

Presented at the
Literacy Policy Forum
Washington, D.C.
May 22, 1992

Matching Literacy Testing with Social Policy: What are the Alternatives?

Richard L. Venezky
University of Delaware
and National Center on Adult Literacy

OVERVIEW

The Departments of Education, Labor, and Health and Human Services now administer a variety of social programs that incorporate literacy training. Adult Basic Education (ABE), English as a Second Language (ESL), JOBS, workplace literacy, family literacy, library literacy, and correctional institution education are the primary examples of these programs, but others exist that either provide literacy instruction directly or that incorporate literacy as a component of another social service. Central to the proper administration of these programs is appropriate information on the needs and characteristics of the clients for these services, the quality of services actually delivered, and the impact of the services upon the individuals who receive them. Literacy tests are one of the most widely used (and controversial) means through which information on all three of these topics is obtained. Although tests are not and should not be the only means through which information is gathered, they are by tradition the primary procedure through which comparative analyses have been performed. From national surveys of literacy, such as the National Adult Literacy Survey (NALS) now underway, to locally developed screening and placement instruments, a melange of literacy tests now exists, providing data with widely varying degrees of reliability, validity, and comparability.

Literacy tests are one of the most widely used (and controversial) means through which information on clients is obtained.

NATIONAL CENTER ON ADULT LITERACY
UNIVERSITY OF PENNSYLVANIA
3910 CHESTNUT STREET
PHILADELPHIA, PA 19104-3111
TEL (215) 898-2100
FAX (215) 898-9804

POLICY BRIEF

Document No. PB92-1

This brief synthesis is an initial attempt to delineate the central issues in literacy testing and to define options that might provide an improved data base for social policy development. The plan of this paper is to attend first to basic issues in literacy testing: the perceived needs for literacy funding and the potential outcomes of literacy programs. Then various types of testing needs are discussed, along with options for satisfying these needs. Finally, the measures or scales used for reporting literacy levels are examined.

This paper is an attempt to delineate the central issues in literacy testing and to define options that might provide an improved data base for social policy development.

BASIC ISSUES

Several key assumptions underlie current literacy policies, some of which have not been subjected to sufficient scrutiny. Although some of these assumptions have been studied and others are the subjects of ongoing studies, many questions remain. The first assumption is that a large number of adults in America need and would profit from literacy training. Although few disagree with this assumption in its general form, little agreement exists on the number of adults who need assistance or on their physical and mental characteristics. One of the barriers to resolving this debate is the lack of differentiation for literacy needs. On an absolute level, almost everyone could benefit from further literacy instruction, even lawyers and doctors. For policy-making, however, the most important needs are those that will allow people to be self-sufficient economically, to manage their homes, and to be functioning members of their communities.

Little agreement exists on the number of adults who need assistance or on their physical and mental characteristics.

However, even within this domain, needs differ widely. The ESL needs of professional immigrants from Eastern Europe are vastly different from the basic literacy needs of some middle school dropouts, and these differ from the literacy needs of high school graduates who want entry to advanced technical skill training. The types of instruction that each of these populations need to qualify them for the same employment opportunities and the same levels of home and civic functioning have different resource and administrative demands. Effective policy requires more than a combined count of all who might profit from further literacy instruction.

The types of instruction needed for various populations have different resource and administrative demands.

A second assumption is that if adults could only be persuaded to attend adult basic education programs, or other types of literacy instruction, they would acquire the skills they are presumed to be lacking. This assumption is relatively untested. We know little about the effectiveness of different literacy programs and have limited experience in measuring literacy program performance (Darkenwald, 1986). Almost all program output measures are of academic or functional skills, yet most researchers agree that attitudes and beliefs also deserve attention. Even the roles of some types of programs are in doubt. Library tutoring programs, for example, usually provide services to adults at the lowest ability levels, using relatively untrained volunteers for tutors. Do we expect appreciable literacy skill gains from such programs or do we

We know little about the effectiveness of different literacy programs and have limited experience in measuring literacy program performance.

expect more confidence building and survival skill acquisition? To what degree should these programs prepare and encourage adults to attend ABE classes?

The third and most important assumption is that adults who acquire whatever is being taught in literacy programs will be better equipped for high-skilled jobs in the labor market, will function better as parents and home managers, and will participate more fully in civic and community affairs. Many literacy programs operate on an open entry, open exit basic, wherein students define their own goals. Although this policy may be successful in encouraging adults to attend programs they might not be comfortable in under different operating procedures, it does not ensure that adults will be striving for the levels of literacy that they need and could attain. Adults with low literacy levels tend to lack confidence in their own abilities to learn. Without assistance in defining what their needs might be, they may aim for far less than what is necessary for their own success.

More serious is the lack of basic understanding of the literacy demands of home and civic functioning. We assume, for example, that one goal of literacy instruction is to improve parenting, particularly for fostering the literacy development of children, yet we have only crude estimates of the amounts and types of reading, writing, and mathematics that are required to succeed at this task. Like most of the other goals of literacy instruction, parenting is dependent upon much more than literacy. In some cases, literacy is only a proxy for some complex of skills and may not be necessary, given compensating conditions. For example, many immigrant and refugee parents in America who are not literate in English are nevertheless quite effective in fostering their children's education (Caplan, Choy, & Whitmore, 1992).

Having programs wherein students define their own goals does not ensure that adults will be striving for the levels of literacy they need and can attain.

Like most of the other goals of literacy instruction, parenting is dependent upon much more than literacy.

THE ROLE OF TESTING

The primary reasons for literacy testing, from which data are regularly accessed by policymakers, are: (1) national and state population assessments (e.g., NALS), (2) individual skill assessments, (3) placement procedures, some of which involve testing, and (4) program evaluation.

Population assessments

National and state surveys of literacy involve two types of data collection: background information, generally gathered orally, on educational, economic, health, language, and civic activities and experiences; and skill abilities as measured by some nationally standardized instrument. Since the early 1970s, tests of functional abilities have been favored for population surveys of adult literacy. These tests tend to have a high face validity for adults because they draw their items from everyday literacy tasks that are familiar to

In most literacy testing, no systematic effort is made to assess writing or mathematical ability.

most English-speaking adults. The most recent population surveys, the Young Adult Literacy Survey (YALS) (Kirsch & Jungeblut, 1986) and the National Assessment of Literacy Survey (NALS) currently underway, were constructed around short answer responses to everyday reading tasks. Although some arithmetic ability is required in one set of tasks and some responses require writing, no systematic attempt is made to assess these skill domains. In addition, both skills are so embedded within reading contexts that it is not always possible to determine where an item difficulty rests.

In contrast, national surveys of elementary and secondary level literacy abilities are based upon basic reading, writing, and mathematical skill definition; e.g., the National Assessment of Educational Progress (NAEP). Although the YALS included a small group of items drawn from the NAEP reading assessment for linkage purposes, the two types of surveys differ sufficiently to make them noncomparable. That is, one can determine how well a young adult might do on the NAEP reading scale, based upon responses to the NAEP items in the YALS, but one cannot from these data equate scores on the NAEP reading assessment with scores on the YALS document, quantitative, or prose scales. This incompatibility between the NAEP reading assessment and the YALS (and NALS) impedes longitudinal analyses of functional literacy abilities. For example, reading scores from the NAEP high school reading tests cannot be compared to functional literacy scores from the same individuals, obtained in the years after leaving high school.

There is a basic incompatibility between the NAEP and the YALS/NALS assessments that impedes longitudinal analyses of functional literacy abilities.

Skills assessments

Most tests used for pre- and post-testing of adult skill abilities in literacy programs have been basic skills tests, most often developed from similar tests for children. A typical adult reading test, for example, will test various types of comprehension (e.g., literal, inferential, critical), plus vocabulary and perhaps structural analysis (e.g., prefixes, suffixes). These tests tend to report scores in grade-level equivalents, although other reporting procedures are possible. Grade-level equivalents are questionable scalings for elementary and secondary level assessment; for adults they are even more problematic (Sticht, 1988; Kirsch & Guthrie, 1977-78; Ryan & Furlong, 1975). A sixth grader who reads at a sixth-grade level and an adult who reads at a sixth-grade level usually have widely different reading abilities and require different forms of instruction, yet by grade-level measures they are classed as identical. In addition, grade-level designations for adults are misleading in that they assume the average experiential and background knowledge levels of children at the designated grade points. Although no widely accepted alternative scaling for adult abilities has yet been developed, the need for such a replacement is strong.

Grade-level designations for adults are misleading in that they assume the average experiential and background knowledge levels of children at the designated grade points.

If functional literacy tests were simply alternative means for tapping the same skills as those assessed by basic skills tests, the two could be equated. However, for adults in ABE programs, the correlation between the document scale of the Test of Adult Literacy Skills (TALS) (Simon & Schuster, 1990) and the reading comprehension scale of the Tests of Adult Basic Education (TABE) (CTB/McGraw-Hill, 1987) is less than .70. A regression analysis found that the best predictor of document scale performance on the TALS was the TABE Mathematical Concepts and Applications; the TABE Reading Comprehension scores also made a significant but far smaller contribution to the predicted scores (Sabatini, Venezky, & Bristow, 1992). Functional literacy tests like the TALS require not only general reading comprehension, but also problem solving abilities.

At present, functional literacy tests are not derived from theoretical models of skill ability. Instead, they are developed from matrices of text types and everyday literacy tasks. For example, one might decide that want ads in newspapers are an important text type (or genre) for assessment. Many different tasks might then be defined for such texts, based upon authentic use of advertisements. A want ad requesting a housecleaner and stating an initial salary of \$10.00 per hour and a phone number might be coupled with any of the following questions: (1) What phone number should you call to respond to the ad? (2) If you took the job and worked one day each week for four hours, how much would you earn in eight weeks? (3) If you worked at this job for a year and received positive reviews of your work from your employer, what salary would you expect for the next year if the average wage increase in your category were 6%?

These questions vary not only in complexity but also in the skills required. It would be misleading to claim that adults could or could not handle want ads on the basis of their answers to these questions. In this example, only one want ad is proposed for testing; others of greater or lesser complexity could also be developed. More importantly, since no analysis has been done of the skills required to do such tasks, the instructional implications of such assessments are not obvious. Considerable research remains to be done on the interaction between text and task difficulty and on the skills that are involved in functional literacy tasks.

The incompatibility between basic skills tests and national literacy surveys places a special burden on policymakers. The NALS results, which will be available within a year, will not be compatible with scores reported by literacy programs that use basic skills tests, and while NALS scores could be converted to a crude grade-level equivalency, this would have little validity in that most of the NALS tasks are not taught extensively in school. Furthermore, the scores would still not be any more comparable than math and reading scores that were mapped onto grade-level scales.

One option for resolving this incompatibility is to shift national adult literacy surveys to basic skills tests. Another option is to develop a stronger theoretical base for functional literacy tests, one that could allow instructional interpretations of test results, and use this type of test for both national and program assessments.

Functional literacy tests are not simply alternative means for tapping the same skills as those assessed by basic skills tests.

Functional literacy tests are derived from matrices of text types and everyday literacy tasks, not from theoretical models of skill ability.

The instructional implications of functional literacy assessments are not clear.

The incompatibility between basic skills tests and national literacy surveys places a special burden on policymakers.

Placement Tests

Many ABE programs currently use tests like the TABE for placement. At one site, for example, new students are first given the TABE locator, which requires a minimum of 35 minutes for administration. Based upon their locator scores (vocabulary and arithmetic), the students then receive four additional tests at a particular difficulty level. These tests require an additional 2 hours and 25 minutes of testing time. From the resulting scores, placements are then made into one of three levels of ABE classes or into a GED class. Total student testing time is about 3 hours; grading and other administration time for the staff might be 6 hours or more for each group of 20 students tested.

One problem with placement testing is that many adults who are entering literacy programs for the first time have poor test-taking skills. They also tend to have low self-confidence and associate testing with their generally negative feelings about formal schooling. These factors lead to low reliability for tests given prior to instruction. Tests given after several weeks of instruction might yield more reliable scores, but of course would not be useful for placement.

But given the small number of placement levels involved in adult programs, the need for three hours of placement testing is questionable. Furthermore, a study currently underway that is comparing placement predictions for the TABE locator, TABE reading, and TALS document and quantitative scales, has found no advantage for the full TABE reading battery or either of the TALS scales over the TABE vocabulary locator test (Venezky, Sabatini, & Bristow, in prep.). That is, the TABE vocabulary locator, which requires about 17 minutes for administration, predicted the actual placements for 125 adults in ABE and GED classes as well as or better than the TABE reading and vocabulary tests, either separately or combine, or either of the TALS tests.

Given these results and the highly negative impact of formal testing on adults returning to instructional programs, alternatives to lengthy formal placement testing need to be considered. One option is self-assessment. Techniques might be developed for assisting adults in assessing their own literacy abilities through sample texts and reading tasks. This might also be done through exposure to the materials used in the first two weeks of each placement level. A second option is a self-administered test that an adult might do at home, at work, or even in a library, with or without the administrative assistance of a family member or friend. Given that the adult must live with the consequences of the test, motivation for cheating would probably be quite low.

Current placement testing methods may take up to three hours of student time to place them into just four levels of instruction.

A student's poor test-taking skills may lead to low reliability for tests given prior to instruction.

Alternatives to lengthy formal placement testing need to be considered.

Program Evaluation

One reason for administering formal test at the beginning of an instructional program is to meet reporting requirements for student progress and program outcomes. Typically, basic skills gain scores are reported. There are many problems with this, and the results, as currently reported, have relatively limited use for policymakers. First, the pre-test scores are based on a test that is enrolled in a program. For many adults this is the first encounter with formal testing since dropping out of high school (or grade school). Scores tend, therefore, to be artificially low: adults lack test-taking skills, have low self-confidence, and lack practice with some literacy skills. After even a few weeks in class, scores may rise significantly due simply to exposure to test formats, printed vocabulary, and the like. Reports from New York City ABE programs show relatively large first year gains (9.9 months average), but less than half of this amount for a second year in a program (3.3 months average) and even less for those who remain for a third year (3.0 months average) (Literacy Assistance Center, 1991).

Second, almost all test score reporting is for reading comprehension, yet most ABE courses stress a variety of skills, including basic mathematics, reading vocabulary and comprehension, and writing, as well as specific survival skills (applying for a job, interacting with your child's school, etc.). This means an incomplete match between course content and test content. Part of the problem with current practices is that ABE programs are attempting to serve two distinct goals. One is to prepare adults who have not had the advantages of formal education or did not make full use of them to progress in a formal educational system. This path might lead to a GED and, for some, to community college or a four-year institution. The second goal is to give adults practical skills they need for self-esteem, work, home, and civic needs. In this path, less attention is given to academic skills, which means less time to abstractions, theories, and the like. (For a discussion of functional vs. general or academic literacy instruction, see Sticht, Armstrong, Hickey, & Caylor, 1987).

For evaluation of progress, several procedures should be considered. For academic programs, what is important is to know how many eventually reach their academic goals (GED, college, etc.) and in how much time. This information would allow policymakers to determine the overall effectiveness of the system as well as the relative effectiveness of different programs in helping students reach their academic goals. The GED examination and college entry requirements generally serve as checks on actual skill attainment, although with open admissions, it may also be important to track progress within post-secondary education.

Students enrolled in courses that stress practical skills could be assessed at the termination of their instruction with functional literacy tests. To measure actual progress, post-test scores might be compared to scores obtained after a month or so of enrollment. But we have little agreement on what average gains to expect for

Pre-test scores tend to be artificially low; gain scores may then be artificially high.

There is an incomplete match between course content and test content.

Like national surveys, program evaluations often ignore writing and mathematics.

different entry-level abilities and instructional regimens. And gain scores do not translate easily into policy. A gain of 50 points on the TALS document scale, for example, might be a positive sign for someone who entered a program with low reading ability and re-enrolled for the next higher level of instruction. For a person who entered with the same low level of ability and planned not to proceed to the next level, this may not be a positive sign, particularly if the exit level of functioning were still too low to ensure adequate handling of print material for home, work, and civic life.

An alternative is to report a combination of exit scores and enrollment statistics, including the percentage of students who enroll in the next higher level course, when one is available. This approach requires further work, however, on matching functional literacy scores with levels of functioning in society. What policymakers need to know, for those who enroll in literacy courses, is how many stay long enough to reach desired levels of functioning. At present many students who enroll do so on a casual basis, attending sporadically and often leaving without formally withdrawing.

Another complication to consider is that what a functional literacy test or a basic skills test assesses may not be what particular courses teach or what particular students want to learn. For efficiency we would like to have assessment instruments that could inform instruction as well as policy. However, this may not be possible. Intelligent policy requires data comparable across time and geographic region. Furthermore, policy related results must be interpretable in relation to societal goals. Students who enroll in literacy courses to learn how to read to their children may not care at the end of instruction what their reading levels are relative to national standards. However, policymakers do. One option is to divide students according to their expressed goals into general literacy students and specialized skill (or functional literacy) students. Then, only the former would be tested with nationally standardized instruments. The latter group might not be assessed at all or be assessed through alternative means (e.g., Lytle & Wolfe, 1989). A second option is to require that all students be tested on exit with nationally standardized instruments as a condition of enrollment. This by itself, however, will not solve the problem created by those who drop out before exit testing.

Finally, we should note that so far we have discussed primarily the assessment of cognitive skills. Attitudes and beliefs are also important for literacy development, yet we have limited instrumentation for valid and reliable assessment in this domain. One task that remains is to synthesize the literature on attitudes and beliefs to determine what, if anything, needs to be assessed and what the complexities are in doing this assessment.

What policymakers need to know, for those who enroll in literacy courses, is how many stay long enough to reach desired levels of functioning.

We have limited instrumentation for valid and reliable assessment of attitudes and beliefs.

Score Reporting

As stated earlier, grade-level reporting for test scores needs to be reconsidered. If it is an adequate scale for reporting adult abilities, then it needs to be justified both empirically and logically.

Grade-level reporting for test scores needs to be reconsidered.

If not, then a scale (or scales) that applies more directly to adult functioning levels needs to be developed. Such a scale might be developed around materials and tasks that represent different levels of functioning in work, home, and civic life. Alternatively, the scale might be built around knowledge of cognitive processing—the skills and strategies required for different levels of ability. A sufficient base already exists for initiating work on the latter. Studies of word identification reaction times, automaticity in decoding, reading flexibility, and metacognition, in particular, provide insights into some of the cognitive skills required for reading (e.g., Vellutino & Scanlon, 1987; Perfetti, 1985; Stanovich, 1982). A targeted research program could pursue these and other leads with adults, not just for reading but also for writing.

The differences between these two approaches to the development of an alternative literacy scale for adults should be subjected to critical examination. The materials and tasks approach will be easier to relate to home, work, and civic functioning; the cognitive processing approach will be easier to relate to elementary and secondary assessment and to instruction. The former will need to be re-evaluated as the literacy demands of everyday life change while the latter will require an additional step to relate it to everyday literacy demands.

The materials and tasks approach will be easier to relate to home, work, and civic functioning; the cognitive processing approach will be easier to relate to elementary and secondary assessment and instruction.

CONCLUSIONS

We have attended in this synthesis to placement, achievement, program, and national survey needs in adult literacy. Among the issues we feel need to be explored in the development of more informative literacy testing are:

- the fundamental differences between skill-based tests and functional literacy tests;
- the varying goals and needs of different types of literacy programs;
- the lack of empirical support for grade-level score designations for adults;
- the limited ability discrimination required for placement into literacy courses;
- the difficulties in interpreting gain scores; and
- the limited attention currently given to testing of writing and mathematics.

What we have not discussed in depth is testing the basic hypothesis behind literacy funding—that program impacts are sufficient to justify the types and amounts of funding currently allocated. This is a complicated question that can only be answered after the issues raised here are resolved. Without an accurate measure of individual progress that links to national trend data, and without linking both to functional levels, little can be decided about long-term impacts of programs. Even with these measures, other factors would need to be considered. The mix of available jobs, how well employers utilize workers with improved skill levels, social mobility, and all of the factors that impinge upon family stability would need to be considered.

For example, adults may become more skilled through particular literacy programs, but because of the strength of the dollar abroad or other factors, jobs may not be available for many of the students wanting to enter the labor market. Therefore, by some evaluation measures these literacy programs would be deemed unsuccessful. Alternatively, workers may, through workplace programs, acquire new abilities but rigid work organization may not allow these new abilities to be utilized and consequently program impacts on salary, performance, and safety might be attenuated. Similar arguments can be made about long-term impacts upon home and civic functioning. Literacy is not the only factor that determines how well a person functions in these areas. Multiple indicators will probably be needed for all impact evaluations, with careful modeling of the influence of uncontrolled factors.

Without an accurate measure of individual progress that links to national trend data, and without linking both to functional levels, little can be decided about long-term impacts of programs.

Literacy is not the only factor that determines how well a worker functions.

References

- Caplan, N., Choy, M.H., and Whitmore, J.K. (1992). Indochinese refugee families and academic achievement. *Scientific American*, 266 (No. 2), 36-42.
- CTB/McGraw-Hill (1987). *TABE: Tests of adult basic education*. Monterey, California: Author.
- Darkenwald, G.G. (1986). *Adult literacy education: A review of the research and priorities for future inquiry*. N.Y.: Author.
- Kirsch, I., and Guthrie, J.T. (1977-78). The concept and measurement of functional literacy. *Reading Research Quarterly*, 13 (No. 4), 485-507.
- Kirsch, I., and Jungeblut, A. (1986). *Literacy: Profiles of America's young adults*. Report No. 16-PL-01. Princeton, NJ: National Assessment of Educational Progress.
- Literacy Assistance Center (1991). *Analysis of New York City's 1989-1990 adult literacy data base*. N.Y.: Author.
- Lytle, S.L., and Wolfe, M. (1989). *Adult literacy education: Program evaluation and learner assessment*. ERIC Clearinghouse on Adult, Career, and Vocational Education, Center on Education and Training for Employment, The Ohio State University, Columbus, Ohio.
- Perfetti, C. (1985). *Reading ability*. N.Y.: Oxford University Press.

- Ryan, T.A., and Furlong, W. (1975). Literacy programs in industry, the armed forces, and penal institutions. In J.B. Carroll and J.S. Chall (Eds.), *Toward a literate society*. N.Y.: McGraw-Hill.
- Sabatini, J., Venezky, R.L., and Bristow, P. (1992). *Probing the relationship between functional and general adult literacy tests*. Technical Report 92-2.1C. National Center on Adult Literacy, University of Pennsylvania, Philadelphia, PA.
- Simon & Schuster (1990). *ETS tests of applied literacy skills*. N.Y.: Author.
- Stanovich, K. (1982). Word recognition skill and reading ability. In M. Singer (Ed.), *Competent reader, disabled reader: Research and application*. Hillsdale, NJ: Erlbaum.
- Sticht, T.G. (1988). Adult literacy education. In E. Rothkopf (Ed.), *Review of research in education*, Vol. 15. Washington, D.C.: American Educational Research Association.
- Sticht, T.G., Armstrong, W.B., Hickey, D.T., and Caylor, J.S. (1987). *Cast-off youth: Policy and training methods from the military experience*. N.Y.: Praeger.
- Vellutino, F.R., and Scanlon, D.M. (1987). Phonological coding, phonological awareness, and reading ability: Evidence from a longitudinal and experimental study. *Merrill-Palmer Quarterly*, 33, 321-363.
- Venezky, R.L., Sabatini, J., and Bristow, P. (in prep.) *Prediction of class placement from different functional and general adult literacy tests*. University of Delaware, Newark, Delaware.

This work was supported by a grant from the National Center on Adult Literacy at the University of Pennsylvania, which is part of the Educational Research and Development Center Program (grant No. R117Q00003) as administered by the Office of Educational Research and Improvement, U.S. Department of Education, with federal co-funding by 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 funding agencies.

For more information on NCAL activities and publications, contact:

Sandra Stewart
Manager of Dissemination
National Center on Adult Literacy
University of Pennsylvania
3910 Chestnut Street
Philadelphia, PA 19104-3111

Phone: 215-898-2100
FAX: 215-898-9804