: Harvard University Press.

Lesh, R., & Lamon, S. J. (1992). Assessing authentic mathematical performance. In R. Lesh & S. J. Lamon (Eds.), *Assessment of authentic performance in school mathematics* (pp. 17-62). Washington, DC: American Association for the Advancement of Science.

Lindquist, M. M. (1994). Lessons learned? In I. Gal & M. J. Schmitt (Eds.), *Proceedings: Conference on adult mathematical literacy* (pp. 91-98). Philadelphia: University of Pennsylvania, National Center on Adult Literacy.

Linn, R. L. (1990). Essentials of student assessment: From accountability to instructional aid. *Teachers College Record*, *91*(3), 422-436.

Lloyd, P., & Mikulecky, L. (in press). Numeracy skills and numeracy needs in workplace contexts. In I. Gal (Ed.), *Numeracy development: A guide for adult educators*. Cresskill, NJ: Hampton Press.

Lytle, S. L., Belzer, A., & Reumann, R. (1992). *Invitations to inquiry: Rethinking staff development in adult literacy education* (Technical Report No. TR92-2). Philadelphia: University of Pennsylvania, National Center on Adult Literacy.

Mathematical Sciences Education Board (MSEB). (1994). *Measuring what counts*. Washington, DC: Author.

Mayer, E. (Chair). (1992). *Putting general education to work. The key competencies report*. Australia: The Australian Education Council for Vocational Education, Employment and Training.

Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, *18*(3), 5-11.

Mikulecky, L., & d'Adamo-Weinstein, L. (1991). Evaluating workplace literacy programs. In M. C. Taylor, G. R. Lewe, & J. A. Draper (Eds.), *Basic skills for the workplace* (pp. 481-499). Toronto: Culture Concepts Inc.

National Council of Teachers of Mathematics (NCTM). (1995). *Assessment standards for school mathematics*. Reston, VA: Author.

National Council of Teachers of Mathematics (NCTM). (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.

Nunes, T., Schliemann, A. D., & Carraher, D. W. (1993). *Street mathematics and school mathematics*. New York: Cambridge University Press.

O'Neil, H. F., Allred, K., & Baker, E. L. (1992). *Measurement of workforce readiness competencies: Review of theoretical frameworks*. Los Angeles: University of California at Los Angeles, National Center for Research on Evaluation, Standards, and Student Testing.

Padak, N. D., & Padak, G. M. (1991). What works: Adult literacy program evaluation. *Journal of Reading*, *34*, 374-379.

Perkins, D. N., & Salomon, G. (1989). Are cognitive skills context bound? *Educational Researcher*, 18(1), 16-25.

Perry, M. (1991). Learning and transfer: Instructional conditions and conceptual change. *Cognitive Development*, *6*, 449-468.

Resnick, L. B. (1987). *Education and learning to think*. Washington, DC: National Academy of Sciences.

Schmitt, M. J. (1994). The ABE math standards project: Adapting the NCTM standards to adult education environments. In I. Gal & M. J. Schmitt (Eds.), *Proceedings: Conference on adult mathematical literacy* (pp. 99-105). Philadelphia: University of Pennsylvania, National Center on Adult Literacy.

Secretary of Labor's Commission on Achieving Necessary Skills (SCANS). (1991). *What work requires of schools: A SCANS report for America 2000*. Washington, DC: U.S. Government Printing Office.

Scribner, S., & Stevens, J. (1989). *Experimental studies on the relationship of school math and work math* (Technical Paper 4). New York: Teachers College, Columbia University, National Center on Education and Employment.

Shepherd, L. A. (1991). Psychometricians' beliefs about learning. *Educational Researcher, 20*(7), 2-16.

Shope, M. A. (in press). Conducting task analysis in the workplace. In I. Gal (Ed.). *Numeracy development: A guide for adult educators*. Cresskill, NJ: Hampton Press.

Stenmark, J. K. (Ed.). (1991). *Mathematics assessment: Myths, models, good questions, and practical suggestions*. Reston, VA: National Council of Teachers of Mathematics.

Sticht, T. G. (1990). *Testing and assessment in Adult Basic Education and English as a Second Language programs*. San Diego, CA: Applied Behavioral and Cognitive Sciences, Inc. (Available through the U.S. Department of Education, Division of Adult Education and Literacy, Washington, DC).

Sticht, T. G., & McDonald, B. A. (1993). *Automotive trades information processing skills: Mathematics*. Westerville, OH: Glencoe.

Tests of Adult Basic Education. (1987). *Examiner's manual*. Monterey, CA: CTB/McGraw Hill.

Venezky, R. L. (1992). *Matching literacy testing with social policy: What are the alternatives?* (Policy Brief PB92-1). Philadelphia: University of Pennsylvania, National Center on Adult Literacy.

Venezky, R. L., Bristow, P. S., & Sabatini, J. P. (1993). *When less is more: A comparative analysis for placing students in adult literacy classes* (Technical Report TR93-8). Philadelphia: University of Pennsylvania, National Center on Adult Literacy.

Venezky, R. L., Wagner, D. A., & Ciliberti, B. S. (Eds.). (1990). *Towards defining literacy*. Newark, DE: International Reading Association.

Webb, N. L. (1992). Assessment of students' knowledge of mathematics: Steps towards a theory. In D. A. Grouws (Ed.), *Handbook for research on mathematics teaching and learning* (pp. 661-683). New York: Macmillan.

Willis, S. (Ed). (1990). *Being numerate: What counts?* Hawthorn, Victoria: Australian Council for Educational Research.

Wolf, A. (1991). Assessing core skills: Wisdom or wild goose chase? *Cambridge Journal of Education*, *21*(2), 189-201.