



National Center for the Study of Adult Learning and Literacy

**THE OUTCOMES AND IMPACTS
OF ADULT LITERACY EDUCATION
IN THE UNITED STATES**

by

Hal Beder
Rutgers University, Graduate School of Education

NCSALL Reports #6

January 1999

The National Center for the Study of Adult Learning and Literacy
Harvard Graduate School of Education
101 Nichols House, Appian Way
Cambridge, MA 02138

The work reported herein is supported by the Educational Research and Development Centers Program, Award Number R309B60002, as administered by the Office of Educational Research and Improvement/National Institute on Postsecondary Education, Libraries, and Lifelong Learning, U.S. Department of Education, through contract to Harvard University. However, the contents do not necessarily represent the positions or policies of the National Institute on Postsecondary Education, Libraries, and Lifelong Learning; the Office of Educational Research and Improvement; or the U.S. Department of Education, and you should not assume endorsement by the Federal Government.

TABLE OF CONTENTS

Executive Summary	4
Preface	6
Chapter 1: Introduction	7
Framing the Issue	8
Accountability	8
Program Improvement and Planning	8
Goals	9
Measurement	9
Capacity	11
Judgment	12
Chapter 2: Methodology	13
Goals	13
Research Strategy	13
Chapter 3: Analysis	15
National Studies	15
National Evaluation of Adult Education Programs (NEAEP), 1990-1994	21
The 1980 National Evaluation: An Assessment of the State-Administered Program of the Adult Education Act	23
A Longitudinal Evaluation of the Adult Basic Education Program, 1973	26
The 1968 Evaluation	29
State Studies	29
The Washington Workforce Training Study	31
The Tennessee Longitudinal Study	34
The California Adult Learner Progress Evaluation (CALPEP)	34
The New Jersey Study	36
The Utah Study	39
The Maryland Study	41
The 1980 Tennessee Study	42
The Ohio Study	43
The Wisconsin Study	45
Welfare Studies	48
The California GAIN Study	48
The Texas JOBS Program Evaluation	51
Steps to Success	53
Workplace Literacy	55
Mikulecky and Lloyd	55
Manufacturing and Financial Services in the Chicago Area	57
Workplace Literacy across the Three Phases of Textile Manufacturing	58
Wisconsin Workplace Partnership Training Program	59

Central Labor Council and the Consortium for Worker Education	61
Family Literacy	63
The National Even Start Program	63
The National Center for Family Literacy	66
Chapter 4: Conclusions	68
How Effective is the Adult Literacy Education Program?	68
Employment	70
Better Jobs for Those Who Are Employed	71
Increased Earnings	71
Continued Education	71
Termination/Reduction of Welfare Dependence	71
Self-Reported Gains in Basic Skills	72
Tested Basic Skills Learning Gain	72
GED Acquisition	72
Self-image	73
Impact on Children's Education	73
Attainment of Personal Goals	73
Conceptual, Design, and Methodological Problems	74
Inaccurate and Incomplete Data	75
Self-report Data	75
Lack of Controls	76
Valid and Reliable Tests	76
Poor Quality Reports	76
Lack of Standards	77
Implications for Policy	77
Finding the Answer	77
Economic Impact	77
Basic Skills	78
Self-image	78
Children's Education	78
Personal Goal Attainment	79
Accountability Standards and Judgments	79
Recommendations	79
An Effective Outcome and Impact Reporting System	80
A National Longitudinal Assessment	
Improved State and Local Outcome Studies	

References

EXECUTIVE SUMMARY

This study investigated the outcomes and impacts of adult literacy education through a qualitative assessment of the outcomes and impacts research conducted since the late 1960s. Outcomes are the changes in learners that occur as a result of their participation in adult literacy education. Impacts are the changes that occur in the family, community, and larger society as a consequence of participation.

The goals of the study were to make reasoned inferences about the effectiveness of adult literacy education in the United States; to identify common conceptual, design, and methodological problems inherent in the outcome studies conducted; to raise and discuss issues for policy; and to make recommendations for the design and conduct of future outcomes studies.

Through a comprehensive literature search, approximately 115 outcomes and impacts studies were identified. All were obtained in either hard copy or microfiche and the 68 of those that were found to have an outcomes component were abstracted. Each study was then evaluated according to the following criteria: the study included an outcome/impact component; the report was adequately documented with respect to design and methods; there was an adequate number of cases; the sampling plan was adequate; data collection procedures were adequate (i.e., were not tainted by substantial attrition or biased by other factors); objective measures, rather than self-report, were used to measure outcomes; measures, especially tests, were valid and reliable; the research design included a control or comparison group; and inferences logically followed from the design and data.

Based on this evaluation, 23 studies were selected as being the most credible, and case studies were prepared for each. Studies are presented in five categories: national, state-level, welfare, family literacy, and workplace literacy. From the 23 studies, inferences about program effectiveness are made.

Program Effectiveness

The 23 case studies represented evidence rather than proof of impact, and, like evidence in a trial, their findings were weighed in order to reach reasonable conclusions. Weighing had two dimensions. The first was the extent to which the various studies converged or diverged in respect to their findings on specific outcome/impact variables. Consensus across studies pointed toward effectiveness/ineffectiveness, while lack of consensus suggested an inconclusive resolution. The second dimension was the credibility of the individual studies. When arriving at conclusions, more credible studies were weighed more heavily than less credible studies.

The conclusions set forth are deemed to be reasonable inferences from the findings reported in the case studies. They do not represent proof. Indeed, it is unlikely that any conceivable study or studies could arrive at certainty. Table 1 presents the data used for this analysis.

In interpreting Table 1 and the conclusions made from it, three caveats are in order. First, the variables included are those studied by a sufficiently large number of studies to enable reasonable conclusions. However, variable definitions and their units of measure vary among studies. In some studies, for example, learning gain is measured by the CASAS, while in others the TALS or TABE are used. Second, if a given study reported a gain, the gain is listed as positive (y) in the table irrespective of the size of the gain or the quality of the study's methodology. In some cases the gains reported as positive are quite small, and in some cases the limitations of the study render claims of gains suspect. Third, the totals are aggregates of studies conducted at different times and on different populations of adult literacy learners, welfare clients and employees being examples. Drawing conclusions from such aggregates presumes that doing so is both valid and meaningful.

From the case studies, as summarized in Table 1, the following conclusions were made about the effectiveness of the adult literacy education program in the United States:

1. In general, it is likely that participants in adult literacy education receive gains in employment.
2. In general, participants in adult literacy education believe their jobs improve over time. However, there is insufficient evidence to conclude that participation in adult literacy education causes job improvement.
3. In general, it is likely that participation in adult literacy education results in earnings gain.
4. In general, adult literacy education has a positive influence on participants' continued education.
5. Although the evidence suggests that participants in welfare-sponsored (e.g. JOBS Program) adult literacy education do experience a reduction in welfare dependence, the evidence is inconclusive as to whether adult literacy education in general reduces welfare dependence for participants.
6. Learners perceive that participation in adult literacy education improves their skills in reading, writing, and mathematics.
7. As measured by tests, the evidence is insufficient to determine whether or not participants in adult literacy education gain in basic skills.
8. In general, adult literacy education provides gains in GED acquisition for participants entering at the adult secondary (ASE) level.
9. Participation in adult literacy has a positive impact on learners' self-image.
10. According to learners' self-reports, participation in adult literacy education has a positive impact on parents' involvement in their children's education.
11. Learners perceive that their personal goals are achieved through participation in adult literacy education.

In the final chapter, conceptual, design, and methodological problems inherent in the studies are discussed and implications for policy are presented, including recommendations for:

1. relevant and measurable outcome standards and a feasible impact reporting system;
2. a comprehensive national longitudinal evaluation measuring long-term impact; and
3. systematic funding and improvement of state and local outcome studies.

PREFACE

This report has four chapters. The first, the Introduction, frames the issues that provide focus to the report. The second describes the study's methodology. The third chapter, which is by far the longest, presents case studies of 23 studies of outcomes and impacts. The final chapter presents conclusions, implications, and recommendations.

The case studies presented in Chapter 3 represent the data upon which the study's conclusions are based. While from a researcher's perspective it is critical that they be included in the report, it is recognized that this material may be of more interest to researchers than to many policy makers and practitioners. For those who are less interested in the case studies, it is recommended that Chapter 3 be read last.

THE OUTCOMES AND IMPACTS OF ADULT LITERACY EDUCATION IN THE UNITED STATES

CHAPTER 1: INTRODUCTION

FRAMING THE ISSUE

This paper examines the outcomes and impacts of adult literacy education in the United States. Before this examination can begin in earnest, however, it is important to frame the issue. What are outcomes and impacts? Why is it important to examine them?

Outcomes are the changes that take place in learners as a result of their participation in adult literacy education. Outcomes imply cause and effect—participation in adult literacy education is the cause; measurable changes in knowledge, skills, attitudes, and behavior are the effects. Impacts are the changes in the family, community, and society in general that result from participation in adult literacy education, and, as with outcomes, they too imply cause and effect. As Brizius and Campbell (1991) note, assessments based on outcomes and impacts are distinct from assessments based on program processes, inputs, and outputs. Assessing inputs, for example, often entails detailed descriptions of learners' characteristics to determine if programs are serving adequate numbers of learners and the kinds of learners programs intend to serve. Similarly, an input assessment might examine teachers' characteristics, such as their years of experience, their attitudes, or their levels of certification, in order to assess whether the teaching force is competent. An assessment of output analyzes the products of adult literacy education. Typical output variables include the number of learners served during a given period, how learners were instructed, and the program retention rate.

Many of the evaluation studies that have been conducted in adult literacy education in the past 30 years have primarily, or exclusively, focused on inputs and outputs rather than on outcomes and impacts, and there are several reasons for this. First, for the most part, the regulations and standards that programs have been required to meet in order to acquire state and federal funding have been based on inputs and outputs. Thus programs routinely gather and report input and output data, although the accuracy of these data has sometimes been questioned (Condelli & Kutner, 1997). Secondly, input and output data are relatively easier to collect than outcome and impact data. Although input data can be routinely collected during the learner intake process and output data can be collected from program records such as teachers' attendance reports, gaining accurate and meaningful outcome and impact data generally requires special evaluation research, which requires resources local programs lack and states are reluctant to allocate.

Although input and output assessments are often useful, clearly outcome and impact assessments represent the "bottom line" with respect to determining the effectiveness of adult literacy education because, even when inputs and outputs are adequate and impressive, there is no guarantee that learners have actually learned and that society has reaped the benefits policy makers expect. Credible measurement of the outcomes and impacts of adult literacy education is critically important for at least two reasons: program accountability and program planning and improvement.

ACCOUNTABILITY

As Merrifield (1998) argues in the companion piece to this report, which addresses accountability, in the past decade accountability has emerged as a critical concern of policy makers, particularly those policy makers who exercise control over resource allocations. Indeed, accountability has been a concern for elementary and secondary education, as witnessed by the movement to develop accountability standards led, in part, by the National Council of Teachers of Mathematics. It has also been a major concern for higher education, public health, and social work.

In a report that focuses on accountability and public policy, Brizius and Campbell (1991) define accountability as follows:

Performance accountability is a means of judging policies and programs by measuring their outcomes or results against agreed upon standards. A performance accountability system provides the framework for measuring outcomes—not merely processes or workloads. (p. 5)

As Brizius and Campbell suggest, accountability must be based on outcomes rather than on a mere reporting of inputs, outputs, or descriptions of program processes. The reason is simple. For the purpose of accountability, an adequate judgment of a policy, program, or system of programs must be based on the cause and effect relationship between what the program or system does and the benefits to individuals and society it produces. This relationship cannot be determined unless the actual benefits (that is, outcomes and impacts) to individuals and society are measured and are measured in a way that allows decision-makers to logically infer cause and effect.

Any discussion of accountability raises the questions of accountability to whom and for what. In adult literacy education, these questions have by no means been resolved. Many claim that adult literacy education should be primarily accountable to learners since learners are the clients of adult literacy education. Further, it is argued that, since participation in adult literacy education is voluntary, unless learners are able to meet their own goals, they will be reluctant to participate, and if they do happen to participate, they will quickly drop out.

Others claim that adult literacy education should be accountable to society in general since society in general finances the program through tax dollars. Given our political system, Congress is the surrogate for society in general at the national level, and state legislatures are the surrogates at the state level. From a practical point of view, Congress and state legislatures are the bodies that allocate resources for adult literacy. Since with the purse goes power, the will of these legislative bodies can hardly be ignored. Debate that revolved around the Careers Act of 1996, which was not passed, and HR 1385, which was passed by the House of Representatives in the spring of 1997, suggests that Congress is increasingly conceiving adult literacy education to be part of the nation's workforce readiness system. To this extent, accountability has increasingly been framed in human capital terms. Can adult literacy education be accountable to adult learners and to society as represented by Congress and state legislatures? If the answer is no, which locus of accountability should prevail?

PROGRAM IMPROVEMENT AND PLANNING

Outcome assessment is a critical tool for program planning and improvement, for, quite obviously, weaknesses cannot be corrected and strengths cannot be capitalized upon unless they are systematically identified. When the literature on outcome assessment is considered, however, it must be concluded that outcome assessment has not been a widespread and systematic strategy for program planning and policy formation in adult literacy education. Since the inception of the Adult Education Act, there have been but a handful of credible state-level outcome assessments in adult literacy education, and, although the U.S. Department of Education has commissioned three national assessments of the adult literacy education system, these evaluations have focused primarily on input and output description, and the outcome data in each of them were so seriously flawed that sound inferences have been impossible. Systematic evaluation at the local level is rare indeed. As this report will demonstrate, the relatively few formal outcome assessments that have been conducted at the state and national levels are of such questionable validity that they are nearly useless for program planning and policy formation.

Demonstrating accountability and use of outcome assessment in program planning and improvement has been particularly difficult in adult literacy education for several reasons that will be discussed here.

GOALS

A classic evaluation begins with a delineation of program or system goals. Next, the goals are broken down into a series of objectives, and the objectives are “operationalized” into a set of variables for measurement. One of the problems inherent with outcome assessments in adult literacy education is that there is simply no consensus regarding what adult literacy is or what its goals should be. While some believe that adult literacy should be defined as the acquisition of a set of generalized skills revolving around reading, writing, and computation, or, as with the National Adult Literacy Survey (Kirsch, Jungeblut, Jenkins, & Kolstad, 1993), prose literacy, document literacy, and quantitative literacy, others argue that literacy cannot be defined outside the context in which it is used. Accepting the context-based definition, there are multiple literacies rather than a single literacy. Perhaps because it has been so difficult to operationalize the construct of adult literacy education for measurement, there has been a tendency to select more global goals as the basis for outcomes and impacts in accountability and program improvement. For example, in a policy paper commissioned by the U.S. Department of Education’s Office of Vocational and Adult Education, Condelli and Kutner (1997) identify seven relatively global outcome accountability measures: economic impact, credentials, learning gains, family impact, further education and training, community impact, and customer satisfaction. Except for learning gains and perhaps credentials, none of these goals has to do *directly* with the skills and knowledge expected from adult literacy education.

A focus on global goals raises the issue of what one can realistically expect adult literacy education to achieve. Adult literacy education is an educational program. Thus it is obviously reasonable to expect learners to learn, that is, to acquire knowledge, skills, and new perspectives as a consequence of their participation. But to what extent can adult literacy education be expected to produce outcomes that are at best indirectly related to what is learned in an adult literacy program? For example, since there are a multiplicity of variables that affect whether a person who completes adult literacy instruction will gain employment, and since many have to do with the state of the economy and the life situation of the individual, can adult literacy education reasonably be held accountable for its graduates’ employment status? If the answer is yes, given the other potent variables that affect learners’ economic gain, how much gain should be expected purely as a consequence of participation in adult literacy education?

MEASUREMENT

It is axiomatic to say that an outcome variable cannot be used for the purposes of outcome assessment unless it can be adequately measured. Yet measurement in adult literacy education has proved to be quite problematic. Tested learning gain is a case in point. Tested learning gain is considered to be an important outcome measure in this study because it has been employed in many studies of outcomes and impacts and because it is presumed by many to be a direct measure of instructional gain. Yet the measurement problems associated with tested learning gain are substantial.

First of all, there is controversy over whether learning gain can be adequately measured through standardized testing. For example, Fingeret and Drennon (1997) suggest that, because the shared meanings learners ascribe to symbols vary in different cultural contexts, literacy should be considered as practices that differ as context changes. If the very definition of literacy varies substantially with context, it would be difficult and probably impossible to measure something so specific with a generalized standardized test. Then there is the issue of test sensitivity. One of the contradictions in outcome measurement is that detailed qualitative studies such as those conducted by Fingeret and Danin (1991) and Fingeret (1985) show that learners do report literacy gains that are important to their lives, whereas studies that use standardized tests such as the TABE tend to show small, and in some cases no, gains. The reason may be that many of the personally important gains learners achieve are too small to be recorded on standardized tests. This may be especially true for beginning readers. For example, Heath (1983) notes the important impact of being able to write a simple memory list or note to one’s children for adults who could not previously perform these tasks. Yet despite the impact on learners’ lives, these gains would probably not register on most standardized tests.

Because there are multiple definitions of adult literacy education, and because under the Adult Education Act states have great latitude in establishing testing policy, there is no single standardized test that is appropriate or in use. Indeed, as the National Evaluation of Adult Education Programs found (Development Associates, 1992), of those programs that test regularly, 68 percent used the Test of Adult Basic Education (TABE), 12 percent used the Slosson Oral Reading Test (SORT), 21 percent used the Adult Basic Learning Examination (ABLE), 20 percent used the Wide Range Achievement Test (WRAT), 14 percent used the Comprehensive Adult Student Assessment System (CASAS), and 31 percent used locally developed tests. Over one third of the programs did not test at all at student intake. Clearly the lack of a common measure for tested learning gains makes accountability for learning gain difficult, especially at the national program level.

Just as serious as the lack of a common measure is the appropriateness of the measures available. For a test of learning gain to be appropriate, it must reflect what is taught in instruction. Yet what is taught in instruction can vary widely among programs and states. An example that highlights the problem is the fact that some states require all programs to pretest and posttest using a standardized test, the TABE being the most common. Although the TABE may be appropriate for those programs that use a general, basic skills approach to literacy, it is not appropriate for those programs that gear instruction to a particular context. Workplace literacy, which focuses instruction on the tasks learners use in the workplace, is an excellent example of a contextualized approach. Because workplace literacy instruction is not geared to the specific skills the TABE measures, the TABE is likely to underestimate literacy growth, and when basic skills-oriented programs are compared to workplace literacy programs on the TABE, workplace literacy is likely to score lower regardless of the quality of instruction.

Lack of adequate controls in measurement is another serious problem. A commonly used measure of economic gain, for example, is increased income for employed learners who have completed the program. Increases in income, however, are affected by many powerful factors that have nothing to do with literacy acquisition—factors such as the rate of inflation, increases in the minimum wage, the strength of the economy, and job tenure. When increases in income are measured by pre-program and post-program measures, or more simply by a post-program design, it is simply impossible to infer that the increase (or decrease, for that matter) was caused by participation in adult literacy education. The only way causality can be inferred is through the comparison of those who completed adult literacy education with a group of like individuals who did not participate in adult literacy education. Then the difference between the groups with respect to income gained can be inferred to have been caused by participation. The point is that, unless it is known with confidence that participation in adult literacy education caused a particular outcome, such as the achievement of increased income, little is known, and if little is known, how can reasonable policy be made? For many reasons that will be discussed later, however, it is so difficult to create adequate meaningful comparison groups in the real world of adult literacy education that few outcome studies have ever done it.

Measures that rely on learners' self-reports also raise an issue. There is nothing inherently wrong with self-report measures. Indeed, researchers often have no alternative but to use self-report. When learners are asked to self-report on items that are essentially objective, such as their gender and age, it can be assumed that truthful responses are being given. However, when learners are asked to self-report on questions for which some answers are socially acceptable and other answers are not, response bias becomes a concern. Response bias is also a concern for questions that require a long period of recall, questions that ask for information the learner might not know "off hand," and questions that are vague.

Asking learners to self-report on their reasons for dropping out is a case in point. Dropping out of school carries with it a social stigma. Thus, when asked this question, there is the potential for learners to respond with socially acceptable answers, such as sickness, lack of transportation, etc., and avoid socially unacceptable responses, such as lack of learning progress or conflicts with teachers and/or other students. Because self-report measures are susceptible to response bias in many cases, objective measures are usually preferable, and outcome studies that rely extensively or exclusively on self-report must be regarded with a degree of suspicion.

CAPACITY

The issue of capacity has two dimensions with respect to outcome assessment: the extent to which programs have the capacity to systematically gather the data necessary for outcome-based evaluation and the extent to which programs have the capacity to achieve goals established for them as policy.

From 1990 to 1994, the National Evaluation of Adult Education Programs (NEAEP) conducted an assessment of the federal adult literacy education program. In the end, the evaluation proved to be as much an assessment of the system's capacity to provide useful outcome information as it did an evaluation of the system itself. Excerpts from NEAEP's Executive Summary (Young et al., 1994b) demonstrate this point:

For inclusion in the evaluation, each program had to receive financial support through the basic grants provisions of the Adult Education Act. However, there is no available list of grantees, and funding practices and definitions vary among the states. (p. 6)

Information at the state level on local programs varies considerably in content and quality. Some programs did not have information on the composition of staff or the nature of instruction provided at different sites. Nor did many programs have any precise idea of the number of adults newly enrolled each year or the number of different individuals enrolled at any given time or over the period of a program year. (p. 7)

Key personnel and the location of instructional sites changed during the course of the study in many projects. Within the first six months of data collection, for example, 16 percent of program directors trained in the requirements of the study had departed, sometimes because their jobs had vanished. (p. 7)

The experience of the NEAEP and others who have attempted to conduct outcome evaluations partially or totally for the purpose of accountability demonstrate the following problems with the system's capacity to generate accurate outcome data:

- Unlike elementary, secondary, and higher education, where learners enroll and complete at discrete times, adult education programs tend to maintain open enrollment and learners attend irregularly. This makes it very difficult to collect complete data on learners at specific time intervals.
- The dropout rate for adult literacy education is quite high, approximately 60 percent after 12 weeks (Young et al., 1994b). Furthermore, it is generally not known if non-persisters represent completions of learners who have met their goals, or whether non-persistence represents dropout for other reasons. Because non-completers are difficult to follow up, "experimental mortality" typically interjects a strong source of bias into learner data.
- Program staff lack expertise in data collection, particularly testing. Furthermore, programs typically lack the resources to insure accurate data collection. Thus a significant amount of the data collected at the program level is of questionable accuracy.

Program capacity has a second dimension. If adult literacy education is to be expected to meet outcome standards, then the system must have the capacity to meet outcome expectations. However, because of low per-student expenditures, a reliance on part-time staff, fragmentation of service, and structural marginality in comparison with elementary, secondary, and higher education, the adult literacy education system's capacity to meet the expectations of either learners or the public is low (Beder, 1996). Indeed, as the NEAEP found, mean expenditures per student per year are about \$258, and over 80 percent of the teachers of adult literacy work part time (Development Associates, 1992). To the extent that system capacity is low, achievement of outcome accountability must include a greater allocation of resources to program development as well as increases in the quantity and quality of outcome evaluation and the development of pertinent outcome standards.

JUDGMENT

As with any form of evaluation, outcome assessment requires judgment regarding whether an obtained outcome is exemplary, adequate, or sub-standard. A comparative approach to judgment entails comparing the results of one program or system with others. An external standards approach entails judgment with respect to established standards. Unfortunately, in adult literacy education, there has been little basis for judgment. First, as this report will show, because the outcome evaluations that have been conducted vary so widely in methods and measurement, and because their internal and external validity has been almost universally questionable, there is little basis for comparative judgment. Second, specific external standards have not as yet been established. For example, learning gain is a fundamental outcome that might be expected from participation in adult literacy education. Yet we do not know what constitutes adequate learning gain on any particular measure.

Unless reasonable judgments can be made from outcome assessment, there is no meaningful way to separate good practice from poor practice, and the opportunity for program improvement and development is lost.

CHAPTER 2: METHODOLOGY

GOALS

This study has four primary goals:

- To make reasoned inferences about the effectiveness of adult literacy education in the United States.
- To identify common conceptual, design, and methodological problems inherent in the outcome studies conducted.
- To raise and discuss issues for policy.
- To make recommendations for the design and conduct of future outcome studies.

RESEARCH STRATEGY

Based on an earlier analysis of the outcome literature on adult literacy education (Beder, 1991), it was not anticipated that a definitive outcome study (or studies) of adult literacy existed from which logical conclusions regarding program effectiveness could be inferred. When the outcome literature was reviewed in conjunction with this study, this, in fact, proved to be the case. Consequently, the strategy here was to analyze a wide range of outcome studies in order to make reasonable inferences about effectiveness from patterns of findings while taking research limitations into account. The analysis was qualitative in orientation. Although it was hypothetically possible to conduct a quantitative meta-analysis for some outcome variables—tested learning gain, for example—it was determined that valid data from a sufficient number of studies did not exist. Furthermore, statistical information critical for a quantitative meta-analysis was generally not reported for the existing studies.

The first step was to identify a pool of research studies conducted since the inception of the federal adult literacy program that were available in the public domain and potentially included outcome assessment components. ERIC and other abstracting services were searched using numerous descriptors, and state and national policy makers were consulted regarding studies the search might have missed. The initial search identified approximately 115 studies. Next, abstracts were reviewed to determine which studies in the initial pool did, in fact, include outcome components. These studies were ordered in hard copy when available and secured in microfiche when hard copy was not available. Sixty-eight studies that included an outcome component were identified and acquired for assessment. Subsequently, each study was abstracted and evaluated according to the following criteria:

- The study included an outcome/impact component.
- The report was adequately documented with respect to design and methods.
- There were an adequate number of cases.
- The sampling plan was adequate (i.e., it could and did result in external validity).
- Data collection procedures were adequate (i.e., they were not tainted by substantial attrition or biased by other factors).
- Objective measures, rather than self-report, were used to measure outcomes.
- Measures, especially tests, were valid and reliable.
- The research design included a control group.
- Inferences logically followed from the design and data.

No study fully met these criteria. Finally, those studies that were assessed as being the most credible based on the above criteria were selected for in-depth analysis. In selecting the most credible studies, the researchers initially experimented with a numerical rating system that assigned points for each study selection criterion. However, this method was abandoned when it proved to be ineffective. Under the numerical rating system, for example, it was possible for a study that included a control group, used objective measures, and had an adequate number of cases, but that also made poor or incomplete inferences, to score inappropriately high. Conversely, a methodologically more limited study that made intelligent and sound inferences despite its acknowledged limitations could score inappropriately low. In

the end, a more holistic system of selection was employed in which two researchers read all outcome and impact studies identified by the project, developed a list of the studies they considered most credible, and resolved differences by discussion. There was 100 percent consensus on the studies eventually selected.

In order to represent a broad range of adult literacy contexts, studies were organized into five categories: national, state-level, workplace literacy, welfare, and family literacy. The quality of studies varied among these categories. For example, although there had been a proliferation of workplace literacy studies since 1991, when the federal workplace education program was established, many local-level workplace education studies were poorly conceived and conducted on meager resources. In contrast, because of the resources available, the national studies tended to be a higher quality. While it was originally anticipated that 10 to 12 studies would be selected for in-depth analysis, 23 studies were eventually selected. Studies that focused on English as a second language, the incarcerated, and the handicapped were excluded from analysis. Abbreviated reports of the 45 studies that were not selected for in-depth analysis are included in a separate Appendix.

The data for the 23 case studies that follow were acquired from published research reports that vary in completeness and clarity. Hence the case studies necessarily vary accordingly. For many of the studies, measuring outcomes and impacts was but one of several research objectives. In such cases, only the portions that pertain to outcomes and impacts are reflected in the case studies.

CHAPTER 3: ANALYSIS

NATIONAL STUDIES

There have been three national evaluations of the federal adult literacy program funded under the Adult Education Act. The first was conducted in 1973 (Kent, 1973), the second in 1980 (Young et al.), and the most recent between 1990 and 1994 (Development Associates). In addition, there is a 1968 study that evaluates components of the adult literacy education program during the period when the program was administered by the Office of Economic Opportunity, prior to the enactment of the Adult Education Act (Greenleigh Associates, 1968). There has also been a national assessment of the Even Start Program (St. Pierre et al., 1993) which will be reported in the section on family literacy. Although most of the national evaluations focused primarily on input and output analyses, each included at least one component which dealt with outcome and impact concerns. Studies are reviewed chronologically, the most recent coming first.

1. The National Evaluation of Adult Education Programs (NEAEP) 1990-1994

NEAEP Reports:

Development Associates. (1992). *National Evaluation of Adult Education programs first interim report: Profiles of service providers*. Arlington, VA: Author.

Development Associates. (1993). *National Evaluation of Adult Education programs second interim report: Profiles of client characteristics*. Arlington, VA: Author.

Development Associates. (1994). *National Evaluation of Adult Education programs third interim report: Patterns and predictors of client attendance*. Arlington, VA: Author.

Young, M. B., Fitzgerald, N., & Morgan, M. A. (1994a). *National Evaluation of Adult Education programs fourth interim report: Learner outcomes and program results*. Arlington, VA: Development Associates.

Young, M. B., Fitzgerald, N., & Morgan, M. A. (1994b). *National Evaluation of Adult Education: Executive Summary*. Arlington, VA: Development Associates.

Reanalysis:

Cohen, J, Garett, M., & Condelli, L. (1996). *Reanalysis of data from the National Evaluation of Adult Education Programs: Methods and results*. Washington, DC: Pelavin Research Institute.

The NEAEP was the most comprehensive and expensive evaluation—the initial contract was for \$2,839,740—of the federal adult literacy education program yet conducted. Data collection began in 1990, and the final Executive Summary was issued in 1994. NEAEP's results are presented in four separate interim reports and a final executive summary. The first report provides a descriptive analysis of service providers, the second describes client characteristics, the third focuses on predictors of client attendance, and the fourth is concerned with learner outcomes and program results. Virtually all the outcome and impact data are found in the fourth report and Executive Summary. In addition to the reports published by the NEAEP, the report of a reanalysis of the NEAEP data conducted by the Pelavin Research Institute is highly relevant to this analysis.

Data Collection Procedures

NEAEP utilized seven procedures to collect data (Young et al., 1994a). The Universe Survey was designed to collect descriptive data from all federally funded service providers. Administered by mail in 1990, the Universe Survey collected data from 2,619 programs, 93 percent of the universe.

The Comprehensive Program Profile was used to collect more detailed descriptive data from a smaller number of programs ($n=131$). Results were weighted to permit generalizability to the universe. For programs that provided client data from more than one site, site-level data rather than program-level data were used in analysis.

The Client Intake Record A was completed for each sampled student at the time of intake by program staff trained by the researchers. Client Intake Record A collected basic demographic and client-related program information. Analysis is based on 22,548 respondents from 116 local programs.

The Client Intake Record B was completed for all sampled clients who completed at least one instructional session. It collected more detailed information, such as reliance on public assistance, living arrangements, employment status, and reasons for participation. Data were collected from 13,845 learners in 108 programs.

The Client Update Record provided instructional and attendance data. It was completed by program staff at five- to eight-week intervals during a period that started at program entry and ended up to 18 months later. Data were collected from 18,461 learners in 110 programs.

The Client Test Record yielded learners' test scores on either the Comprehensive Adult Student Assessment Survey (CASAS) or the Test of Adult Basic Education (TABE). Pretests were administered near the inception of enrollment and posttests were administered at varying intervals during instruction. Pretest scores were obtained from 8,581 learners in 88 programs and posttest scores were obtained from 1,919 learners in 65 programs.

The Telephone Follow-up Survey yielded data on the quality of instruction, reasons for termination, and the results of instruction from a sample of clients six months after they had terminated instruction. Data were obtained from 5,401 clients in 109 programs. Most of the data reported from the telephone survey pertains to 4,653 subjects who had attended at least three classes. As the NEAEP notes:

As described in Appendix C, the weighting of the telephone survey essentially adjusts the results for non-response bias. After weight adjustments were applied to the telephone survey, respondents were very similar to the clients of the national sample who began instruction, except for having received fewer hours of instruction and having been out of adult education for six months. (Young et al., 1994a, p.46)

Commenting on the telephone survey, Cohen, Garet, and Condelli (1996) noted that "our review of the telephone survey found that the survey respondents differed from non-respondents in many ways, suggesting that estimates from this data ought not to be generalized to the population of clients" (p. xi).

Although the NEAEP provided a detailed and comprehensive description of program characteristics, staff characteristics, instructional practices, and learner attendance patterns, it is the outcome data that is of concern here. Most outcome data were collected from the Client Record Forms, the Client Test Records, and the Telephone Follow-up Survey and can be found in the Fourth Report (Young et al., 1994a).

Problems and Issues

The NEAEP study is particularly important for two reasons. First, many of the findings—especially the descriptive information on programs, staff, and learners—are useful in their own right. Second, and perhaps ultimately more important, the problems the NEAEP encountered provide us with a detailed case study of the problems of conducting large-scale outcome and impact assessments in adult

literacy education. In fact, the problems experienced by the NEAEP in data collection proved to be so severe that the validity of the NEAEP's findings were called into question by the U.S. Department of Education, which commissioned the study. Because of these reservations, a second contract was awarded to the Pelavin Research Institute to "conduct a comprehensive review of the study methodology, quality of data, and statistical methods used in prior analysis; and to validate reported findings, make needed corrections and conduct new analyses" (Cohen, Garet, & Condelli, 1996, p. vi). For this reason, and because the NEAEP's limitations are well documented, the problems and lessons associated with conducting the NEAEP will be discussed here in detail. As the reader will discover, to one degree or another many of the issues raised by the NEAEP pertain to most of the outcome and impact studies in adult literacy conducted to date.

At the heart of the NEAEP was to be a set of approximately 150 local programs that were to provide detailed program information and longitudinal (18-month) information on learners, including pre- and posttests. To enable generalization from programs and their clients to the entire federal adult literacy education program, participating programs were to be selected using a probability of selection proportionate to size methodology (Development Associates, 1992). In accord with this methodology, 18 programs with 20,000 or more participants were so large that they were automatically selected for the study (i.e., they had a probability of selection equal to 1).

Operational Definitions

Although this procedure appears to be simple, in reality it was not. First there was the issue of defining what constituted a "program." The NEAEP defined a program as the administrative unit that served as the sub-grantee for federal funds. As such, the City of New York and the Los Angeles Unified School District were defined as programs. As the NEAEP explained:

Receipt of such funds, however, did not adequately target the administrative entities that should be included in the evaluation, for example, more than one grant may be made to the same administrative agent, such as separate grants for the ABE, ASE and ESL instructional components. Or sometimes a basic grant is awarded to a regional administrative service agency that has several subgrantees, some of which may be local school districts and other community-based organizations; and grantees exercise varying degree of administrative control over the service delivery agencies. (Young et al., 1994a, p. 4)

The definition of "client" was also an issue. A client was conceptually defined as one who received services directly supported by the Adult Education Act. Yet because of multiple funding streams that flowed to local programs from state funds, JOBS, JTPA, and private sources, it was often difficult to determine if a given client was wholly, partially, or not at all supported by Adult Education Act funds. Learners who changed from classes supported by one funding source to classes supported by another funding source during the course of the evaluation caused further problems in definition. When a client became a client for the purposes of the study was also problematic. Although eligibility for the study was intended to occur at client intake, the NEAEP discovered that 16 percent of those who completed intake never engaged in instruction. Furthermore, in its reporting the federal government counts only those learners who have completed 12 hours of instruction as participants. The NEAEP found that 17 percent of those who completed the intake process never achieved 12 hours of instruction (Young et al., 1994a, p. 10).

The point is that it is impossible to select a sample that represents the universe of adult literacy education programs and clients if one cannot precisely define the universe and its components in operational terms.

Program, Subject, and Site Attrition

More serious, perhaps, was program and site attrition from the study. Although the NEAEP provided economic and other incentives for programs to cooperate in data collection, it had no direct authority over the participating programs. As a result:

Our goal was to enlist participation of 150 local programs. When data collection began, 141 programs had agreed to participate, 114 from the initially selected set of 150, 25 first order replacements, and two replacements of replacements. After 10 months of data collection, 2 of the originally selected 150 had terminated operations, 3 had formally withdrawn from the evaluation, and another 5 had failed to submit data. Within the first six months, 16 percent of program directors trained in the requirements of the study had departed, sometimes because their positions had been abolished. (Young et al., 1994a, p. 5)

Program and site attrition from the study and incomplete data supplied by programs wreaked havoc with the NEAEP's weighting protocol based on program size and made the problem of generalizing to the universe complex. As a case in point, the adult literacy education program provided by the Los Angeles Unified School District is so large that it was automatically included in the study. Indeed, Los Angeles served about 10 percent of the national population and was about 72 percent ESL (Cohen, Garet, & Condelli, 1996). However, the data supplied by Los Angeles were submitted late and differed from what might have been expected to such an extent that its accuracy was severely questioned. The problem is that, because of Los Angeles' size, if the data is accepted as accurate and included, NEAEP's findings differ substantially—especially with respect to ESL and variables associated with ESL—from when the data is not included. NEAEP included the Los Angeles data; the Pelavin reanalysis excluded it and adjusted the findings to account for the deletion. Commenting on the problem program and site attrition caused for the weightings, and hence the generalizability of the findings, the Pelavin re-analysis stated:

If relatively complete sample coverage had been achieved, it would have been possible to estimate population parameters using appropriate design weights. Given the substantial non-coverage of clients within sites, however, as well as the substantial non-response at the site and program level, weighted estimates are likely to be biased, although the direction and extent of the bias is difficult to assess. (Cohen, Garet, & Condelli, 1996, p. viii)

With respect to the outcome data collected by the NEAEP, the inability of programs to supply accurate information at specified time intervals was a very serious problem.

Originally it had been hoped that 150 programs would supply pretest and posttest data on participants. Ultimately, 101 agreed. Yet these programs were only able to provide pretest data on 57 percent of the potential base level of clients (11,354 out of 19,796). Furthermore, "the number of matched scores [pre- and posttest] obtained for clients on all tests constituted only 12 percent of the clients originally expected (i.e., 2,333 out of 19,796)" (Young et al., 1994a, p.30). Although posttesting was to occur at fixed intervals, so many posttests submitted were not administered at the stipulated intervals that the NEAEP had to abandon the fixed interval protocol and instead report average hours of instruction for the posttest data. To compound things, after analysis of the test data submitted, the NEAEP discovered that much of it was either incomplete or tainted by ceiling or floor effects. Consequently, of the 19,796 matched pre- and posttests that the NEAEP had originally hoped to obtain, only 614 usable cases were obtained for analysis. Furthermore, 50 percent of the learning gain data on ABE/ASE was obtained from but three sites; 75 percent of the ESL data came from nine sites (Cohen, Garet, & Condelli, 1996). Assessing the validity of the tested learning gain data in relation to the problems encountered in data collection, the reanalysis study concluded:

The implementation of the test plan was also poor, and this data should not be used to assess the capabilities of clients at intake. Some of the key evidence supporting this conclusion includes:

- Only half the clients were pretested, and sites that pretested differed from sites that did not. At sites that pretested only some of their clients, pretested clients differed from those who were not pretested.
- Programs reported perfect exam scores for a substantial proportion of pretested clients.
- Less than 20 percent of eligible clients received a matched pretest and posttest.
- Among clients eligible to be posttested, significant differences exist among those who were and were not posttested.
- The available matched pre- and posttests were concentrated in a very few programs.

These facts render the test data unusable. *Therefore this reanalysis invalidates all of the findings concerning test results from the original analysis.* (Cohen, Garet, & Condelli, 1996, p. xi)

Logistical Constraints

NEAEP's experience suggests that there were at least two reasons why it is difficult to collect accurate and timely data from local programs when local program staff are assigned to keep records and administer tests. First, the workforce in adult literacy education is predominately part-time. Indeed, the NEAEP found that over 80 percent of the teachers teach part-time (Development Associates, 1992). It is difficult to expect part-time staff who devote most of their time to instruction and have little training or experience in research to accurately collect the kind of data required of a credible outcome and impact evaluation. More important, perhaps, is that most adult literacy education programs maintain open enrollments, attendance tends to be irregular, and dropout rates tend to be high. These factors make it very difficult to collect data from learners at regular intervals, and the high attrition rates inject the threat of attrition bias.

Findings

The outcome and impact variables measured by the NEAEP included tested learning gain, clients' self-report of learning gain, employment, further education, clients' assessment of personal goal attainment, and how often clients read to their children.

Tested Learning Gain

Because the NEAEP lacked the resources to administer a common test to all subjects, and because variation in instructional goals and processes among programs made it inappropriate to administer a common test, the NEAEP used the tests that were already being administered in the programs selected. Although there are a number of instructional tests employed by local programs, there were but two tests that were in sufficient usage to yield enough cases for the national evaluation—the California Adult Student Assessment System (CASAS), which was used to measure learning gain for ESL, and the Test of Adult Basic Education (TABE), which was used for adult basic education (ABE) and adult secondary education (ASE). As already mentioned, the NEAEP was able to obtain only 614 valid cases from an intended 19,796 potentially available cases. Because intervals between pre- and posttests varied, the mean hours of instruction between pre- and posttests was reported. For ESL students, who on average received 120 hours and 14 weeks of instruction between the pre- and posttests, the learning gain on the CASAS was five scale points. ABE students received a mean of 84 hours of instruction between pre- and posttests and attended for an average of 15 weeks. On average their gain was 15 points on the TABE. Adult secondary students received a mean of 63 hours of instruction and gained 7 points on the TABE. All gains were statistically significant (one sample *t*-tests) at the .001 level (Young et al., 1994a).

What do these results mean? It is difficult to say. First there is the issue of the data's validity. Although Development Associates took care to "clean" the test data, when only 614 acceptable cases were obtained from a potential of over 19,000 cases, validity is still a very major issue. The second issue has to do with standards. What is an acceptable learning gain after the mean hours of instruction reported? We do not know. Development Associates (Young et al., 1994a) notes that ABE students gained from an average 6.1 grade level on the TABE to a 7.4 grade level and that adult secondary students gained from a 8.5 grade level to a 9.3 grade level, but does this represent high, low, or moderate gain? Moreover, are grade level changes, as normed on elementary and secondary school students, valid indicators of learning gain in adult literacy education? Many would say they are not.

The NEAEP was only able to obtain matched pre- and posttests on 12 percent of the potential number of clients expected. Thus a high proportion of those who were pretested were not posttested. Did those who were both pretested and posttested differ from the original sample in important ways? Were they more able or less able? This is an important issue, especially in adult secondary education. There is anecdotal evidence that some of the most able adult secondary education students leave programs after but

a few hours of instruction to take, and eventually pass, the GED. Were those ASE students who were available for posttesting the least able of those adult secondary education students who enrolled? We do not know. Finally, how do we know that the gains reported resulted from adult literacy education instruction? The only way to answer this question convincingly would have been to have had a control group with which to compare the gains. Without a control group, it cannot be determined whether the gains were real or due to other factors, a particular concern in ESL where presumably students were interacting with the English language outside of class.

Through multiple regression analysis, the NEAEP went on to determine which background, attendance, and program variables influenced learning gain for ABE, ASE, and ESL. Two findings from this analysis are of particular import. First, almost two thirds (61 percent) of the variance for learning gain in reading for ABE was found to be accounted for by the pretest score in reading, which may be taken as a general measure of ability. For ESL the score was 48 percent of the variance, and for ASE the score was 19 percent. These findings suggest that, especially with respect to adult basic education, a learner's initial "ability" is a potent predictor of learning gain. Second, total hours of instruction were not shown to be related to tested learning outcomes for either ABE or ASE. This finding runs counter to conventional logic which would assume that the more time adult literacy students spend in class, the more they would learn. The NEAEP did not offer explanations for the finding, but assuming that the finding is valid, the relationship between tested learning gain and hours of instruction certainly warrants more research.

Most of the findings that pertain to employment, learners' goals, and further education were derived from NEAEP's telephone survey of former learners who had not attended for six or months. From 109 local programs, 5,401 clients responded, 86 percent of whom had attended for at least three sessions. Learners were asked if participation in the program had helped them improve their basic skills. For ESL, 44 percent responded "a lot" for reading and writing, 26 percent for mathematics, 48 percent for speaking and listening, and 62 percent reported that they had been helped a lot in at least one skill level. For ABE the figures for "a lot" were 50 percent for reading and writing, 51 percent for mathematics, 48 percent for speaking and listening, and 68 percent reported that they had been helped in at least one skill level. For ASE the comparable figures were 45 percent reading and writing, 49 percent mathematics, 45 percent speaking and listening. Sixty-three percent reported they had been helped "a lot" in at least one skill level. The NEAEP examined the extent to which learners' reports of the benefits they attained in reading coincided with tested improvement in reading and found convergence in 58 percent of the cases.

Employment and Further Education

With respect to employment, 63 percent of the learners reported that they were unemployed when they entered the program and 69 percent were employed at the time of follow-up. However, without a control group it is impossible to determine whether these statistically significant, but modest, gains were due to instruction or to other unknown reasons. In fact, when NEAEP asked those who became employed between enrollment and follow-up if what they learned in the program helped them get a job, a majority (57 percent) said no, suggesting that learners perceived that factors other than adult literacy education were critical for job acquisition. With respect to further education, of those learners who did not possess a high school diploma at intake, at follow-up 18 percent were enrolled in further education (11 percent post-secondary, 6 percent GED, 1 percent other), 44 percent had no plans to enroll, and 38 percent expected to enroll within a year.

Self-image and Learner Satisfaction

Although enhancing learners' self-image and self-esteem are not stated goals of the federal adult literacy education program, improved self-image is a common variable in outcome and impact studies. The NEAEP found that 65 percent of the learners reported that they felt better about themselves at follow-up. These data are mitigated by the fact that the follow-up sample included both respondents who had terminated because they attained their goals and respondents who had dropped out for other reasons. When respondents were asked at follow-up why they left the program, 41 percent were designated by NEAEP as

having “left satisfied,” 45 percent were designated as having left for the sake of outside events beyond their control, and 7 percent left for instructional factors. On the average, satisfied respondents had substantially more hours of instruction than respondents who had left for other reasons.

Strengths

- With respect to its variables, data sources, and scope, the NEAEP is the most comprehensive national evaluation of the federal adult literacy program to date. It provides a rich descriptive picture of adult literacy education in the United States.
- The NEAEP’s reports are clearly written, adequately documented with respect to methods and procedures, and honest in their portrayal limitations.

Limitations

- Attrition of programs, sites, and subjects from the study compromised the original sampling plan based on program size. Although the NEAEP adjusted the weightings to account for program and subject attrition, NEAEP’s adjustments were criticized in the reanalysis study. Thus the validity of the NEAEP with respect to the outcome data is problematic.
- Because of data collection problems, only a small proportion of the intended posttest data was obtained and the intervals between pretest and posttest varied widely. The accuracy of the test data has also been questioned. The data collection problems associated with the test data were so severe that it is questionable whether any reasonable inferences can be made from these data.
- Lack of a control group casts doubt on an inference that the outcomes obtained were caused by participation in adult literacy education.
- With the exception of the learning gain data, outcomes are based on learners’ self-reports of benefits acquired.

When the limitations of the NEAEP are considered with respect to outcome findings, it must be concluded that at best the findings are suspect and at worst they are unusable. Perhaps because the NEAEP cost almost \$3 million and took four years to complete, the evaluation was subject to a considerably higher level of scrutiny than the national evaluations that preceded it. Because the data collection methods of the earlier evaluations are in many ways similar to the NEAEP, the suspicion lingers that the earlier national evaluations, too, were flawed in ways similar to the NEAEP and that these flaws either went unnoticed or were not reported.

2. The 1980 National Evaluation An Assessment of the State-Administered Program of the Adult Education Act

Young, M., Hipps, J., Hanberry, G., Hopstock, P., & Golsmat, R. (1980). *An assessment of the state-administered program of the Adult Education Act: Final report*. Arlington, VA: Development Associates.

As with the NEAEP, the 1980 national evaluation of adult literacy education was conducted by Development Associates under contract from the U.S. Department of Education. Work commenced in 1978, most data were collected in 1979, and the final report (Young et al., 1980) was issued in 1980. The general purposes of the evaluation were: “(a) to provide an analytic description of the Grants to States program, with particular emphasis on program participants and (b) to identify a set of impact measures that could be studied in a longitudinal design” (Young et al., 1980, p. 1). Because the evaluation was primarily descriptive, the collection of comprehensive, generalizable outcome and impact data on participants was not a major objective of the study.

State-level data for the evaluation were gathered through a mailed survey to state directors of adult education in the 50 states. The response rate was 100 percent. Local program data were collected from a national probability sample ($n=420$) of local programs in 47 states. Four hundred and four local directors responded, for a response rate of 96 percent.

The outcome and impact data for the study were collected from a separately drawn sample of 110 local programs stratified according to type of sponsoring agency and size of program. At each local program, data were collected from directors, a randomly selected sample of teachers, and present and former participants. The intent of the evaluation was to collect data from learners who had completed at least one adult education course or who had terminated the program. Thus subjects were selected who had completed a course the semester prior to learner data collection.

Each randomly selected teacher from among the 110 selected programs was asked to list the students he/she had taught during the spring of 1979. In order to obtain a sufficient number of learners per site, a minimum of 25 participants were selected at each site. If the program selected had five or more teachers, five learners were selected from each teacher's class list at random. Two kinds of data were obtained about learners. Descriptive data regarding learner characteristics and participation were collected from program records during site visits to the 110 programs. The second kind of data, upon which the outcome findings rely, was collected from selected learners through interviews that were generally conducted over the telephone. Although at least three attempts were made to reach each sampled learner, the final response rate was but 38 percent (1,177 cases out of a sampled 3,061 cases). Less than half (43 percent) the valid cases were still active in the program at the time of data collection. Because of the relatively low response rate, and because it was determined that those interviewed differed from those who could not be contacted, no attempt was made to generalize the findings to adult literacy education in general in the United States.

Findings

Outcome variables measured included participants' reports of personal goal attainment, self-reported gains in self-concept, basic skill acquisition, getting along better with one's family, getting a better job, and further education.

In the interviews, learners were asked if they had reached, or were in the process of reaching, the goals they had set for themselves when entering the program. Forty-two percent reported that they had successfully reached their goals (ABE=44 percent, ASE= 49 percent, ESL=24 percent). Thirty-eight percent reported that their goals had been partially obtained (ABE=36 percent, ASE=28 percent, ESL=56 percent), and 17 percent claimed that their goals had not been met (ABE=16 percent, ASE=21 percent, ESL=14 percent). Secondary analysis showed that goal attainment was positively associated with several program variables, including having spent at least one year in the program, being a participant at the time of data collection, and having attended classes more frequently than the norm. With respect to demographic variables, females were more likely to report both successful and unsuccessful goal attainment. Perceived goal attainment increased with age.

Regarding specific outcomes gained, 84 percent of the respondents reported that participation had improved their self-concepts (10 percent reported no improvement), 75 percent reported that they had improved in reading (22 percent reported no improvement), 69 percent reported that they had improved in mathematics (27 percent reported no improvement), and 66 percent perceived that they had improved in writing (29 percent reported no improvement). Fifty-one percent believed that their family relations had improved because of participation (33 percent reported they had not), 25 perceived that their life skills had improved (62 percent reported they had not), and 18 percent reported that participation had helped them get a job (69 percent reported it had not).

With respect to future educational plans, 57 percent of those interviewed said that they planned to enroll or had already enrolled in adult education courses and 24 percent reported that they were uncertain of their future educational plans. Eighteen percent reported that they planned no future education. Fifty-eight percent of the respondents said that they planned to enroll in schooling other than adult education, and 23 percent said they were uncertain about such plans.

Strengths

- The 1980 national evaluation is comprehensive, particularly with respect to descriptive information.
- The report is clearly written, and methods and procedures are adequately documented.

Limitations

- There is a paucity of outcome data.
- Because of a low response rate to the participant interviews, and because respondents were shown to have differed substantially from non-respondents in important respects, the outcome data from the 1980 national evaluation cannot be generalized to adult literacy education in general with any degree of confidence. Because of this limitation, other limitations such as lack of a control group and a reliance on self-report data are rendered moot.

3. A Longitudinal Evaluation of the Adult Basic Education Program, 1973

Kent, William P. (1973). *A longitudinal evaluation of the Adult Basic Education Program*. Falls Church, VA: System Development Corporation.

The 1973 evaluation of the federal adult literacy program funded under the Adult Education Act commenced during the middle of 1971 and terminated more than two years later. As with the 1980 national evaluation and the NEAEP, the 1971 national evaluation collected a great deal of descriptive information regarding the federal program. However, unlike the other national evaluations, the 1971 national evaluation focused primarily on learner outcomes. A major difference between the 1971 and other evaluations, however, lies in the groups it excluded. Because the Adult Education Act at that time restricted instruction to adults at the pre-secondary level, the study was limited to learners with eight years or less of schooling. For logistical reasons, the study excluded ESL programs, migrant programs, and Native American programs, and, because at that time the priority population for adult literacy education was defined as learners between age 16 and age of 44, persons under 16 and over age 44 were also excluded. Although it had originally been planned to establish control groups for the study, these plans were abandoned when they proved not to be feasible.

Because it was impossible to acquire an accurate count of the number of students meeting the study's criteria, it was decided to use states as the primary unit in sampling. A two-way stratification protocol was adopted, based on grouping states by geographic region and the proportion of black students enrolled. The 50 states were classified into an 8x8 matrix and 16 cells were drawn from the total "with each cell's probability of being drawn proportional to the number of students assigned to the cell....Finally one state from each of the selected cells was drawn to enter the sample, the probability of being drawn proportional to the number of students" (Kent, 1973, pp. 3-8). From each of the selected 16 states, programs were selected using a proportionate-to-size methodology, and from each program, classes were randomly selected so as to produce a sample of approximately 25 students per program. The final sample constituted 92 programs, 206 classes, and 2,318 learners.

The 1971 evaluation collected data from a number of sources, including state directors, local program directors, and learners. Most of the outcome data were collected from a learning gain test and a survey. The learning gain test was first administered in January 1972 and again in May 1972. The survey was administered in February and March 1972 and then again 12 and 18 months later.

Tested Learning Gain

After considering the learning gain tests available for adult literacy education in 1972, the 1973 evaluation selected two tests from level M of the Test of Adult Basic Education (TABE), one measuring reading comprehension, the other measuring arithmetic fundamentals. As the final report states:

The TABE had three levels ("E," Easy; "M," Medium; and "D," Difficult). Level E is suitable for grades 1 through 4; M for grades 2 through 9 and D for grades 3 through 12. Since M covers all grades of interest to this study except 1.0-1.9, it is nearly satisfactory all by itself. (Kent, 1973, pp. 3-13)

The validity and reliability of the TABE components used are not reported. After developing directions and field testing, tests and instructions were distributed to local program directors, and teachers were asked to administer the tests. Only the tests from the learners selected in the sample were used. Of the 1,108 initial tests that were obtained, matching pre- and posttests were obtained for 441 subjects. It is important to note that, strictly speaking, the tests administered were not pre- and posttests since at initial testing learners had already received instruction to varying degrees.

Surveys

The collection of student interview data was subcontracted to a market research firm. Initial student interviews coincided with ABE classes; students were excused from class for 20 minutes to complete the interviews. If necessary, interviewers returned to class several times to interview learners who had been selected as part of the sample. For those learners who could not be reached at class times, a sub-sample of half the absentees was established, and those selected were interviewed in their homes or other places. At least two attempts were made to interview members of the absentee sub-sample. First follow-up interviews were conducted a year after the initial interviews. Seventy-four percent of the initially interviewed learners were interviewed at first follow-up. Second follow-up interviews were conducted 18 months after the initial interviews, generally by phone. The response rate was 79 percent. Of the 1,448 students who were initially interviewed, 1,065 were reached for first follow-up and 844 were reached for the second follow-up. It is not known whether the 844 learners who were interviewed after 18 months differed substantially in important ways from the 604 learners who were originally interviewed but could not be reached 18 months later.

Findings

Learning Gain

When initially tested with components of the TABE, learners scored at grade level 5.4 on reading achievement and 6.4 in mathematics (raw scores were not reported). After the second administration of the test approximately four months later, in which a different test form was used, 26 percent of the students had gained one grade or more in reading, 41 percent had some gain, but less than one grade, and 33 had zero or negative gain. In mathematics, 19 percent gained one or more grades, 46 gained some, but less than one grade, and 35 percent showed zero or negative gain. The proportions of those who gained and those who did not may have been affected by the differing hours of instruction learners had amassed between the first and second test administrations. While almost a fifth of the learners had 39 hours or less of instruction between the first and second testing, another fifth had 80 or more hours of instruction. Average gains for reading were .5 grades after 98 hours and .4 grades after 66 hours. For mathematics the comparable figures were .3 grades both after 98 hours and after 66 hours.

In the 1973 study, learners with the lowest initial scores tended to show the greatest gains. In reading, those whose initial scores were below the fifth grade showed average gains of .8 grade levels, those whose initial scores were at the fifth or sixth grade demonstrated gains of .3 grade levels, and for those who initially scored at the seventh grade or higher the average reading gain was 0 grade levels. For mathematics the pattern was similar. This is an apparent contradiction with the NEAEP which, using multiple regression, found that ABE learners' pretest scores accounted for 61 percent of the variance in posttest scores. There are many factors that may explain this contradiction. One explanation is that, while the NEAEP carefully screened its cases to eliminate those that were likely to have been tainted by floor and ceiling effects, if the 1973 evaluation did so, it is not reported. Thus the findings of the 1973 evaluation may have been unintentionally biased by both floor and ceiling effects. The ceiling effect hypothesis is

supported by the evaluation's finding that, despite the intended exclusion of adult secondary-level learners, 54 percent of the sampled clients had completed nine or more grades of school. As with the NEAEP, the 1973 evaluation showed no linear relationship between hours of attendance and tested learning gains.

Females gained more than men, there was no clear relationship between learning gains and race and age, and there was no consistent pattern between learning gains and previous school experience.

Further Education

When they were initially interviewed, 60 percent of the learners reported that they might attend college at some time and 70 percent reported that they might attend vocational school at some time. However, over time, the percentages decreased substantially for college attendance. At follow-up, only 37 percent said that they might enroll in college and 65 percent said they might attend vocational school.

Helping School-age Children

When they were initially interviewed, 55 percent of the learners reported that they had helped children with schoolwork; at follow-up 58 percent said they had helped children with schoolwork.

Employment and Earnings

Assessment of the impact of adult literacy education on employment and earnings was a major objective of the 1973 national evaluation. When initially interviewed, 55 percent of the learners reported that they were working, 26 percent were receiving public assistance, and 58 percent reported that they had some earnings. A year later, 63 percent were working, 24 percent were on welfare, and 66 percent had some earnings. Eighteen months later, 65 percent were working, 22 percent were on welfare, and 70 percent had some earnings. Overall, there was a 10 percent increase in employment over 18 months and a 12 percent increase in the incidence of some earnings. For those who were employed at the time of initial interviews, over 18 months, mean monthly earnings increased from \$336 to \$407, mean hourly earnings increased from \$2 to \$2.36, and mean hours of hours worked per week increased from 37.1 to 39.1. Thus for 18 months, mean monthly earnings increased 21 percent. Extrapolating to the entire adult basic education population, the 1973 evaluation reported that the increases would have amounted to \$46 million per year (pp. 2-24)

The gains in earnings reported by the 1973 evaluation are very large. Acknowledging that inflation and wage fluctuations may have accounted for perhaps 5 to 6 percent of the increase, the 1973 evaluation reported that "in analyzing the gains indicated by the foregoing figures, several considerations suggest that the gains are real rather than a product of selective interviewing, increases in hours of work, or inflation" (Kent, 1973, pp. 2-23). Unfortunately, the only way to determine with confidence whether these gains were real would have been to compare the gains of participants with a control group whose members differed from participants only with respect to participation in adult literacy education. Although such a control group was planned for the evaluation, it was never implemented.

Learners who had received pay increases were asked how much they thought ABE had helped in getting the pay increase. Fifteen percent responded very much, 20 percent responded some, 12 percent responded a little, and 52 percent responded not at all. Thus even if the large increases in pay reported by the 1973 study were real, almost two thirds of the respondents perceived that participation in ABE played either no role or a minor role in obtaining the increases.

With the exception of the tested learning gain data, most of the outcome data is based on learners' self-reports. This interjects a source of bias that becomes apparent in the data reported on attendance. Learners were asked, "During which months did you attend the adult basic education at least one time during the month?" When these data were compared to teachers' reports, there were wide discrepancies between students' and teachers' reports and these discrepancies tended to increase over time. In November 1971, 81 percent of the students reported they had attended at least once, while 85 percent of the teachers

reported that the student in question had attended at least once. In June of 1972, however, although 38 percent of the students reported attendance, only 17 percent of the teachers reported attendance.

Strengths

- As with the NEAEP and the 1980 national evaluation, the 1973 national evaluation was comprehensive in scope.
- The report is clearly written and methods and procedures are adequately explained.
- The outcome data is longitudinal.

Limitations

- Because of age and grade level exclusions in the 1973 evaluation, comparison with other national evaluations is problematic.
- Because the tests used for tested learning gain were extrapolated from the TABE, and because a single level of the TABE was used for all respondents, the validity and reliability of the test data are issues.
- As with the NEAEP there was some program attrition from the study and there was substantial respondent attrition. Although the study attempted to compensate, the extent to which attrition compromised the proportionate-to-size sampling plan is an issue. If the sampling plan was seriously compromised, the generalizability of the findings is questionable.
- Respondent attrition leads one to question the internal and external validity of the outcome findings.
- Posttests were administered approximately four months after pretests. Whether this is a reasonable time over which to expect meaningful learning gains is an issue.
- With the exception of the test data, outcome findings are based on learner self-report.
- The study lacked a control or comparison group.

4. The 1968 Evaluation

Greenleigh Associates, Inc. (1968). *Participants in the field test of four adult basic education systems: A follow-up study*. New York: Author.

The earliest evaluation of adult basic education that could be considered national in scope was conducted by Greenleigh Associates prior to the advent of the Adult Education Act. At that time, the federal adult literacy education program was administered by the United States Office of Economic Opportunity. The evaluation, which was conducted between June 1966 and January 1968, is the follow-up to a field test of adult basic education programming conducted from March 1965 to May 1966 in three states—New York (three counties), New Jersey (two counties), and California (one county). Participants in the field test, who had been selected to assure representation on age, gender, and ethnic background, had all scored below the fifth-grade level on a standardized test. Initially 540 participants were randomly placed in 36 classes.

With respect to the outcomes portion of the evaluation, the objectives were to determine the extent to which reading skills learned during the field test had been retained and to assess the effects of the educational experience provided in the field test on income, housing, family relations, health, and motivation.

In order to permit comparisons, participants in the follow-up study were assigned to one of three groups based on the following characteristics. Participants (Group I) were those who had completed 17 weeks of adult basic education in conjunction with the field test. Non-participants (Group II) were those who qualified to participate by virtue of the fact they had scored below grade level 5 on the reading test but had dropped out during the first two days or had declined to attend classes. It should be noted that Group II did not constitute a true control group since there was no random assignment to treatment and control groups involved. In fact, Group II members obviously differed from Group I participants with respect to their motivation and/or ability to participate in adult basic education. As the authors suggest, it was

possible that while “Group I to a significant extent were continued in adult basic education following the field test...Group II were pressured to enter basic education or seek employment with the threat of dismissal from welfare for noncompliance. Apparently, Group II persons sought and obtained employment rather than engage in education (Greenleigh Associates, 1968, p. 30).” The third group (Group III), designated as “overqualified,” were those who did not participate in the field test since they had scored above 5.9 on the standardized reading test.

Two methods were used to collect data—a series of two interviews conducted 6 and 12 months after the field tests and a standardized reading test, the Gray’s Oral Paragraphs Test. The test was administered twice following the interviews.

Five to six interviewers, most of whom were trained social workers, were recruited for each state. All received one week’s training prior to the interviews and administration of the test. For the most part, interviews and tests were administered in subjects’ homes. Although it took an average five to six contacts to produce a single interview, during the first set of interviews 1,641 cases were obtained from a sample of 2,003. Three hundred and sixty-two of those who were scheduled for first interviews ultimately could not be reached. During the second round of interviews 1,425 cases were obtained and 216 of those scheduled for interviews could not be reached. It is likely that the evaluation was affected by attrition bias. Although the universe was 15 percent white, 14 percent male, and 66 percent over age 30, the first interview participants were 20 percent white, 18 percent male, and 73 percent over age 30. Of those who were interviewed in both the first and second interviews, 18 percent were male and 74 percent were over age 30; of those who were administered the first interview, but not the second, 23 percent were male and 66 percent were over age 30. Qualitative as well as quantitative data were collected during the interviews, and the qualitative data are used to elaborate and refine the quantitative findings reported.

Upon the recommendation of the New York State Education Department, the Gray’s Oral Paragraphs Test was used as the reading test for the study. Because this test was designed for, and normed on, children, the validity of the test for an adult population is an important issue. More appropriate, adult-oriented reading tests were simply not available at the time of the study.

In an attempt to verify the data gathered in interviews, caseworkers were also interviewed. However, this component of the evaluation was fraught with problems. As the authors note:

In many cases case workers were completely unfamiliar with their clients, had never seen them, and were obviously unacquainted with the field record. This occurred because turnover among caseworkers was unusually high, and reorganization often caused reshuffling of case loads. (Greenleigh Associates, 1968, p. 19)

Findings

Economic Impact

Most participants in the study experienced no change in their welfare status between the first and second interviews (Group I=86 percent, Group II=77 percent, Group III=86 percent). A small number (Group I=2 percent, Group II=3 percent, Group III=2 percent) of those who were not initially on welfare became welfare participants. A majority of those in each group who were not on welfare during the first interview were still not on welfare during the second interview. Only a small group who were on welfare during the first interview had been removed from welfare by the second interview (Group I=5 percent, Group II=5 percent, Group III=5 percent).

Qualitative data collected during the interviews suggested that:

According to the interviewers in each state during both interview periods, the overwhelming majority of the study population asserted their desire to gain their independence from welfare.

However, they could not envision how to accomplish this goal. The move into gainful employment required the solution of critical difficulties, in most cases related to problems of adequate child care. (Greenleigh and Associates, 1968, p. 27)

The employment rate for participants (Group I) at the time of the first interviews was 17 percent while the employment rate for this group at the second interview was 20 percent. Nine percent of the participants in adult basic education who were unemployed during the first interview gained employment by the second interview (Group II=4 percent, Group III=9 percent). Ten percent of participants were employed during both periods (Group II=17 percent, Group III=11 percent), and 75 percent were unemployed during both periods. Six percent of those participants who were employed during the first interview became unemployed by the second (Group II=9 percent, Group III=5 percent). The authors caution that, because Group I members were enrolled in basic education, they were not necessarily available for employment.

Social Impact

The 1968 evaluation showed substantial impact on social participation variables for all groups. In respect to reported participation in community organizations, at the time of the first interview 12 percent of participants reported that they had participated in community organizations (Group II=7 percent, Group III=17%), while at the time of the second interview, 31 percent of the participants reported participation in community organizations (Group II=22 percent, Group III=31 percent) for a net gain of 19 percent for participants.

Education

At the time of the first interview, 20 percent of those initially designated as participants in adult literacy education (Group I) were still attending, 44 percent had completed the program, and 16 percent had dropped out. Of those who participated in a first interview, 67 percent were still attending a program, 9 percent had completed, and 24 percent had dropped out by the time of the second interview. Participants were asked in what areas they had gained as a result of participation, and, of the 10 options listed, the greatest gains were reported in reading skills (29 percent first interview, 23 percent second interview), arithmetic skills (21 percent first interview, 19 percent second interview), and self-confidence (13 percent first interview, 11 percent second interview).

Initial scores obtained at the beginning of the field test on Gray's Oral Reading Paragraphs Test were 2.9 for participants (Group I), 2.9 for non-participants (Group II), and 8.0 for the overqualified group (Group III). During the first interviews, approximately 12 months later, scores were 4.0 for group I, 4.0 for Group II, and 8.4 for group III. Although the test scores may well have been biased by attrition effects and the validity and reliability of the test are significant issues, it appears that those who participated in adult literacy education after the field test did not score higher on the test than those who did not participate. At the end of the second interviews, which took place 6 months after the first, participants scored 4.0, non-participants scored 4.0, and Group III scored 8.2. Thus again there was no appreciable gain between participants and non-participants. It was also found that the level of teacher preparation, defined as high school graduate, college graduate, or certified teacher, had no effect on learning gain. Likewise, the type of reading system used (American Incentive to Read, Science Research Associates, Mott Basic Language Skills Program, Systems for Success) had no effect on learning gain.

Strengths

- The 1968 evaluation is comprehensive with respect to the variables included and the *n* is large.
- The report is clearly written and is well documented with regard to methods and procedures.
- The study is longitudinal.
- Qualitative data enhance and enrich the report.
- Comparison groups were included.

Limitations

- Because the adult literacy education program as constituted in 1968 differs markedly from the program as constituted today, findings cannot be generalized to the present. Because the states and counties studied were not representative of the universe, findings cannot be generalized to adult literacy in general in 1968.
- The comparison groups are not true control groups. Because Group II, designated as non-participants, was drawn from those who elected not to participate in classes or dropped out, this group may be biased at least with respect to its motivation and/or ability to participate in adult literacy education.
- The test used, the Gray's Oral Reading Paragraphs Test, was designed for, and presumably normed on, children. As such its validity for adults is questionable.

STATE STUDIES

Although since the late 1960s many states have published input-output studies describing their adult literacy education programs, few states have conducted formal outcome assessments. Only nine state outcome studies were assessed as being sufficiently credible for inclusion in this report: California, New Jersey, Maryland, Ohio, Tennessee (two studies), Utah, Washington, and Wisconsin. Studies will be reported in chronological order, the most recent listed first.

5. The Washington Workforce Training Study

Washington State Training and Education Coordinating Board. (1997). *Workforce training results net impact and benefit cost evaluation*. [Meeting NO. 52, March 25, 1997]. Author.

Because the final report of the Washington Workforce Training Study was not available at the time of this review, the above preliminary report was used for our analysis. The purpose of the Washington study was to determine whether the outcomes for participants in workforce training programs were similar to the outcomes for those who did not participate in training. Workforce training programs were operationally defined as post-secondary training at community colleges and technical colleges, adult basic skills education for participants who enrolled for work-related reasons, Job Training Partnership Act (JTPA) Title IIa programs for disadvantaged adults, JTPA Title IIc programs for youth, and secondary vocational-technical education. The analysis here will focus on the adult basic skills program.

The study, which was conducted by the Battelle Memorial Institute, employed a matched comparison group analytic design. To examine both short and longer term effects, two groups of program completer/leavers were studied—those who exited during 1991-1992 and those who terminated in 1993-1994. Sample selection began with the acquisition of lists of all program completer/leavers for the years studied. As the report acknowledges, this was problematic for adult basic education since it was impossible to separate completers from dropouts. Hence, completer/leaver for adult literacy was defined as anyone who stopped attending and did not return for at least one year.

The selection of a comparison group began with a list of employment service registrants who had not participated in any of the workforce education programs studied. Through a logistic regression procedure, comparison group members were selected who were "equivalent" to treatment group members with respect to age, race, ethnicity, gender, education, region, employment history, earnings, and receipt of welfare and unemployment insurance. Although the study's authors reported that this procedure was generally successful in matching treatment group participants with comparison group members, the match for the 1993-1994 cohort of adult basic skills participants was less successful due to a relatively small pool of potential comparison group members.

Data for the study were obtained from records compiled by the workforce programs in which participants enrolled, the state employment service, and welfare and unemployment insurance offices. Data were collected in 1995. This represented a seven- to nine-month period between termination from the adult literacy education program for the 1993-1994 cohort and a three-year period for the 1991-1992 cohort.

It is important to note that the adult literacy education subjects of the study had all been enrolled in programs sponsored by community and technical colleges and had all enrolled for reasons related to work. This included only about one third of the total of adult literacy education students (including ESL) served by Washington's community and technical colleges. In addition, the study was limited to those who had enrolled only in basic skills and, therefore, it excluded learners who had also enrolled in vocational education. Finally, as the report mentions, construction of an adequate comparison group for adult basic skills was difficult because there was no information about the basic skills levels or English language proficiency of comparison group members.

Findings

Seven to nine months after exiting from adult literacy education, 52.8 percent of the comparison group were employed and 44.9 percent of the participant group were employed, the difference being 7.9 percent in favor of the comparison group. Three years after leaving adult literacy education, 49.4 percent of the comparison group were employed and 45.7 percent of the participant group were employed for a difference of 3.7 percent, again in favor of the comparison group. Participation in adult literacy education was, therefore, found to have a negative effect on employment.

Seven to nine months after terminating the adult literacy education program the mean hourly wage for participant group members who were employed was \$8.22, while for comparison group members hourly earnings were \$7.90, the difference being \$.33. Three years after termination, participant group members were earning \$9.00 and comparison group members were earning \$9.05. The difference of \$.05 was not statistically significant. Adult literacy education had a small positive short-term effect and no long-term effect on hourly earnings.

In the short-term, former participants worked an average of 355 hours per quarter and comparison group members worked 326 hours for a difference of 29 hours in favor of the participant group. In the long-term, participant group members worked 393 hours and comparison group members worked 363 hours for a difference of 30 hours in favor of the participant group. Participation in adult literacy education had a positive effect on hours worked.

With respect to mean quarterly earnings, seven to nine months after leaving the adult literacy program, participant group members earned \$2,994 and comparison group members earned \$2,635 for a difference in favor of the participant group of \$359. Three years after termination from the program participant group members earned \$3,653 and comparison group members earned \$3,361 for a difference of \$292 again in favor of the participant group. Participation in adult literacy education had a positive effect on quarterly earnings.

In the short-term, 9.3 percent of the participant group received AFDC, 23.2 percent received food stamps, and 23.9 percent received medical benefits. The figures for comparison group members were 6.3 percent for AFDC, 19.2 percent for food stamps, and 18.5 percent for medical benefits. The differences between the groups were 3 percent for AFDC, 4 percent for food stamps and 5.4 percent for medical benefits. In all cases, the incidence of public assistance was higher for participant group members than for comparison group members. Three years after termination, 6.8 percent of the participant group members received AFDC, 18 percent received food stamps, and 18.9 percent received medical benefits. For comparison group members, the percentage for AFDC was 5.9, the percentage for food stamps was 18.8, and the percentage for medical benefits was 17.8. Although the small difference of .9 percent was statistically significant for AFDC, the differences for food stamps and medical benefits were not significant. Participation in ABE had a negative effect on receipt of public assistance in both the short and long term.

With respect to benefits and costs, the report states that "the ratio of participant benefits to program costs is therefore, without considering impacts on social-welfare expenditures, \$416 to \$1,261" (p. 11).

Strengths

- The report is sufficiently clear and detailed with respect to methods, procedures, and findings.
- Direct measures, rather than learner self-report, were used.
- The study includes a relevant comparison group.
- The sample size is large and adequate for the analyses performed.
- The study measures both long- and short-term effects.

Limitations

- Although the use of a comparison group enhances the study, the comparison group is not a true control group. As the report acknowledges, the matching procedures for adult literacy education were imperfect because the adult literacy rate for the comparison group was not known. As a result, it is not known for certain whether the differences between the participant and comparison groups can be attributed to the effects of adult literacy education or should be attributed to the non-equivalence of the participant and comparison groups.
- Data were collected from program records. The report acknowledges some problems with employment and earnings data gathered from state unemployment insurance records and data collected from welfare records.
- The participant group was comprised of leaver/completers and the study does not distinguish between the two. Thus the proportion of successful completers of adult literacy education to learners who dropped out before their goals had been achieved is not known.

6. The Tennessee Longitudinal Study

Merrifield, J., Smith, M., Rea, K., & Shriver, T. (1994). *Longitudinal study of adult literacy participants in Tennessee: Year two report*. Knoxville, TN: Center for Literacy Studies, University of Tennessee.

Merrifield, J., Smith, M., Rea, K., & Shriver, T. (1993). *Longitudinal study of adult literacy participants in Tennessee: Year one report*. . Knoxville, TN: Center for Literacy Studies, University of Tennessee.

The Tennessee Longitudinal Study was one of the most ambitious state studies, and, had its ambitions been realized, it might have been the best state study yet conducted. The goals of the five-year effort were:

To expand our understanding of how participation in literacy programs changes adults' quality of life....2. To examine the influence of community and programmatic contexts of the individuals in the study, within which they change skills, perceptions and attitudes, and to explore the meaning for them of these changes, 3. To provide findings for policy makers and program developers to inform development of future adult basic skills programs. (Merrifield et al., 1993, p. 9).

The study was longitudinal in design, and a qualitative component was planned to examine changes in the individual, community, and program contexts.

The study focused on ABE level one participants, those who had scored below grade level 5.9 on the ABLE Test. The sampling plan, which used a paired comparison method, began with the selection of program sites. Six demographic variables were identified as being relevant to the study: percent non-white, percent of families living in poverty, percent population change, percent high school graduates, percent urban, and median years of school completed. Tennessee counties were then sorted into three demographic regions by rural/urban. This produced six sets of counties. Then the means for the six demographic variables were computed for each of the six sets of counties. Counties with the most variables that fell within a half standard deviation from the means on the six variables were selected as possible sites. Next, the number of level one ABE students for each potential county site was computed as a step toward developing a sample of about 240 learners. It was found, however, that only the larger urban counties had

sufficient numbers of level one students to yield a sample of this size. As a result, some smaller counties were combined. Eight county research sites were thus identified.

The data collected in the first year of the study were to serve as a baseline for longitudinal follow-up data collections in subsequent years. Data were collected through personal interviews conducted by interviewers trained by the project. One hundred and thirty-three interviews were obtained, well short of the 240 interviews intended. Several reasons for the failure to obtain interviewing goals were posited:

Several of the sites reported increases in the number of ABE Level-2 and ASE new entrants, but reported few new ABE Level-1 entrants in that year. Since only those scoring at 5.9 or below in reading qualified for the study, not all of the ABE Level-1 students qualified for the study, our initial projections (based on ABE-1 totals) were probably optimistic. We also excluded some students who were in correctional facilities, nursing homes and other students aged 75 and over, and students who were mentally retarded. (Merrifield et al., 1993, p.16)

Although the 133 learners who were included in the baseline cohort were sufficient for meeting the first year's basically descriptive project goals, it was recognized at the outset that, given expected attrition in the baseline cohort, sufficient numbers of subjects might not be available in future years to meet the goals of the longitudinal study.

In the second year, longitudinal comparisons were to be made between the baseline cohort for year one and the same subjects a year later. However, as expected, the Tennessee study experienced significant attrition between years one and two, and was left with a sample size of 70 in the second year. Most of the attrition came from one large urban site, thus skewing the second year sample. Moreover, comparisons between the first and second year data showed that the composition of the second year sample had changed substantially as a result of attrition. While for the first year sample 30 percent of the subjects resided in East Tennessee, in the second year 44 percent resided in East Tennessee. Other variables for which there were differences between years included race (first year blacks=58 percent, second year blacks=47 percent) and employment (first year employed=33 percent, second year employed=46 percent).

Unfortunately, the amount of attrition and the significant differences between the year one and year two data render meaningful inferences from the data problematic. To compound the problem, the sample size for the second year data was so small that many important analyses were impossible. Although it was intended that the follow-up data for the study would be collected one year after the baseline data, in actuality follow-up data were collected between 12 and 20 months after the baseline data. Thus the time interval between baseline data collection and follow-up data collection differ.

The Tennessee Study used two instruments to collect data. The first was a 117-item focused interview guide that included variables pertaining to socio-economic well-being, social well-being, personal well-being, and physical well-being. The survey was well-conceived, detailed, and comprehensive. Yet, because of its length, the instrument required about an hour to administer. To measure self-esteem, the Rosenberg Self-Esteem Scale was used. The Rosenberg is of acceptable validity and reliability and has been used in other studies of adult literacy education. It is one of the shorter self-esteem instruments.

Findings

The outcome findings, which pertain to changes in the first year cohort over a year's time, are found in the second year report (Merrifield et al., 1994).

Learners' Assessments of Their Participation and Additional Education

Ninety-one percent of the respondents reported that participation in adult literacy education had made a difference for them or helped them achieve their goals. Forty-nine percent reported that participation had made a difference in reading, writing, and math. Findings for "making a difference" in

the other areas listed were: learning in general, 18 percent; everyday literacy skills, 13 percent; self-confidence, 10 percent; getting/improving a job, 8 percent; helping children, 1 percent; being with people, 1 percent. Seventeen percent reported that since enrolling in ABE classes they had enrolled in other educational activities.

Employment

The overall employment gain for the group was 9 percent. Of those who were unemployed and looking for work when they enrolled in the ABE program, almost half had secured employment by the second year interview. However, 20 percent of those who were employed in year one had lost their jobs by the second year. As the second year report states:

The bad news is that all those who gained a job left the ABE program. That is bad news for a couple of reasons. First, for the most part they did not participate long enough to be likely to make great skill gains, given that they were reading at the fifth grade level or below. Second, the jobs they got were not very good jobs. Average wages were only \$4.87 per hour, worse even than the wages of those employed at the baseline (\$6.07 per hour). (Merrifield et al., 1994, p.54)

Self-Esteem

Self-esteem was measured by the Rosenberg Self-esteem Scale. Small but statistically significant changes were found in self-esteem between the baseline and follow-up data collections (baseline mean=3.63, follow-up mean=3.83, difference=.20). Differences were lower for those who were no longer active in the program (.16) than for those who were still active (.28). Differences were the least for those with minimal participation (-.13), greatest for those with moderate participation (.32), and in between for those with substantial participation (.27).

Having become recently employed proved to be an important factor in explaining differences in self-esteem. The difference in Rosenberg scores for the recently employed was .31, while the difference for the continued employed was .22, for the unemployed looking for work .21, and for the unemployed not looking for work .15. Race was not an important factor in changes in self-esteem. Women's self-esteem (difference=.31) increased substantially more than men's (.07).

In addition to receiving the Rosenberg instrument, participants were asked if their feelings about themselves had changed. Seventy-seven percent reported that they felt better about themselves, 20 percent said they felt the same, and 3 percent reported that they felt worse. Thirty-nine percent reported that being able to read better caused the change and 29 percent reported that attending the ABE program had caused the change.

It should be noted that there is a discrepancy between the changes recorded on the Rosenberg Scale and the changes in self-esteem reported by learners. While the changes recorded by the Rosenberg are quite small, the self-reported changes are quite large.

Marriage, Family, Community Involvement, Everyday Literacy, and Health

No significant changes were found in positive marital relations between baseline and follow-up. Changes were found with respect to activities with children. Although helping children with school work decreased 5 percent between baseline and follow-up (44 percent baseline, 39 percent follow-up), visiting a child's teacher four or more times increased from 28 percent to 46 percent. Checking on a child's progress in school increased from 50 percent to 75 percent. Attending school activities increased from 50 percent to 60 percent, and attending four or more school activities increased from 40 percent to 61 percent.

Although voter registration increased 5 percent between baseline and follow-up, the percent that voted in the last presidential election decreased 8 percent. Seventy-nine percent of those who reported that they had changed in the way they felt about their community reported that the change in feeling was positive. Of the eight questions related to community involvement, only one showed significant changes

(“as a citizen I can bring about needed change in government”), and that change was negative (42 percent decrease between baseline and follow-up).

Changes in attitudes toward literacy increased and, of the 84 percent who reported that they had seen changes in their everyday literacy usage, 51 percent noted increases in reading and writing. Use of the public library increased, as did incidence of visits to the local public health clinic.

Strengths

- In design, the study was sound. Because of the longitudinal design, long-term as well as short-term changes could have been ascertained had the study been conducted as planned. The qualitative component afforded the possibility of yielding new insights into the meanings learners ascribe their participation in adult basic education. The sampling plan was adequate to produce findings with external validity.
- The survey was detailed and comprehensive. Variables were sufficiently numerous and relevant to permit detailed and sophisticated analyses.
- The reports are clear and well documented.

Limitations

- Because funding for the study was withdrawn before its completion, complete follow-up data from planned data analysis cohorts were not collected. More importantly, the qualitative component was never implemented.
- Learner attrition from the study over time resulted in an unacceptably small sample size for the follow-up samples. This precluded detailed and sophisticated analyses. Learner attrition may also have resulted in biased data for the analysis of changes that occurred over time.
- The time period between the baseline and follow-up data collections varied significantly in duration.

Like the NEAEP, the Tennessee Longitudinal study is an excellent case study of what can go wrong with well-intentioned and well-designed outcome studies in adult literacy education.

7. The California Adult Learner Progress Evaluation (CALPEP)

Solorzano, R. (1989a). *An analysis of learner progress from the first reporting cycle of the CALPEP field test: A report to the California State Librarian*. Pasadena, CA: Educational Testing Service.

Solorzano, R. (1989b). *An analysis of learner progress from the second reporting cycle of the CALPEP field test: A report to the California State Librarian*. Pasadena, CA: Educational Testing Service.

CALPEP was an outcome assessment of the California Literacy Campaign conducted by the Educational Testing Service. It differs from the previously reported outcome evaluations in two important respects. First, the California Literacy Campaign is a tutor-based adult literacy education program conducted under the auspices of the California State Library System. Hence all outcomes are the product of one-on-one tutoring. Secondly, the variables in the study relate almost exclusively to reading and writing behavior. They include reading habits, writing habits, reading levels relative to reading goals, writing levels relative to writing goals, overall reading level, overall writing level, learners' perception of reading progress, effect on job status, and learners' reasons for leaving the program.

Data were collected with an intake form entitled “Where We Started.” Follow-up versions of the intake form were administered semi-annually and at exit from the program. Tutors and learners completed the forms together. Data were reported only for learners with matched data, that is, learners who completed both intake and follow-up forms. Data were collected from 53 of the 65 libraries participating in the California Literacy Campaign. This represented 733 learners from whom 354 matched sets of data were

obtained. The amount of time that the 354 learners in the study had participated in tutoring ranged from three to five months.

The first CALPEP report (1989a) presents data from the first six months of the assessment. The second report (1989b) reports data from the first year of the assessment. Findings are drawn from the second report, which reports changes over the duration of a year.

Findings

Reading frequency was measured as the frequency that 14 types of reading material were read. Differences between baseline and follow-up were recorded as follows: less often than at baseline measurement, more often, or the same. For the majority of learners, reading frequency did not change, but for those whose frequency did change, reading frequency generally increased rather than decreased. More than 20 percent of the learners increased their frequency of reading books, mail/bills/letters, labels and instructions, TV guides, newspapers, and magazines. Changes in the difficulty learners experienced in reading materials were also reported for the 14 types of reading material. Again, the largest group demonstrated no change in difficulty in reading. However one quarter of the learners reported less difficulty reading mail/bills/letters, labels/instructions/work-related materials, and books.

For changes in overall reading behavior, learners were asked to indicate how often they read materials in general outside the tutoring sessions. Response options ranged from “a few minutes” to over four hours per week. As the report states:

In general, learners at the lower-frequency end increased their outside reading habits while those at the upper end did not demonstrate much change in their outside reading habits. In effect, a plateau seemed to be reached after a certain amount of time. (Solorzano, 1989b, p. 10)

General reading frequency increased over time. As with reading behavior, the majority of learners reported no change in writing behavior over a year.

CALPEP attempted to measure learners’ self-perceptions of their reading and writing growth by the changes in the responses to five items that were assumed to measure improvement. For reading growth the response options were: I can’t read; I can read, but only simple things; I can read, but I can’t understand; I can read, but not under pressure; I can read, and I like to read. For writing the items were: I can’t write; I can write, but just letters and words; I can write, but only simple things; I can write, but I can’t spell; and I can write, and I like to write. For reading, more learners perceived they had improved than perceived that they had stayed the same or regressed. For learners who initially said they could not read, at follow-up 56 percent responded with “I can read, but only simple things” and 20 percent responded with “I can read and I like to read.”

For writing, of those who initially indicated they could not write, 36 percent improved to writing letters and words, and 20 percent improved to writing simple things. Seventy-three percent of the learners reported that participation in the California Literacy Campaign had helped them in their jobs.

Strengths

- The assessment is sufficiently simple in its instrumentation and administration to be used effectively in a tutor-based context.
- The report is adequately documented with respect to procedures and methods.

Limitations

- The findings are based on learner self-report.
- The findings pertain to tutor-based programming. Because tutor-based programming is generally one-on-one in orientation and because tutors are generally less well-trained than regular adult literacy education instructors, the study cannot be generalized to adult literacy education in other contexts.
- Matched sets of data could be obtained from less than 50 percent of the sample.

- The variables included in the study are based narrowly on reading and writing behavior. Thus no data are available on personal and social impact.
- The findings of the report are portrayed primarily in graphic format, and the few tables presented are difficult to interpret.
- Lack of a control or comparison group makes it difficult to infer causality.

8. The New Jersey Study

Darkenwald, G., & Valentine, T. (1984). *Outcomes and impact of adult basic education*. New Brunswick, NJ: Rutgers University, Center for Adult Development.

The goals of the New Jersey Study were:

1. To determine the impact of adult Basic Education in New Jersey in terms of (a) attainment of students' own goals for participation and (b) program effects on tangible indicators of social and economic well being. 2. To ascertain the nature and impact of costs and benefits for New Jersey's adult high school completion program, including a comparison of the GED and adult high school options. 3. To design a model, instrumentation and procedures for ongoing statewide student follow-up. (Darkenwald & Valentine, 1984, p. 2)

Relevant outcome variables were identified through a review of the literature and through consultation with the project's advisory board. Because of cost limitations, the study excluded ESOL and special populations such as prisoners and the mentally retarded.

For the outcome component of the study, a random sample was selected using the probability proportionate to size model. Seeking a sample of about 400 students enrolled in 10 programs, the enrollments of all New Jersey programs were divided into clusters of 40 students and each cluster was assigned a number. Then 10 clusters of 40 students were randomly selected from all the clusters. Nine programs were thus represented. One program, a very large one, was selected twice, and 80 students were entered into the sample rather than 40.

Once programs were selected, permission to collect data was sought and all programs agreed. Then, for each program, a list was compiled of learners who had enrolled no later than October 1983 and had achieved 12 hours of instruction by November 5, 1983. From the lists of eligible students, 40 were randomly selected from each program (80 from the program selected twice). Data were collected in April/May 1984; thus all learners had been enrolled seven to eight months at data collection. Data were collected through telephone interviews using a 27-item survey that included several open-ended questions that were subsequently inductively coded. The survey was field-tested with groups of 10 students and modified accordingly. Two hundred and ninety-four responses were received for an unadjusted response rate of 74 percent. When invalid cases were removed from the sample (e.g., unlisted or no phones) the adjusted response rate was 97 percent.

The benefits of the high school completion component of the study included two groups: 1. learners who had participated in publicly-funded adult literacy programs and had achieved high school certification by passing the GED tests and 2. learners who had completed the New Jersey Adult High School program. Through the New Jersey Adult High School, learners earn regular school district diplomas. For sample selection, 300 GED graduates were randomly selected from a list of those who had passed the GED tests between January and April 1982. All had completed adult literacy instruction between 14 and 18 months previously. Because a list of adult high school graduates could not be obtained, subjects were randomly drawn from five (of the nine) programs that had been selected for the outcome study and operated an adult high school. Data were collected through a short mailed survey for which the unadjusted response rate was 50 percent and the adjusted response rate was 64 percent.

Findings

Adult Basic Education

In interpreting findings it should be noted that, in its April/May 1984 data collection, the New Jersey Study surveyed all students who had enrolled in the preceding October. Thus the April/May 1984 data collection included the 60 percent of students who were no longer attending the program at that time.

Employment

Seven to eight months after they had entered the program, the net gain in employment for all respondents was 12.5 percent. Of those who were unemployed and seeking work at the beginning of the study, the net gain in employment was 16.4 percent.

Of those who were employed at the beginning of the study, 18 percent reported that they had changed jobs by the end of the study, and of these, 61 percent said that they had obtained a better job. Sixty-five percent felt that their job performance had improved, 42 percent reported that they had received a raise, 14 percent said that they had received a promotion, and 57 percent indicated that their job security had improved.

Of those who were seeking employment at the time of data collection, 79 percent reported that they believed that their participation in ABE would improve their employment prospects.

Basic Skills

Learners were asked whether participation in ABE had helped them to become better readers. Eighty-nine percent responded yes. They were then asked if reading outside class had enabled them to do something they could not do before. Two thirds responded yes. Then learners were asked to name some of the things they could now do with their acquired reading skills. One hundred and fifty-eight answered with 345 ways in which they had employed their reading skills. Reading newspapers, reading magazines, and reading books were mentioned by 20 percent or more of the learners.

Sixty-three percent indicated that classes had helped improve their writing, and use of writing outside of the classroom was reported by 49 percent of the respondents. Eighty-five percent of the learners said that participation in ABE had helped their math skills and 58 percent indicated that they used their newly acquired math skills outside the classroom. Finally, respondents were asked if participation had helped them with other things. Social studies (26 percent) and interpersonal skills (20 percent) were mentioned most often. Only 4 percent mentioned job-seeking skills.

Personal Goals

Respondents were asked to what extent participation in ABE had helped them to reach their own personal goals. Thirteen percent responded with "totally," 49 percent answered with "a lot," 26 percent said "some," and 12 percent responded "a little" or not at all. Twenty-one percent reported that the class had helped them earn high school certification, 67 percent indicated that the class had not, and 12 percent said that they already had a high school diploma.

Further Education

Sixty-three percent of the learners reported that they planned to enroll in further education and training in the future, 18 percent said they did not plan to engage in further education, and 19 percent indicated that they were uncertain. Computing (20 percent), secretarial studies (18 percent), and allied health (17 percent) were the most commonly mentioned areas of planned future study. Twenty-three percent planned to study in a community college, 19 percent planned to study in a public vocational-technical school, and 11 percent planned to study at a four-year college.

Public Assistance

Twenty-seven percent of the learners reported that they had received some sort of public assistance since the preceding October. Of these, two thirds said the amount of public assistance had remained the same, 19 percent said that the amount had decreased, and 15 percent reported that public assistance had been eliminated. Of those for whom public assistance had terminated, 44 percent said the reason was job acquisition, and 9 percent reported that the reason was increased income. Nearly half (48 percent) indicated that the reason was other than the reasons listed.

Other Outcomes

When asked in general whether they felt better about themselves as a result of attending the adult education program, 92 percent responded in the affirmative. Of those who had children, 75 percent reported that they helped children with homework more, 81 percent said that they talked to their children more about school, 73 percent reported that their children had developed a better attitude towards school, 75 percent reported that their children were getting better grades, and 50 percent indicated they had become more involved with the schools.

In closing, the New Jersey study's survey asked learners to specify the single most important benefit they obtained from participation. Thirty-nine percent indicated academic skills, 32 percent said enhanced self-confidence, 10 percent indicated a GED or high school diploma, 9 percent reported job-related benefits, and 5 percent indicated enhanced personal skills.

High School Completion*Employment and Earnings*

Of those graduates who were initially unemployed and seeking employment, 58 percent obtained employment. This gain is offset by 15 percent of graduates who were initially employed but lost their jobs. Of those who were initially employed part-time, 44 percent obtained full-time employment. Seven percent of those who were employed full-time at the outset of the study became employed part-time.

Of those who were employed both at the beginning and end of the study, 45 percent reported that they had obtained better jobs and 93 percent reported gains in earnings. In addition, 29 percent indicated that they had received promotions, 78 percent reported that they were more likely to keep their jobs, and 76 percent said that they were able to do their jobs better.

For graduates who were employed full-time at the beginning and end of the study, the increase in weekly take-home pay was \$64. For graduates employed part-time, both initially and at the termination of the study, the increase was \$34. For graduates who were employed part-time but obtained full-time employment, the gain was \$88. For the total sample, the increase in take-home pay was \$26. Average number of months employed increased from 7.7 to 9.0.

Other Benefits

Incidence of welfare enrollment decreased 45 percent. Twenty-nine percent of graduates enrolled in college and 31 percent enrolled in a trade or technical school. On a 5-point scale, the mean score was 3.7 to the question, "The classroom instruction I received was very helpful in preparing me to take the GED test."

It should be noted that the economic benefits gained in 14 to 16 months after receiving a high school diploma for the New Jersey study's population were considerably higher than those reported by other studies. Given the study's limitations, the reasons for these abnormally high gains are not known.

Strengths

- In its description and explanation of research methods and procedures, the New Jersey study is a model of clarity and completeness. Considerable attention was paid to explaining technical complexities in terms the lay person could understand.
- The response rate for the ABE study was extremely high in comparison to other studies, and the response rate for the “benefits to high school completion” component was quite good.
- The variables and survey items eliciting them were well-conceived and clear.

Limitations

- Because there was no control group it is difficult to infer that participation in adult literacy education or high school completion caused noted outcomes.
- Differences are not measured differences in a pre-survey and a post-survey. Rather, differences are based on respondents’ recall of their status in the October preceding the April/May data collection.
- All findings are based on learner self-report.
- With respect to the ABE study, although the time period covered by the study was six to seven months, the actual hours of instruction received by respondents may have been relatively short. Although the mean hours of instruction received per respondent over the duration of the study is not reported, an attrition rate of 60 percent is reported and average hourly class attendance by month is reported as: October=32 percent, November=26 percent, December=13 percent, January=14 percent, February=13 percent, and March=15 percent.

9. The Utah Study

Mahaffy, J. E. (1983). *Impact evaluation of adult basic education program outcomes: ABE Evaluation Project final report*. Bozeman MT: Montana State University Department of Educational Services.

The purpose of the Utah study was to determine the impact of participation in adult basic education. Impact was measured on two dimensions: impact on the quality of learners’ lives and financial impact. Impact on students’ lives was measured by such variables as gains in knowledge, skills, community participation, self-confidence, and further education. Financial impact was measured by gains in employment and income.

The sample for the study included two groups—currently enrolled learners and learners who had left the program within the previous five years. Because former students were defined as those who terminated the program for any reason, it cannot be assumed that they left the program because of completion. On the average, currently enrolled students had received 214 hours of instruction and former students had attained 338 hours of instruction. The time of termination of learners in the former student group varied from one year prior to measurement to five.

For both groups, the design called for random selection from a sample of programs stratified according to program size. Nine programs were selected for the study from the 19 operating at the time, and students and former students were randomly selected for participation. However, because many of those selected could not be contacted and others refused to participate, the random selection procedure failed to secure a sufficient number of subjects to conduct the study. As a result, additional students were selected in the order that they appeared on the lists used for selection. “Ultimately 516 former and current students were interviewed, 328 of the 381 former students desired and 188 of the 190 current students desired” (Mahaffy, 1983, p.18). Thus, although programs were selected randomly, the research subjects were not. Current students differed from former students on several demographic dimensions including gender and age. Data were collected with a focused survey instrument administered by personal interview at program sites and by interviewers trained by the project.

Findings

Nineteen percent of current students and 72 percent of former students reported that their goals had been achieved. The difference is not surprising given that one would expect current students to terminate once their goals had been achieved.

Basic Skills

Perceived improvement in basic skills was measured on a 3-point scale that ranged from “a lot” to “little or none.” For mathematics 52 percent of the current students and 47 percent of the former students responded “a lot.” Comparable figures for other basic skills areas were: reading (current=43 percent, former=37 percent), writing (current=43 percent, former=41 percent), and communication (current=28 percent, former=53 percent). Thus for all areas except communication, current students reported greater improvement in basic skills than former students.

Perceived improvement in life skills areas was also measured. Using the same scale, subjects responded “a lot” as follows: how to find and keep a job (current=21 percent, former=16 percent), how to get the most value from your money (current=21 percent, former=25 percent), rights and responsibilities as a citizen (current=25 percent, former=27 percent), knowledge of community resources (current=13 percent, former=24 percent), and health care for self and family (current=7 percent, former=9 percent). Current students were more likely than former students to respond “a lot” to use of the library and to reading magazines and books. Former students were more likely to participate in community organizations and to vote. There were no differences between the groups with respect to attending parents’ meetings at school and talking to school officials about their children.

Further Education

Sixteen percent of the current students had enrolled in other educational programs prior to enrollment in ABE. Thirty-six percent of former students had enrolled in other educational programs after termination from ABE. However, only 17 percent of the former students indicated that ABE had influenced their desire to enroll in other education programs. Eighty-four percent of current students reported a desire to enroll in additional educational programs and 86 percent of former students indicated a desire.

Personal Lives

Learners were asked to report on perceived changes in their personal lives on a 5-point scale with “much better” anchoring the high end of the continuum. Based on responses of “much better,” former students were slightly more satisfied with their lives in general than current students (former=63 percent, current=57 percent). Former students were also more likely to report that they were getting on well with their jobs (former=30 percent, current=18 percent), had higher levels of self-confidence (former=67 percent, current=59 percent), and were more satisfied with their family lives (former=52 percent, current=45 percent). There were no significant differences between the groups with respect to getting along with other people and feeling good about one’s self. Former students were more likely to report that as a result of participation in ABE their relationship with their children had improved and that their children were more interested in school. Seventy-three percent of former students and 21 percent of current students reported positive changes in their lives from participation in ABE.

Financial Impact

At the time of the interview, 28 percent of current students were receiving some sort of government financial assistance. The rate for former students was 25 percent. At the time of the interview, 32 percent of current students were employed and 49 percent of former students were employed. The mean annual income for current students was \$6,422 and the comparable figure for former students was \$7,483. Twelve percent of current students and 57 percent of former students reported that their financial condition had improved.

Strengths

- The report is comprehensive and clearly written. Methods and procedures are adequately explained. Findings are presented in detailed tables.
- The study included a wide range of variables.
- By and large the number of subjects is adequate.
- The study includes a comparison group.

Limitations

- All findings are based on learners' self reports. The recall time for former students could have been as long as five years.
- The comparison groups are conceptually flawed. If one wanted to measure the impact of adult literacy instruction, a group that had received no instruction would logically have been compared to a group that had received substantial instruction. However, the current student group had received adult literacy instruction. In fact, at the time of data collection, current students had received an average of 214 hours of instruction while former students had received 338. More important, the former student comparison group included students who had terminated the program for some reason. Thus this group includes successful completers as well as program dropouts and there is no way to distinguish between the two.
- Hours of instruction are not controlled.

10. The Maryland Study

Walker, S. M., Ewart, D. M., & Whaples, G. C. (1981). *Perceptions of program impact: ABE/GED in Maryland*. [Paper presented at the Lifelong Learning Research Conference, College Park, MD, February 6 and 7, 1981]

The purpose of this study was to assess learners' perceptions of ABE/GED program impact in Maryland. The subjects were 120 participants in ABE/GED programs who, at the time of data collection, had received an average of six months of instruction. Forty-five percent of the subjects had been enrolled for three months or less. It was originally intended that five subjects would be randomly selected from a series of ABE programs stratified according to program size in three Maryland counties, but because some selected subjects refused to participate and others requested to participate, the random selection protocol was not strictly followed. Of the 120 subjects who volunteered for the study, adequate data were collected from 116.

Data were collected by graduate student interviewers using an instrument based on prior research and discussions with learners, administrators, and teachers. Interviews ranged from 13 minutes to an hour and averaged about 20 minutes. Interviews were conducted near the end of the semester, and at that time, some classes were so small that another class had to be selected from the same center.

Findings*Employment*

Of the subjects who were unemployed and the 20 percent who were homemakers, 84 percent desired to obtain employment in the future and 85 percent believed that participation in ABE would help them do so. Of those who were employed, two thirds believed that ABE would help them gain promotion. 63 percent perceived themselves as being able to do their jobs better.

Basic Skills

Seventy-six percent of the students reported that they could read better as a result of participation, 81 percent perceived that they could write better, 90 percent believed that their computational skills had improved, and 63 percent reported that they were better shoppers. Math was the most commonly mentioned most useful thing learned (52 percent).

Community Involvement

About half the participants reported that they belonged to at least one community organization, although 70 percent of the learners said that participation in ABE had not affected their membership in community organizations. Half the subjects indicated that they had recently used the library. Fifty-three percent expressed some interest in politics and 35 percent claimed that they were more interested in politics since having participated in ABE. Fifty-eight percent reported that they were more aware of community services.

Attitudinal Change

Eighty-nine percent of the students reported that they felt different about themselves as a result of participation. Of those with school-age children, 52 percent reported that they were more confident about discussing problems with their children's teachers.

Personal Relationships

Of the 86 percent of respondents who had children and reported that their children asked for help with homework, 61 percent indicated that they were more confident about giving help with homework.

Continuing Education

Ninety-one percent of the learners said that they planned to continue their education once they had completed the ABE program.

Strengths

- The report is clearly written. Methods and procedures are adequately described and the report is honest about its limitations.
- The study is comprehensive with respect to the variables studied.

Limitations

- Because the research subjects were not selected randomly, and because all subjects came from only three Maryland counties, the results cannot be generalized to Maryland or to any larger unit.
- Almost half the research subjects had been enrolled for three months or less. It is questionable whether there was a sufficient amount of instruction gained between enrollment and data collection for significant impact to be measured.
- The study lacked a control or comparison group. Thus it is questionable to infer that participation in ABE caused the changes noted.
- All findings are based on learners' self reports.

11. The 1980 Tennessee Study

Jones, P., L., & Petry, J., R. (1980). <i>Evaluation of adult basic education in Tennessee, 1980</i> . Memphis, TN: Memphis State University, College of Education.

The 1980 Tennessee Study measured learners' perceptions of their educational experience in Tennessee ABE programs. Data were collected with a 26-item instrument. Variables were selected from the research literature and responses were recorded on 5-point Likert scales. Variables included self-expression, self-concept, family life, life in general, leisure, relationships with others, and relationships with society. Thus, unlike most of the other impact evaluations assessed in this report, the 1980 Tennessee Study focused primarily on the affective dimension.

To select a sample, first a sample of programs across the state was identified (the report does not say how). Then each program was mailed a packet of surveys and the supervisor was asked to identify 25 students who would receive the survey. Each supervisor was asked to administer the survey personally. Information was received from 72 of the 89 programs then active in Tennessee. From a potential of 2,225 responses, 1,623 responses were received for a response rate of 73 percent.

Findings

Learners scored 3.9 on the 5-point scale measuring self-expression. Students who had been in the program longer, who were more affluent, and were older scored higher on this dimension. Students scored 3.9 on the scale measuring self-concept. Those who had participated in the program longer than 18 months scored higher on self-concept. Subjects scored 3.7 on the scale that measured whether they perceived that the ABE program had helped them to be more confident about their family relationships. Again, learners who had participated in the program longer scored higher. A composite of six items measured life in general, and subjects scored 4.0 on this measure. On the scale that measured whether learners' interests had expanded to make their leisure time more meaningful, the mean score was 4.0.

With respect to relationships with others, learners scored 4.0 on the scale used to measure this dimension. The mean score for the scale that measured social responsibility was 3.8. Students scored 4.1 on the scale that measured their perceptions of the value of their education.

Strengths

- In general, the report is clear, and methods and procedures are adequately reported, although descriptions of methods are less detailed than in most of the other studies reviewed here.
- The sample size is adequate ($n=1,623$), and the 73 percent response rate to the survey is relatively high.

Limitations

- Rather than being randomly selected, the subjects for the study were selected by ABE supervisors. This interjects the possibility of selection bias.
- Data from specific survey items seem to have been collapsed into the summative scale scores reported, and this results in a loss of detail.
- All data are based on self-report.
- There is no control group.
- All subjects were enrolled ABE students. Because 91 percent had been active in the program for six months or less, it is questionable whether there was sufficient time for substantial impact to occur. Because all subjects were enrolled students, data were not collected from those who had terminated the program either because they had achieved their goals or had dropped out.

12. The Ohio Study

Boggs, D. L., Buss, T. F., & Yarnell, S. M. (1979). Adult basic education in Ohio: Program impact evaluation. *Adult Education*, 29(2), 123-140.

The objective of the Ohio study was to determine the extent to which the goals of the Adult Education Act, the federal legislation that funded ABE in Ohio, were being achieved. Specifically the study sought to ascertain whether, as a result of participation in ABE programs, learners' occupational

status had improved, whether learners had been assimilated into society, and whether learners' personal goals had been met.

The population for the study was learners who had terminated ABE in 1973-1974. Follow-up data were collected three years later. Data were collected with two telephone surveys, one of which was administered to former students, the other of which was administered to adults who were eligible for ABE but had never participated. Although the comparison group is an important feature of the study, it is important to note that the comparison group is not a true control group as there was no random assignment to instruction and to control groups. Indeed, because the comparison group was representative of the general population eligible for ABE, and because the treatment group was representative of ABE participants, the two groups are different with respect to socio-demographic composition.

A multistage sample design was used to create a sample representative of the approximately 27,000 ABE students in 1973-74. First, a stratification protocol was designed based on ABE program size, municipality size, gender, age, and race. Then, from these strata, 12 ABE programs were randomly selected to participate in the study. Of the approximately 3,500 former ABE students who had been enrolled in these programs, 1,200 had valid telephone numbers or addresses. Data were collected by telephone by trained interviewers in April and May of 1977. Three hundred and fifty-one valid responses were received. The average completion time per interview was 13.6 minutes and, on average, it took four telephone calls to secure a usable survey. The estimated sampling error was 5.3 percent.

A multistage sampling design was also used to establish the comparison group of adults who were eligible for ABE but had not participated. Initially the comparison group population was stratified into 88 counties. Then the counties were grouped into 11 sampling areas in order to allow for urban-rural distribution. From telephone books acquired by the study, 1,500 comparison group subjects were randomly selected. Subjects were surveyed by telephone between May and July 1977. Over 90,000 calls were made at an average completion time per interview of 13 minutes. The completion rate was 13 to 1 and the estimated sample error was 2.4 percent.

Findings

Involvement

Three years after they had left the ABE program, former participants scored significantly higher than non-participants on use of the library, reading magazines, use of social services, community activity, reported self-confidence, and communication. Non-participants scored higher on continuing their friendships.

Employment/Home Ownership

Former ABE students scored higher than non-participants on incidence of employment, being promoted, and perceived job security. Non-participants were more likely to own their own homes than former ABE students, although the difference between the groups was small.

Children/School

Former ABE students scored higher on attending parents' meetings. The differences between former ABE students and non-participants were not significant with respect to helping children with homework. Non-participants scored higher than former ABE students on communication with the schools.

Voting

Although former ABE students scored higher than non-participants on being registered to vote, there were no significant differences with respect to having voted in 1972 or 1976.

Education

Former ABE students were more likely to be enrolled in school and to be contemplating future enrollment than non-participants.

Learners' Personal Goals

Of those who stated that their goal at enrollment in the ABE program was to improve their English (27 percent of the sample), 96 percent reported that they had achieved the goal. Of the 77 percent of the sample who had as their goal improving math, 97 percent reported that they had achieved this goal. Sixty-two percent of the sample indicated that improving their reading was a goal at entry; 96 percent reported that they had reached this goal. Of the 62 percent of the sample that indicated that obtaining a GED was one of their goals, 40 percent reported goal attainment.

Strengths

- The study design was systematic, and the design and procedures are clearly described in the report.
- The variables studied are comprehensive.
- The number of subjects is adequate for detailed analysis.
- The study includes a comparison group.

Limitations

- Because the Ohio study is one of the few that includes a comparison group, the limitations of the Ohio comparison group warrant discussion. There are two potential reasons for including control or comparison groups in an impact design: 1. to enable researchers to conclude that participation in adult literacy education caused a particular impact and 2. to compare performance of adult literacy education students with a particular reference group with known characteristics. In the Ohio study, former ABE students were compared with non-participants. Unfortunately, this comparison does not allow us to infer that ABE caused the differences between the groups that were found. The reason is that non-participants differ substantially from participants in ways that have nothing to do with ABE instruction. For example, if Ohio is similar to the rest of the country, adults who are eligible for ABE but have never participated are likely to be significantly older than participants. Are the differences noted in the study due to participation or are they the result of the differences in age between the two groups? We do not know, and, not knowing, the inference that ABE caused the difference cannot be made with great confidence.
- Although use of telephone interviews appeared to be an efficient way to collect data, those without telephones or valid addresses could not be reached. Of the 3,500 former ABE students selected for the study, only 1,200 had valid phone numbers and only 351 were eventually reached. Because only about 1 in 10 of the original sample could be reached, the potential for response bias is high.
- All data were based on self-reports, and the recall period for former ABE students was three years.
- The study does not control for the amount of instruction learners received.

13. The Wisconsin Study

Becker, W. J., Wesseliuss, F., & Fallon, R. (1976). *Adult basic education follow-up study 1973-75*. Kenosha, WI: Gateway Technical Institute.

The objectives of the Wisconsin study were to measure student perceptions of the program, to identify economic and employment impact, to assess learners' attainment of personal goals, to assess further education and social impact, and to measure exit level reading ability.

The study's description of design and procedures begins with the acknowledgment that certain contextual features of adult literacy education had to be taken into account in the design of the study. They were: lack of uniform grade levels, non-English speaking learners who were not literate in their native language, varying attendance, open enrollment, multiple goals for participation, and varying reading ability. Although these contextual constraints are well known by most experienced researchers of adult literacy education, the Wisconsin study is one of the very few that has acknowledged the problems these features pose in study design. The authors of the study note that, although a control group would have been desirable, establishment of a control group was beyond the means of the study and that, while collection of pre-data at the time of enrollment would have been desirable, such data were not available to the researchers.

Data were collected from a sample of Gateway participants who had terminated in 1973, 1974, and 1975. Gateway was a comprehensive adult literacy program that operated learning centers and classes in prisons, communities, training schools, institutions for the handicapped, and ESL. Of the learners sampled in the study, 133 attended learning centers, 72 attended classes, and 65 attended both learning centers and classes. In 1973 1,340 learners were enrolled, in 1974 1,133 learners were enrolled, and in 1975 the program served 1,703 learners.

To select a sample, first a 10-percent sample was randomly drawn from participant lists, and the 111 students who were still active in the program were eliminated. The resulting sample was then divided into four groups based on hours of instruction attained (category 1= 0 to 24 hours, category 2=26 to 50 hours, category 3=51 to 100 hours, category 4=100 or more hours of instruction).

Each selected student was notified of the study by letter and was told that an interviewer would contact them if they indicated by return postcard that they were willing to participate in the study. Terminated students who volunteered were placed on one list ($n=306$). In addition two other 10-percent samples were randomly drawn to serve as replacements if students from the first list could not be contacted ($n=612$). From the initial list and the replacement lists, a total of 593 former students were contacted, and in total, 270 usable interviews were completed (response rate=.45). Sixty-four learners had completed in 1973, 95 had completed in 1974, and 111 had completed in 1975. In 1973, 67 percent of the sample had 50 or less hours of instruction. The figures for 1974 and 1975 are 40 percent and 49 percent respectively.

The survey instrument was based on input from teachers, aides, and counselors and researchers. Most interviewers were ABE professionals and paraprofessionals who were trained in a one-day session. Data were collected between March and June of 1976. In total, of 589 anticipated interviews, 273 were completed and 319 could not be completed. Seventy percent of the non-completions were due to the inability to secure up-to-date addresses. As part of the interview, to measure reading level respondents were asked to take a one-page excerpt from the reading section of the Wide Range Achievement Test (WRAT). Interviewers were trained in the administration of the test.

Findings

Learners' Personal Goals

Taken as a whole, 64 percent of terminated learners for all three years rated the program as having been very useful, 34 percent rated it as helpful, and only 3 percent rated it as not being helpful. Those with 100 or more hours of instruction were most likely to rate the program as being very helpful. Improving reading, math, and writing were the three most commonly given goals of participation. Ninety percent reported that the program had helped them in reading, 82 percent indicated that the program had helped them in math, and 83 percent said that they had been helped in writing. Of the 44 percent of the sample that indicated that earning a GED was one of their goals, 27 percent reported that they had passed the GED tests. Of the 152 respondents who reported that they wished to enter another educational program upon completion of the ABE program, 33 percent did enter another program, and 54 percent of them responded that the ABE program had adequately prepared them for the additional educational experience. When

asked if they had completed what they wanted to accomplish, 62 percent reported that they had not. Eighty-two percent indicated that they were satisfied with their accomplishments.

Employment

The most frequently mentioned job-related reason for participation (34 percent of sample) was to learn a specific skill for a specific job. Thirty-six percent of those who indicated this reason reported that they had learned a skill sufficiently to qualify for the job they wished. Half the respondents had applied for jobs since termination from the ABE program, and of these, 81 percent had had job interviews. Of those who had applied for jobs since termination, 82 percent said they had had either full-time or part-time employment. Of those who had been employed since termination, 41 percent reported that ABE did help them get a job. Twenty-three percent reported that they terminated the ABE program because they obtained a job. In total, two thirds of the respondents reported no change in employment status since terminating from the program. Twenty-three percent increased employment and 11 percent experienced a decrease in employment.

Family Relationships

Learners with school-age children were asked if participation in ABE had helped them understand their children's teachers and school better. Sixty-four percent replied in the affirmative. Sixty percent indicated that participation had not encouraged their participation in teacher conferences and PTA, and 57 percent responded that participation had not changed the way they worked with their children. Twenty-nine percent responded that, since participating in ABE, household members got along better with each other.

Perceptions of the Value of the ABE Program

Learners were asked to indicate the subject that had been most helpful when they were enrolled in ABE, and, of all the subjects listed, math (17 percent) and English (16 percent) were considered to be most important. Forty-four percent of the respondents reported that ABE had helped them to become better off financially; 41 percent mentioned that ABE had helped them manage money better. When students were asked to specify the most important changes from participation, improving self-confidence (23 percent) was the most commonly mentioned change and overcoming shyness (21 percent) was the second most commonly mentioned change.

Community Participation

Eighty-four percent of the respondents indicated no change in their voter registration status, and of those who did become registered to vote, only 5 percent said that they had done so because of ABE. Eighty-five percent indicated no change in community group participation. Of those who did increase community group participation, only 4 percent attributed the change to ABE.

Reading Gains

As part of the interview, respondents were administered the reading portion of the Wide Range Achievement Test (WRAT). These scores were compared with earlier scores obtained at both enrollment in the program and shortly before termination. Seventy-eight percent exhibited an increase of grade level between enrollment and follow-up and 19 percent showed a decline. Seventy-three percent of the respondents reported that they purchased newspapers, magazines, and books prior to participation in ABE, while 84 percent said they did so at follow-up. However, 62 percent indicated that participation in ABE had not influenced their purchase of these reading materials. Fifty-five percent said that they read more since the time they entered ABE, 37 percent said they read the same, and 8 percent said they read less.

As was mentioned in the description of methods and procedures, the sample was divided into groups based on the number of hours of instruction attained. Although data by these groupings are reported throughout the findings, they have not been reported here for simplicity's sake.

Strengths

- The report is extremely detailed in both the descriptions of methods and procedures and in its findings.
- The study includes a multitude of impact variables.
- Sampling procedures are adequate.
- The number of subjects is adequate for most analyses.

Limitations

- All data, except for the WRAT scores, are based on self-report.
- The validity and reliability of the one-page excerpt from WRAT is not known.
- There is no control or comparison group.
- Much of the data pertain to a comparison between learners' goals at the time of entry and at follow-up. However, rather than collecting learner goal data at entry, the study asked respondents to recall their goals at the follow-up data collection. Since follow-up data were collected in 1976, the recall period for the 1973 cohort was three years and for the 1974 cohort it was two years.
- The response rate of .45 interjects the possibility of response bias.
- In the 1973 cohort, two thirds of the former students had achieved 50 or less hours of instruction. This raises the question of whether an adequate amount of instruction had been gained for changes to be measured.
- The findings cannot be generalized beyond the Gateway service area.

WELFARE STUDIES

With the establishment of the federal JOBS program in 1988, the numbers of adult literacy education learners served in welfare-sponsored and -funded programs increased dramatically. Although the specific details and provisions of JOBS programs varied by state, and to some extent by county, a common provision of JOBS-funded adult literacy education was mandatory participation for those who were deemed in need of basic education. In addition, JOBS clients also received support services, such as transportation and child care, and were required to participate in employment readiness activities such as job search. JOBS clients who were not eligible for adult literacy education were either enrolled in vocational education or slotted directly into employment. Most JOBS adult literacy education activities were subcontracted by the welfare system to the traditional providers of adult literacy education—public schools and community colleges.

The relationships between education/training, welfare participation, and work were hotly debated during the national discussion that surrounded welfare reform in 1996, a point of contention being whether it was better to invest in substantial education and training prior to requiring welfare recipients to acquire employment, or whether it was more socially beneficial to move welfare recipients directly into employment with education and training coming afterwards. The Personal Responsibility Act of 1996, which established major reform of the nation's welfare system, promoted the work first, educate second principle as national policy.

The purpose of this section of the report is neither to debate welfare policy nor to discuss how the research literature informs that debate. Rather, the purpose here is to analyze the research conducted on the impact of adult literacy education within welfare-sponsored programs for its contribution to a general understanding of the impact of adult literacy education.

14. The California GAIN Study

Martinson, K., & Friedlander, D. (1994). *GAIN: Basic education in a welfare to work program*. New York: Manpower Development Research Corporation.

The above report is one of several issued by the Manpower Development Research Corporation (MDRC) to convey the findings of its evaluation of the California GAIN program. GAIN (Greater Avenues for Independence) was California's program for increasing employment and reducing dependence

on AFDC (Aid to Families with Dependent Children). GAIN, which was initiated in 1986, became California's JOBS initiative in 1989 and accounted for 13 percent of all federal money allocated to JOBS. MDRC's evaluation of GAIN is exceedingly detailed and refined in its methodology and procedures, and for that reason represents a very important contribution to our understanding of adult literacy education's impact. However, the fact that the study focused on a special population of adult literacy education—clients who were eligible for welfare and were required to participate in adult literacy education—is a limitation that precludes the generalization of its results to adult literacy in general.

Enrollment in the GAIN program began with an initial assessment of eligibility. According to the law, single welfare recipients with children under age six were exempt, but could volunteer for the program. After registration, clients attended an orientation and assessment, which included administration of the CASAS test. Those who scored below 215 on either the CASAS reading or math tests, lacked high school certification, or could not speak English well were assessed as being “in need of basic education” and assigned to programs for either GED preparation, ABE, or ESOL, depending on their skill levels. Participation was mandatory; clients who failed to participate could incur a reduction of their benefits. After completion of basic education, clients participated in a formal assessment where an individual employment plan was developed that could include additional training, unpaid work, or supported work. Most adult literacy education instruction was contracted by GAIN to California's public schools and community colleges.

The GAIN evaluation was conducted in six counties: Los Angeles, Riverside, San Diego, Alameda, Tulare, and Butte. Although there were similarities among the counties with respect to implementation of the law, the law was sufficiently flexible to allow for differences in administration. Together these six counties accounted for about one third of the entire California caseload.

The evaluation employed a random assignment design. Between 7 and 14 months after the county implemented GAIN, all clients who were determined to be mandatory registrants and who attended an orientation session were randomly assigned to either a treatment group, which received GAIN services and was subject to the participation mandate, or to a control group, which did not receive GAIN services and was not subject to the mandate. Control group members were permitted to enroll in adult literacy not sponsored by GAIN if they wished to.

Data Sources included:

- *Baseline data* (n=21,596). Included basic demographic information and CASAS test scores.
- *AFDC and UI records* (n=21,392). Included data on welfare payments and employment earnings.
- *Registrant survey data* (n=3,210, 2,551 valid cases). Administered to a stratified random sample of treatment and controls by phone or in person two to three years after random assignment. The survey measured self-reported participation in a number of activities as well as GED acquisition. Subjects were paid \$25.
- *A literacy test (TALS)* (n=1,719, 1,119 valid cases). The Test of Applied Literacy Skills was administered in the client's home two to three years after random assignment. Subjects were paid \$35. Data were collected from treatment and controls who were not ESOL students.
- *GAIN Program Tracking Data* (n=4,523). Recorded clients' patterns of GAIN participation from case files.
- *Attendance data* (n=1,222). Data on attendance were collected from educational providers.
- *Staff Activities and Attitudes Survey*. This survey was administered to GAIN staff twice, one and two years after GAIN began.
- *Field Research*. Interviews with administrators and teachers to provide descriptive information about GAIN program operations.

Findings

Participation in adult literacy education was mandatory for the GAIN treatment group and was not mandatory for the control group. As might be expected, for the two- to three-year period, the participation rate in ABE/GED for the treatment group (30.2 percent) was significantly higher than for the control group (5.8 percent). Differences in the participation rates of the two groups for vocational (9.4 percent treatment, 9.5 percent control) and post-secondary education (11.6 percent treatment, 9.8 percent control) were not significant. The same relationships were found in length of participation as measured by both the number of months of participation and the number of hours of participation.

GED Attainment

Over the two- to three-year follow-up period, GAIN treatment group participants were considerably more likely to have earned a GED or other high school certification (treatment=7.2 percent, control=1.3 percent), and, not too surprisingly, success in obtaining high school certification was found to be highly related to the learners' initial educational levels as measured by the CASAS test. Perhaps the most striking finding regarding high school certification came from Tulare County, where the difference between treatment and control groups with respect to earning high school certification was 19 percent. This compares with a difference of only 7.7 percent for the next highest county, Alameda. Why was Tulare County's performance on high school certification so outstanding in comparison to the other counties? Although there is no definitive answer offered by the report, the authors note that this "may indicate that the package of educational services offered in Tulare—the GAIN program's emphasis on education, the counseling and close monitoring GAIN provided, and the education services offered by the education providers—contributed to the large attainment impacts in Tulare" (Martinson & Friedlander, 1994, p. 118).

Tested Learning Gain

Learning gain was measured by the document and quantitative literacy components of the TALS. The document literacy section (26 items) measures the skills needed to work with documents such as use of charts and forms. The quantitative literacy section (23 items) measures the ability to perform calculations embedded in text. Although the TALS includes a prose literacy section, it was not used in the GAIN study.

The difference between the treatment group and control group on the combined document and quantitative portions of the TALS was a statistically non-significant 1.8 points, suggesting that, on the whole, no meaningful tested learning gain had occurred two to three years after random assignment to groups. During that time, treatment group members had received an average of 251 scheduled hours of instruction. Differences between counties were striking, however. Although the control groups in Riverside and Tulare actually outperformed the treatment groups, the treatment group in San Diego outperformed the control group by a very sizable 33.8 percent. Commenting on this, the report states that "the county differences would appear instead to reflect differences in the implementation of the basic education program across counties" (Martinson & Friedlander, 1984, p. 131).

As with acquisition of high school certification, initial educational achievement as measured by the CASAS had an important impact on tested learning gain. For those whose initial CASAS scores were 215 or above on both the reading and math tests, the difference between the treatment and control groups was 17.8 points, 19 percent of a standard deviation, while the differences for those who scored below 214 for both tests were -17.1, indicating that the control group outperformed the treatment group.

Strengths

- The report is very detailed in its descriptions of methods, procedures, and findings. Limitations are discussed in-depth.
- The study utilized a true experimental design; subjects were randomly assigned both to treatment and control groups. As a result, solid inferences regarding causality can be made.
- The duration between treatment and measurement (two to three years) is sufficiently long for meaningful impact to have occurred.

- Inferences are based on tested gain rather than on learner self-report.

Limitations

- The subjects of the research were a special population of adult literacy education learners (welfare clients). Participation in the program was mandatory. Hence the study can only be generalized to this population.
- Only two impact variables were studied—tested learning gain and acquisition of high school certification.
- The Test of Applied Literacy Skills used to measure learning gain may not have been sensitive enough to measure the learning gains that occurred.

Comment

The California GAIN study is one of the very few that used a true experimental design, and for this reason we may be more confident that adult literacy education instruction caused the reported impacts than for other studies we have thus far reviewed. However, experimental design studies are not without their own weaknesses. Paramount among these is that we do not know why or how these impacts were achieved. Thus we are left with an unresolved ambiguity. Why did Tulare County produce such high gains on acquisition of high school certification and poor gains on tested learning gain? Likewise, why did San Diego County show very high gains on tested learning gain and low gains on high school certification? Although the authors of the report speculate that to some extent the answers lie with the manner in which the two programs were conducted, we do not know what the relevant program differences were.

The authors of the GAIN study concluded that significant gains were not made in basic skills, and this appears to be true in the aggregate. However, given the large gains in San Diego, an equally reasonable conclusion might have been that gains were made in some programs, but not in others. The conclusion on GAIN, then, depends somewhat on whether one considers the proverbial cup to be half full or half empty.

15. The Texas JOBS Program Evaluation

Center for the Study of Human Resources. (1994). *Texas JOBS program evaluation: Final Report*. Austin, TX: Center for the Study of Human Resources, Lyndon B. Johnson School of Public Affairs, University of Texas at Austin.

Schexnayder, D. T., & Olson, J. A. (1995). *Texas JOBS program evaluation. Second year impacts*. Austin, TX: Center for the Study of Human Resources, Lyndon B. Johnson School of Public Affairs, University of Texas at Austin.

Like the California GAIN evaluation, the Texas JOBS evaluation is a sophisticated and rather complex assessment of a state welfare program. Although much of the evaluation focuses on the economic benefits that accrue to JOBS clients and to the state, the effects of participation in the adult literacy education component of the JOBS program are addressed.

In Texas, during its first two years (1991-1992), 84,000 AFDC recipients were enrolled in the JOBS program. Almost all were women. Upon initial assessment, JOBS clients were separated into three service levels. Level I clients possessed high school certification, had recent job experience, and had few barriers to employment. They were certified as being job ready. Level II clients, who were designated as being less job ready, lacked high school, had less recent job experience, and had more barriers to employment. Level III clients had very little education and substantial barriers to employment, and, because helping them was beyond the resources of the program, they were not served by JOBS. Although Level I clients were generally referred to job readiness and job search activities, they were permitted to participate in education and/or job skills training if they wished. Level II clients were generally assigned to adult education and survival skills training before they received job skills training. More than one third received child care assistance and three fifths received transportation assistance.

In the 1994 report, program impacts were measured a year and a half after program entry. The evaluation used a quasi-experimental design. A treatment sample was created by randomly selecting six samples of 3,300 JOBS clients by calendar year quarter starting with the first quarter of 1990 and ending with the second quarter of 1992 (total $n=19,854$). Clients who entered the program during specific quarters are termed cohorts in the study, and much of the analysis is presented by cohorts.

For the purposes of the study, a client was defined as a female recipient of AFDC with at least one hour of participation in any JOBS component. To actualize the quasi-experimental design, a matched comparison group of approximately 20,000 was selected from non-JOBS AFDC recipients. This was achieved by first matching exactly treatment group members with non-JOBS AFDC clients on Department of Human Services region, service level, race/ethnicity, groups of employment services, and presence on AFDC rolls during the quarter in question. Then the sample was adjusted with a procedure that paired comparison group members with treatment group members who were similar with respect to age, number of children, age of youngest child, and total time on AFDC. A comparison between the treatment and comparison group on these variables showed no significant differences.

Consistent with an evaluation of a welfare program, the impact (dependent) variables used included probability of exit from AFDC, probability of exit from AFDC to employment, probability of employment regardless of AFDC exit, quarterly earnings, and probability of AFDC recidivism. The impact of education, which was measured by the number of hours of participation in high school, GED preparation, ABE, post-secondary education, and self-initiated education, was analyzed through multiple regression analysis. Multiple regression analysis is a statistical procedure that measures the impact of each of a series of independent variables on a dependent variable. In this case, the dependent variables were each of the impact variables listed above. The independent variables were the various components of the JOBS program. They were: education, whether the client had only received assessment, life skills training, job training, job search activities, job assessment, and whether the client was currently engaged in any JOBS activities. Data were collected from JOBS, JTPA, and other program records.

The findings reported here are from the second year impact report (Schexnayder & Olson, 1995). Because about one third of the comparison group members had entered the JOBS program by the time the second year study was conducted, they ($n=6,460$) and their counterparts in the treatment group were eliminated from the study. For the second year sample, the treatment group sample size was 13,196 and the comparison group sample size was 13,303. Impacts were measured between two and two and a half years after entry into the program.

Findings

Thirty-two percent of the Level I clients and 59 percent of the Level II clients had received education. It should be noted that education included high school, GED preparation, ABE, and post-secondary education. Thus it is impossible to identify the specific impact of any of these components.

Termination of AFDC

In total, participation in JOBS had either a negative or statistically non-significant effect on termination of AFDC when measured four quarters after entry. However, "By the last quarter measured, JOBS participation increased AFDC exits by 4-16 percent in all but one group" (Schexnayder & Olson, 1995, p. 12). For both Level I and Level II clients, participation in education had a statistically significant impact on exit from AFDC. Although job training also had a significant impact, life skill education, job search, and job assessment did not.

Exit from AFDC to Employment

A consistently greater percentage of treatment group members than of comparison group members exited AFDC for employment. For both Level I and Level II clients, the differences between the treatment and comparison group were in the 1 to 8 percent range, and the differences tended to increase

over time. Participation in education and in job training had a statistically significant impact on exit from AFDC to employment. Participation in life skills did not.

Employment Regardless of AFDC Exit

One year after entry to the JOBS program, the employment rates for the treatment and control groups were similar and ranged from 38 to 45 percent. Over time, however, the treatment group outperformed the comparison group in all cohorts except one. The impact of education on employment was statistically significant for Level I clients (those who were certified as being job-ready) but not for Level II clients, and the impact of job training was significant for both levels. The impacts of life skills education, job search, and job assessment were not significant.

Quarterly Earnings

Earnings for both treatment and comparison group members increased over time. After 10 quarters, JOBS participants' earnings were significantly higher than comparison group members' earnings for 5 of the 10 cohorts. Adult education had a positive and significant effect on earnings for both Level I and Level II clients, as did job training. The impacts of life skills education, job search, and job assessment were very small and mostly not statistically significant.

Return to AFDC

JOBS participants were more likely to return to welfare after they exited than the comparison group. However, education, job training, and job search reduced recidivism for Level I clients. Participation in job training was the only variable that reduced recidivism for Level II clients.

Strengths

- Although it is complex, the report adequately describes methods, procedures, and limitations.
- The sample is representative and quite large.
- The design includes a relevant comparison group.
- The measures are direct rather than based on self-report.

Limitations

- The research subjects are drawn from a special population (welfare recipients) and thus the findings cannot be generalized to adult literacy education in general.
- Because education is defined as an amalgam of high school, GED preparation, and post-secondary education, the separate impact of these components cannot be determined.

16. Steps to Success

Reder, S., & Wiklund, K. R. (1994). *Steps to Success: Literacy development in a welfare-to-work program*. Portland, OR: Northwest Regional Educational Laboratory.

Established in 1990, Steps to Success was Oregon's largest JOBS program serving clients at two locations—in Multnomah County and the Portland metropolitan area. It was administered by Mt. Hood Community College. After screening, which included a one-hour interview and the Basic Adult Skills Inventory System (BASIS) test, clients were placed in one of two tracks. Those who were assessed to have marketable skills were placed in the Placement Track and participated in a mix of job training, job search, and placement activities.

Those who were deemed to be in need of adult literacy and job training were placed in the Career and Life Planning Track (CLP). Those assigned to the CLP track generally first participated in a four-week course designed to facilitate job readiness and future self-sufficiency. After the initial course, those whose BASIS tests indicated that they needed literacy education participated in ABE/GED. Attendance was

required for a minimum of 15 hours per week. One-on-one instruction in a learning lab and small group instruction was available.

The Steps to Success outcome study included both qualitative and quantitative components, but, because the qualitative component is not extensively reported, the case analysis will focus on the quantitative. The study commenced with the development of a sample. To this end a list was secured of all clients who had participated in ABE/GED ($n=920$). The records for each case included names and addresses, basic demographic information, the BASIS score from intake assessment, and participation in Steps to Success Activities. Letters were then sent to all members of the list inviting their participation in the study and promising a \$25 stipend to those who agreed. Eighteen data collection sessions were offered for the research subjects. Subsequently a second mailing was sent and six additional data collection sessions were scheduled.

During data collection, subjects were administered a questionnaire that compiled participants' activities and perceptions including current education and employment status, additional education and employment since leaving Steps, perceptions of how Steps had helped them in various aspects of their lives, whether Steps might have prepared them better, and perceptions about improvements in their basic skills.

Two tests were employed. The baseline test, the BASIS, is an adaptation of the CASAS Test specially prepared for Steps by CASAS. The second test administered during data collection was the CASAS (level C) itself. Data collection occurred between one and three years after the clients had terminated Steps.

One hundred and nine of the 920 letters sent were returned by the Post Office. Of the remaining 811, 229 agreed to participate in the study, for a response rate of 27 percent. Of those who agreed, data were eventually collected on 163. Comparison of demographic information of the sample to county welfare clients showed that the sample was more predominantly female and less predominantly minority than the general welfare population. A comparison of participants in the study with those who elected not to participate showed that participants were more likely to be female, less likely to be minorities, and had slightly higher initial BASIS test scores.

Findings

Learning Gains

Because the BASIS test and the CASAS test differed in the ranges they measured accurately, the researchers reported two sets of figures: gains for the entire group and gains only for those who fell within the accurate range of the test. For reading, only 53 persons fell within the range of the test. Mean scores were 231.3 on the BASIS pretest and 235.5 on the CASAS posttest for a gain of 4.2 ($p < .001$). Of the 53 subjects, 37 showed gains in reading, 3 showed no difference, and 13 demonstrated declines.

For math, the entire group had pretest scores of 221.1 and posttest scores of 229.0 for a gain of 7.9 ($p < .0001$). One hundred and thirty-five of the 163 cases showed gains, 5 showed no difference, and 23 demonstrated declines. For the group that fell within the accurate range of testing on math ($n=127$), the mean pretest score was 217.9 and the mean posttest score was 226.7 for a gain of 8.8 ($p < .0001$). One hundred and thirteen of the 127 showed gains, 5 showed no difference, and 9 had declines.

Of the 79 subjects who lacked a GED when they entered Steps, 22 (28 percent) had acquired a GED during the follow-up period. When subjects were asked whether Steps had helped them improve their basic skills, for math 37 percent responded with improved "a lot," 41 percent responded with improved "somewhat," and 21 percent said "no change." Scores for reading were 37 percent for "a lot," 42 percent for "somewhat," and 26 percent for "no change." For writing, scores were 29 percent for "a lot," 48 percent for "somewhat," and 27 percent for "no change."

Employment

Forty-nine percent of the respondents reported that Steps had helped them to acquire employment. Thirty-eight percent reported that they were employed, and of these, 57 percent were working full-time and 43 percent were working part-time. Presumably, none of the subjects were employed before they entered Steps and enrolled in ABE/GED.

Further Education

Sixty-seven percent indicated that they had participated in additional training or education after leaving Steps.

Termination of Welfare

A discriminate analysis showed that math learning gain and whether the subject were currently employed were associated with whether a subject was still on welfare after terminating Steps.

Strengths

- The report is very well documented with respect to methodology and study limitations. The discussion of literature on welfare and adult literacy is excellent.

Limitations

- Data were collected on only 163 of the 920 intended subjects. Study participants differed from non-participants with respect to gender and minority status. All subjects volunteered for the study. This raises the possibility of selection bias.
- Only 53 subjects fell within the accurate range of the reading tests and 127 fell within the accurate range of the math tests.
- Data were collected on subjects between one and three years after termination from Steps. Thus without a control group it is difficult to know whether unknown factors, such as continued education, intervened in the gains noted.
- The perception of program success data is based on self-report.

WORKPLACE LITERACY

In recent years there has been a considerable increase in the attention paid to workplace adult literacy education. Much of this attention can be traced to a shift in public policy toward the human capital aspects of adult literacy education—the idea that a primary function of adult literacy is to enhance individual and societal economic productivity. Workplace literacy is defined as adult literacy education conducted in learners' places of employment to enhance individual job performance.

In the early 1990s the number of workplace literacy programs has increased considerably by a provision of the Adult Education Act that mandated a set-aside for federally-funded workplace education programs. As the number of workplace literacy programs increased, so did the number of impact evaluations available. However, as Kutner, Serman, and Webb (1991) and Mikulecky and Lloyd (1996) have noted, most of these evaluations are primarily descriptive and seriously flawed in their research designs and procedures. The workplace literacy evaluations presented here are the most credible of the many evaluations we initially reviewed.

17. Mikulecky and Lloyd

Mikulecky, L., & Lloyd, P. (1996). *Evaluation of workplace literacy programs: A profile of effective instructional practices*. Philadelphia, PA: University of Pennsylvania, National Center on Adult Literacy.

Rather than being a report of a single workplace evaluation, this sought to develop and refine a workplace literacy evaluation model. The report is reviewed here because, in the process of developing the model, several workplace literacy programs were assessed. The focus of the assessment was changes in learners' beliefs about personal effectiveness with literacy, changes in learners' literacy practices, learners' literacy improvement with general and workplace materials, and changes in learners' goals. Data were collected by interview from 10 groups of learners ($n=181$) in six different companies in a pretest, posttest format. Pretest data were collected at the beginning of each course and posttest data were collected toward the end of each course. Learners were asked about themselves, their literacy abilities, their reading and writing practices, how they read print, and their future educational plans.

Data on learners' literacy practices, beliefs, and plans were gathered through open-ended questions. For example, the question posed for learners' literacy practices was, "Tell me the sorts of things you read on the job during a normal week." Reading process was assessed with job-specific scenarios in which learners were presented with job-related materials and asked to explain how they read them.

Quantifiable interview responses were analyzed statistically. For the open-ended questions, categories were allowed to emerge from the data, and, when categories were sufficiently refined to yield an inter-rater reliability of 90 percent or more, they were analyzed quantitatively. For other open-ended questions, a holistic comparison was made between pre- and post-scores, and the change was rated either positive, neutral, or negative.

The evaluation itself was based on measuring the relationship between various course characteristics and on the data collected from learners. Course characteristics measured included total instructional time, topic (workplace and home/family orientation), and talk (discussion of literacy beliefs, plans, and reading writing processes). Data for course characteristics were collected from course syllabi, other descriptive information, and observation. Ratings of course characteristics were made by two researchers who compared results, discussed differences in ratings, and came to consensus. All classes enrolled between 10 and 15 learners and most were 20 to 50 hours in duration. Data were analyzed using analysis of variance (ANOVA).

The six workplace literacy program sites were:

1. A large manufacturing plant that offered GED preparation for 4 hours per week over 6 weeks and ESOL for 8 hours per week for 6 weeks.
2. A women's prison at which staff participated in report writing and instruction that enhanced skills needed for promotion. For report writing, 28 staff participated for three hours per week for 13 weeks; for promotion support, 9 participated for 3 hours per week for 7 weeks.
3. A small insurance company. Twenty learners participated for about 40 hours to improve job-related reading and writing.
4. A hospital in which 19 learners attended a computer-based writing course for 20 hours.
5. A large gasket maker. Ten learners studied job-oriented reading and writing for 50 hours.
6. A company that manufactured electric motors where 33 learners studied reading for 30 hours.

Findings

Findings are aggregated for all six programs studied. Learners with 200 hours of instruction (all subjects were enrolled in one program) made significant gains in practice away from work. Learners with 50 or less hours of instruction did not make significant gains.

With respect to the self-reported sophistication in the reading processes they would employ, those who spent over 70 percent of their time in class in reading and writing had a mean gain of over three times that of learners who spent less than 70 percent of their time on reading and writing.

All learners gained in job-related scenario comprehension, but learners who used job-related examples in class 20 percent or more of the time gained almost twice that of learners who did not. Learners

who had discussions of literacy beliefs and plans as deliberate components of their classes scored three times higher on reading scenarios than did learners who did not.

Learners who had deliberate discussions of literacy beliefs as part of their classes made significant gains in literacy self-efficacy, while learners where such discussions were merely incidental did not.

Learners gained in the area of specific, detailed future plans if they spent over 50 percent of their time in class reading and writing, if they used workplace examples in class at least 20 to 30 percent of the time, if they had deliberate discussions about literacy beliefs and plans, and if they had deliberate discussions about reading processes. Other learners did not make such gains.

Strengths

- The conceptual scheme and variables studied seem particularly relevant for workplace literacy.

Limitations

- The ANOVA is based on the aggregated data from six workplace literacy programs. The small number of programs intersects the possibility that unreported differences among programs may have influenced the findings in unknown ways. For example, findings on the relationship between practices away from work and hours of instruction divide hours of instruction into two categories: 50 and below hours and 200 hours. However, only one program, a large manufacturing plant, offered 200 hours of instruction. Thus it is not known if the ANOVA actually measured hours of instruction or other characteristics associated with this one program.
- The definition and meaning of variables is not always clear.
- The study lacks a control or comparison group.
- To a great extent the study relies on learner self-report, although such data were collected quite systematically.

18. Manufacturing and Financial Services in the Chicago Area

Mrowicki L., Jones, D., Locsin, T., Olivi, L., & Poindexter, C. (1995). *Workplace literacy in a total quality management environment for the manufacturing and financial services industry. Final performance report*. Des Plaines, IL: The Center-Resources for Education.

The goal of this workplace literacy project was to improve the productivity and efficiency of 19 manufacturing companies and two financial institutions by providing education to workers lacking basic skills. Customized curricula and materials were developed. Of the 21 companies that participated, 10 conducted ESOL classes, six had math classes, nine offered reading and writing, one offered only writing, and one conducted a communications class. Overall, 161 classes were conducted. The average number of classes per company was eight; 1,526 learners were served.

Most of the report provides descriptive information on the various companies' activities. The evaluation of the project employed a modification of the Context-Input-Process-Product (CIPP) model of program evaluation. Context was defined as the shared goals and philosophy of key personnel and participants. Input included resources such as personnel, materials, time, and facilities. Process was defined as the extent to which instruction was congruent with project goals. Product was defined as indicators of project effectiveness. Because of limited resources, site visits were restricted to five companies that were considered to be representative of project operations. These five were visited four times and three other sites were visited once. The five companies were: Avon Products, R. Olson Manufacturing, Allied Die Casting Company, E.J. Brach, and First Chicago National Bank.

Data for the evaluation were collected through post-program participant surveys; structured interviews with participants, instructors, and program personnel; observations of instruction; reports of instructor training; participant records; and pre- and posttest scores

Findings

From post-program surveys and on-site interviews it was determined that most learners were satisfied with course content, and the confidence to use learned skills was built among participants. One hundred percent of those surveyed said that they would recommend participation in the program to others.

Pre- and posttests were administered to a limited sample (the selection of the sample, its size, and the test are not described). The mean pretest score was 54 percent and the mean posttest score was 80 percent for a gain of 26 percent. Instructor and company representatives' comments about the program were positive.

Strengths

- The study presents a useful descriptive picture of the program.

Limitations

- The description of methods and procedures is very sparse, and important information such as sample selection procedures, sample size, and the nature of the test used is not explained. As a result, it is difficult to assess the validity of the study.
- Most of the data is anecdotal and based on self-report.
- There are no control or comparison groups.

19. Workplace Literacy across the Three Phases of Textile Manufacturing

Fisk, W. (1994). *Workplace literacy across the three phases of textile manufacturing. National workplace literacy demonstration project. Performance report.* Clemson, SC: Clemson University.

The goals of this workplace literacy project were to promote and improve employees' literacy skills in order to increase worker productivity in a textile manufacturing plant. Task analysis preceded the development of a job-relevant curriculum that focused on math, reading, and communication skills. Instruction was integrated into the three phases of the manufacturing cycle—spinning/ weaving, finishing, and fabrication—and was conducted in two plants. Two hundred and forty participated in classes during the grant period. Ninety-nine of these participated in GED preparation classes.

Although the evaluation design had both formative and summative components, all the impact data are found in the summative section. Data were collected through seven site visits; interviews with project supervisors and participants; and questionnaires administered to teachers, participants, and plant supervisors. Instrumentation included teacher-made tests. In some cases pre- and posttests were administered, in others only posttests were given, and in some cases, no tests were used. Data on the validity and reliability of the tests is not given. A supervisor's questionnaire was developed by the evaluator to measure supervisors' perceptions of project implementation and impact on participants. Employee's and teacher's questionnaires were developed for the same purpose. In addition, a six-item course evaluation form was administered by project staff and the evaluator at course completion. No information is provided regarding sample sizes or survey return rates.

In addition a "long-term" follow-up was conducted 10 months after the termination of classes conducted during the first project grant period with a random sample of 24 employees. A 21-item survey was used for this purpose.

Findings

Plant one: In one department 13 students had completed math pre-and posttests and 17 had completed reading tests. These students demonstrated statistically significant increases in math and

vocabulary. In another department, math scores increased significantly. Some participants exhibited changes in communication behavior after attending class. Many students refused to take the tests.

Plant two: Because this plant changed from classroom instruction to computer assisted instruction, it was difficult to measure tested learning gains of the 113 learners who participated. Learners made significant gains in math, and there were positive changes in learners' attitudes after attending the communications class.

Although it was originally intended to measure the program's impact on productivity, this plan was abandoned because it was thought that the seven- to eight-week classes were of insufficient duration to create measurable impacts on productivity.

With respect to the study conducted 10 months after completion of the first year project (the evaluation reviewed here is of the second year of operation), it was found that, of the 186 fabrication plant project participants, 74 (40 percent) voluntarily enrolled in additional courses for an average of 22 hours of instruction per employee. Of the 24 randomly selected participants who were administered the follow-up survey, data indicated that the project had a positive impact on interactions with co-workers and supervisors on the job, interactions with their families, and satisfaction with their daily lives. Positive work habits were formed and employees accepted increased responsibility on the job, in the family, and in the community.

Strengths

- The report contains a useful descriptive account of workplace literacy activities.

Limitations

- Important information, such as survey return rates, is not supplied.
- Much of the data is self-reported.
- In many cases the sample size is quite low.
- There is no control or comparison group.
- The validity and reliability of the tests used is not reported.

20. Wisconsin Workplace Partnership Training Program

Paris, K. A. (1992). *Evaluation of the third year of the Wisconsin Workplace Partnership Training Program, March 1, 1991 through August 31, 1992*. Madison, WI: Center for Education and Work, University of Wisconsin, Madison.

The goal of the Wisconsin Workplace Partnership Training Program was to provide job-specific basic skills instruction to 3,066 employees at their work sites in order to promote job retention and/or job advancement and increased productivity. At the state level, the program was a cooperative effort between the Wisconsin Board of Vocational, Technical and Adult Education, the Wisconsin State AFL-CIO, and Wisconsin Manufacturers and Commerce. At the local level, there was a partnership between local Vocational Technical and Adult Education colleges, unions, and employers. The program operated in 23 work sites. Instruction focused on job-related reading, verbal and written communication, listening, math, reasoning and problem-solving, and use of English. Four research questions were addressed: 1. To what extent do program participants agree they have achieved academic and job-related objectives? 2. To what extent do local partners agree that participants achieve academic and job-related objectives through participation in the program? 3. Which program objectives do participants and local partners view as most significant? 4. What are some of the best practices exhibited by programs whose participants report the highest mean improvement in academic skills and job performance?

Two surveys were designed to collect data. The first was administered to participants to measure the extent to which they perceived their basic skills and job-related skills had improved. The second was administered to local partners to determine their perceptions regarding how much participants' skills had improved. Because a field test indicated that learners would have difficulty reading the instrument, the survey was verbally administered to participants by the researchers. The participant survey was administered at 10 randomly selected program sites between April and May 1992 to all learners who were present on days of the site visits. One hundred and two participants—3 percent of all enrollees in the program—were interviewed; the response rate was 97 percent.

For the local partners survey, during the site visits surveys with self-addressed envelopes were given to the union liaison to the program, the human resource director, the CEO or highest official at the site, three to six peer advisors identified by instructors, teachers, and the Vocational Technical Adult Education supervisor. Non-respondents were contacted by telephone to promote their participation and data were tabulated both by site and in the aggregate. One hundred and ninety-seven local partners surveys were administered and 160 were returned for a response rate of 81 percent.

Findings

Fifty percent of the learners had attended between one and six months; 24 percent had attended seven to twelve months.

Participants' Views

All respondents indicated the extent to which they had improved their basic skills on a 5-point Likert scale on which 5 indicated "strongly agree." Mean scale scores were as follows: math (4.5), writing (4.2), reading (4.4), speaking (4.0), ESOL (4.6), GED (4.4), and computer (4.6). Similarly, using the same scale, respondents were asked the extent to which they had improved job-related skills. Scores were: job skills in general (4.2), getting along better with employees (3.4), getting along better with superiors (3.9), problem-solving (4.1), quality (4.1), self-image (4.3), eligible for promotion (1.3), and job enjoyment (2.3). Of all the skills listed, both basic and job-related, only eligible for promotion and job enjoyment received mean scores below 3, the scale midpoint.

Ninety-one percent agreed or strongly agreed that they were satisfied with the progress they made in the program. However, only 3 percent of the respondents received promotions and 10 percent were transferred to another job. When months of attendance were analyzed in relation to satisfaction with an analysis of variance, no significant relationship was found.

Local Partners' Views

Local partners were asked about the extent to which they perceived participants' basic and job skills had improved. The same 5-point Likert scale was used to record responses. For basic skills the scores were: math (4.6), writing (4.3), reading (4.4), speaking (3.7), ESOL (3.5), GED (4.4), and computer (4.4). For job skills the scores were: job skills in general (3.8), getting along better with employees (3.5), getting along better with superiors (3.1), problem solving (3.8), quality (3.3), self-image (4.4), eligible for promotion (3.0), and job enjoyment (3.4). Eighty-five percent of local partners believed that participants were satisfied with their progress in the program.

An analysis of variance conducted on the type of local partner respondent showed some significant differences. With respect to mean improvement in total job-related improvement, instructors (mean=4.0) rated learner improvement higher than company officials (mean=3.3). With respect to perceptions that participants are promoted, instructors responded with a mean of 3.4 whereas company officials responded with a mean of 2.3.

Analysis of variance also showed that in general participants' ratings of their improvement in both basic skills and job-related skills were significantly higher than those of local partners. Local partners perceived that participation in the program led to promotion to a significantly greater extent than did

participants, although neither group perceived that the effects of participation on promotion were substantial. Local partners also responded with significantly higher scores for job enjoyment than did participants.

Both participants and local partners were asked which program objectives they felt were the most important. The three highest for participants were computer skills, math, and self-image; for local partners the three most important were self-image, math, and reading.

In general the results of the study suggest that while both learners and local partners were satisfied with the gains learners had achieved in basic and job-related skills, these gains had not been translated into job-promotion benefits for workers.

Strengths

- Explanation of methods and procedures and the presentation of findings are clear and complete.
- The analysis of variance component of the study is a useful elaboration of the basic findings.
- Although the sample size is relatively small, it is adequate and the response rate is high.

Limitations

- Although the 10 sites from which data were collected were randomly selected, the subjects for the study were not.
- All data are based on participant and local partner self-report.
- There are no control or comparison groups.

21. Central Labor Council and the Consortium for Worker Education

Gross, A. L., & Feldman, S. (1990). *Evaluation report: The workplace education program of the Central Labor Council and Consortium for Worker Education*. New York: Center for Advanced Study in Education: The Graduate School and University Center, City University of New York.

This study evaluated the workplace education program of the New York City-based Central Labor Council for Worker Education. Data were obtained from 15 union-based workplace literacy programs for the time period October 1, 1989, to June 30, 1990. During that time, the program enrolled 3,775 learners and conducted 215 different classes. ESOL, basic education, and union-specific skills courses were offered. Basic education classes ranged from basic reading and writing to GED preparation. Union-specific classes focused on job-related literacy skills. Classes ranged from 2 to 49 weeks in duration and met from 4 to 40 hours of instruction per week. Mean contact hours were: ESOL, 62.3; basic education, 70.3; and union-specific, 33.1. Variables measured included learners' attainment of their personal goals and tested learning gain.

Data were learner records compiled by the New York Literacy Assistance Center which contained pre- and posttest scores, information on contact hours, information on attainment of learners' objectives, and background data. The researchers encountered several problems with the data set including: 1. contact hours were not recorded for nearly one third (1,357) of the 3,773 cases; 2. there were substantial data entry errors; and 3. in many cases there were multiple records for students. The implications of these problems are explained by the study's authors in their discussion of findings on learners' personal goal attainment:

A second problem concerns the manner in which an objective was designated as not being achieved. For the $n = 3,259$ students for whom class type was established, only $n = 995$ achieved their stated objective. Of the $n = 2,264$ "non-achievers", $n = 794$ were simply missing from the Impact File, $n = 927$ had a "blank" recorded, and $n = 543$ achieved objectives which did not match the originally stated objective. It is not clear whether all 2,264 cases represent failure to achieve an objective or the failure to record the data. (Gross & Feldman, 1990, p. 23)

The tests used were the Test of Adult Basic Education (TABE) for adult basic education, the JOHN and Davis tests for ESL, and various job-related tests for union-specific education (administered in some but not all cases). Although the validity and reliability of the TABE has been established, its appropriateness for job-related basic skills education is an issue. Information on when and how these tests were administered is not provided. For example, the time intervals between pre- and posttests are not reported.

Findings

Learners' Attainment of Personal Goals

When learners were asked to specify their learning goals, they were presented with 11 response options. For all types of classes combined, the most commonly stated goal (46 percent) was to prepare for a better job. Twenty-four percent of those who stated this goal reported that they had achieved it. The second most commonly stated objective was to learn English (19 percent), and 18 percent of those who stated this objective claimed they had achieved it.

For ESOL classes, the most commonly stated objective was to learn English (60 percent), and 17 percent of those who reported this objective said they had achieved it. To acquire citizenship was the second most important objective for ESOL students (31 percent), and 7 percent reported they had achieved it.

For basic education GED preparation was the most common objective (38 percent), and 13 percent claimed it had been achieved for them. To prepare for a better job (30 percent) was the second most commonly stated objective (30 percent), and 64 percent stated that this objective had been achieved.

For union-specific course learners, to prepare for a better job was the overwhelmingly most common objective (71 percent), and 22 percent reported they had obtained it.

Overall, for ESOL 152 learners achieved their goals and 841 did not. For basic education 339 learners achieved their goals and 402 did not. For union-specific courses 504 learners achieved their goals and 1,021 did not. The overall achievement of objective rate was 31 percent. The proportion of learners attaining the objectives that were presented to them varied greatly by program, with the highest scoring program achieving a 95 percent goal attainment rate and the lowest scoring achieving a 0 percent goal attainment rate. Although on the face of it findings on learner goal achievement seem to suggest rather low goal attainment, the data analysis problems reported earlier may well have resulted in an underestimation of learner goal attainment. It may also have been that the instructional goals of some of the union programs did not match well the goals contained on the response list presented to learners.

Tested Learning Gain

Pre- and posttests were only administered in 9 of the 15 programs. Of the three programs that offered ESOL, two showed statistically significant gains, one did not. Of the programs that offered basic education, three showed gains, one showed a decline and one showed no gain. Of the programs that conducted union-specific education, three showed gains and one did not.

Strengths

- Methods and procedures are clearly explained and the report is honest about its limitations.
- The sample size is adequate.

Limitations

- Problems with data collection and data entry created inaccuracies and missing data in the data set. For this reason, findings on student goal attainment are suspect.

- Learners were presented with a choice of 11 goals as response options for learner goal attainment. How relevant these response options were to the instruction offered is not clear.
- The relevance of the TABE, JOHN, and Davis tests for workplace-oriented instruction is an issue.
- Learners' goals data are based on self-report.
- The study lacks a control or comparison group.

FAMILY LITERACY

In family literacy, low-literate adults and their pre-school children are educated together to create a presumably synergistic effect on educational gains for both. Family literacy takes many forms. Even Start, the federally-funded family literacy program, requires programs to offer parent education, adult literacy education, and early childhood education. Support services such as transportation, home visits, and child care are generally offered as well.

The Kenan Family Literacy Model program is another well-known and widely implemented model for family literacy. Programs that follow this model must provide adult literacy education and early childhood education plus two additional components, Parent Time and Parent Time Together (National Center for Family Literacy, 1992). In Parent Time, parents and their children design and participate in programs of interest to both, such as child nurturing and managing and coping with child behavior. In Parent Time Together, parents and their children play together to promote interaction and improved communication. Provision of pre-employment training and service integration are encouraged, but not required. Family education programs vary in the extent to which adult literacy education is specifically tailored to meet parenting objectives and the degree to which the components of the program are integrated. They also vary with respect to the provision of support services, such as tutoring, childcare, and home visits.

22. The National Even Start Program

St. Pierre, R., Swartz, J., Murray, S., Deck, D., & Nickel, P. (1993). *National evaluation of the Even Start Family Literacy Program: Report on effectiveness*. Washington, DC: U.S. Department of Education, Office of Policy and Planning.

St. Pierre, R., Swartz, J., Gamse, B., Murray, S., Deck, D., & Nickel, P. (1995). *National evaluation of the Even Start Family Literacy Program: Final Report*. Washington, DC: U.S. Department of Education, Office of Policy and Planning.

Even Start is a federally-funded family literacy program that commenced in 1988. In 1991-1992, 234 projects were in operation throughout the 50 states, and 9,690 families received some core services, 13,541 children received early childhood education, and 10,800 parents received adult literacy education or parent education. To be eligible, the law requires that a family have an adult who is eligible for adult literacy education under the provisions of the Adult Education Act, who is a parent with a child under eight years old, and who lives in a Chapter 1 school area. Even Start programs must have at least three components: adult basic education, early childhood education, and parenting education. They are also required to provide home-based services and are encouraged to offer supportive services such as transportation to programs and childcare.

The national evaluation of Even Start was mandated by the authorizing legislation and was contracted to Abt Associates with a subcontract to the RMC Research Corporation. The evaluation began in 1990 and ended in 1993. Four research questions were posed: 1. What are the characteristics of Even Start participants? Who is in the program? 2. How are Even Start projects implemented and what services do they provide? 3. What Even Start services are received by participating families? 4. What are the effects of Even Start on participants? The impact component of the study addressed by question four had two major components. The first, the National Evaluation Information System (NEIS), was a data set that contained descriptive information collected from Even Start local programs. The second was an in-depth

study of 10 programs. The Even Start national evaluation measured effects on children, parents, and families. This report focuses exclusively on the assessment of the adult literacy education component.

The NEIS

The NEIS collected descriptive longitudinal and cross-sectional data on four cohorts of participants, projects first funded in 1989, 1990, 1991, and 1992. For the first two cohorts, data were collected annually through completion of data collection forms, parent interviews, testing of children and adults, and completion of program logs. Data were collected from families at entry to Even Start, at the end of each program year, and at exit from the program. Families who did not remain in the program long enough to be posttested were eliminated from the analyses. Data were collected by local program staff who were trained by the researchers and who were supported by a portion of the local program grant earmarked for this purpose. The CASAS test was administered to participants of adult literacy education. For the second two cohorts, a computerized data collection system was used that allowed local programs to submit their data on diskette for direct computer input. However, because many local program staff were not familiar with computers, many failed to use the system or used it incompletely.

In-depth Study

The in-depth study was designed to complement the NEIS. It was implemented with 10 projects selected from the first cohort of 73 projects on the basis of geographic location, level of program implementation, and willingness to participate in the study. As intended, 40 families for each project were to be randomly assigned to either a participant or control group (20 in each group). However, only five of the ten programs were able to implement the random design. Study participants were pretested with tests that included the CASAS test in the fall of 1991 and were posttested in summer of 1992 (9 months) and spring of 1993 (18 months). Tests were administered by a private contractor. Comparison of the control group ($n=98$) with the participant group ($n=101$) showed no significant differences with respect to subjects' background characteristics. However, comparison of the in-depth study participants with the NEIS sample showed differences, especially with respect to Spanish as the primary language and Hispanic ethnicity.

Effect estimates were based on a regression model in which posttest was used as the dependent variable and pretest and group assignment were used as the independent variables. Effect magnitudes were calculated by subtracting pre-post test gains of the control group from those of the participant group and by then dividing by the standard deviation of the control group pretest scores. Adult literacy education outcome measures used included scores on the CASAS test, GED attainment, and reading/writing activities at home.

The National Evaluation of the Even Start also measured the effects of the program on children using the PreSchool Inventory (PSI) and the Peabody Picture Vocabulary Test (PPVT) as outcome measures. The 32-item PSI measures school readiness skills such as identifying shapes and colors and numerical skills. The PPVT measures receptive vocabulary. In general, the results of the effects on children are complex. At issue for this study is the extent to which parents' participation in the adult literacy education component of Even Start affected their children's gains on the PSI and PPVT. This question was addressed by regressing children's posttest scores on parents' background and program participation variables.

Findings

Tested Learning Gain

For tested learning gain the CASAS Reading Survey was used. This test has four levels with 24 to 40 items per level. The CASAS has been used extensively as a learning gain measure in adult literacy. The most recent national evaluation of the federal adult literacy program and the California GAIN Study were two studies reviewed in this report that have used the CASAS.

In the in-depth study, valid pre- and posttests were obtained from 64 adult literacy education participants and 53 control group members. At the time of the second posttest (18 months), a statistically non-significant difference of 3.7 points on the CASAS was found between the gains of the participant and control groups, indicating that participation in Even Start adult literacy education had no significant effect on tested learning gain.

However, both the participant and control groups gained significantly on the test. This findings was perplexing since the control group presumably received no instruction. Further analysis showed that nearly one quarter of the control group members reported they had participated in adult literacy education at the time of both the pre- and posttests. This and testing error may well have reduced the differences in gain exhibited by the two groups and accounted for the gains among the control group. With respect to the NEIS data, gains were small (4.6 points) but statistically significant after 70 hours or more of instruction and of about one third of a standard deviation in magnitude.

It should be remembered that the in-depth study included all participants who were pre- and posttested regardless of whether they remained in the adult literacy education program, whereas the NEIS sample included only those participants who were still active in the program and had received at least 70 hours of instruction. Both the in-depth study and the NEIS data show a significant relationship between the amount of adult literacy education instruction and gains on the CASAS. NEIS data show mean gains of 3.0 points between 1 and 69 hours of instruction, gains of 4.3 points between 70 and 200 hours of instruction, and gains of 5.2 points for over 200 hours of instruction. NEIS data also show gains were highest for those who entered with the lowest CASAS scores.

GED Attainment

Data from the in-depth study show that 22.4 percent of the participant group and 5.7 percent of the control group members attained a GED over the 18-month course of the study. For the NEIS sample, of those who had at least a ninth-grade education at the time of initial data collection, 14.1 percent had obtained a GED within a year or less. Attainment of a GED was related to grade level at intake, CASAS score, hours of instruction, younger age, and having English as the primary language.

Reading and Writing in the Home

Reading and writing in the home was measured by 13 self-report questions that asked about the frequency parents read various types of reading material and a 11-item self-report measure focusing on writing activity. Frequency was measured on a scale that ranged from 1 to 3. The instrument was administered only to in-depth study participants. At pretest, all study participants scored at about the middle of the measure (2 on the 3-point scale). No significant program effects for either reading or writing in the home were found after 18 months.

Effects on Children's Literacy-Related Skills

Although the amount of time parents participated in parenting education was positively related to their children's PPVT (which measures receptive vocabulary) scores, the amount of time parents participated in adult literacy education was not significantly related. Likewise, there was no relationship between children's scores on the PSI (which measures school readiness) and parents' participation in adult literacy education. Remarking on these findings the study's authors noted:

Parenting education activities are often targeted at developing parents' abilities as teachers of their children, and children's language development is exactly what is taught through many of the parenting activities emphasized in Even Start, such as reading to children. On the other hand, there is no special reason that participation in adult education programs, which focus only on the

development of adult-level skills, should yield immediate benefits to children. (St. Pierre et al., 1995, p.180)

Strengths

- The reports provide an adequate description of methods and procedures and the findings are clearly presented.
- The in-depth study includes random assignment to control and participant groups.
- The NEIS data were systematically collected and are comprehensive in the variables studied. The sample size for the NEIS is large and adequate.
- The study is longitudinal, measuring program effects at 9- and 18-month intervals.

Limitations

- The sample size for the in-depth study was small.
- Random assignment to treatment and control groups was implemented in only five programs for the in-depth study. Thus differences in program characteristics and contexts could have influenced the findings.
- As the report acknowledges, the CASAS test may not have been sufficiently sensitive to measure literacy gain.

23. The National Center for Family Literacy

Hayes, A., E. (1997). *The power of Even Start family literacy: A summary of findings from a follow-up study*. Louisville, KY: The National Center for Family Literacy.

The purpose of this study was to determine the long-term effects for a group of Even Start programs that met the National Center for Family Literacy's quality standards. The rationale for the study was that, because programs evaluated in the National Even Start Evaluation did not necessarily meet the high quality standards established by the National Center for Family Literacy, the effect of high quality standards was "obscured." Program sites were selected through nomination by "leaders in the Even Start community" and by external evaluators familiar with Even Start programs. Data were collected from 15 programs that served a total of 507 children and 508 adults. Data were supplemented by the evaluation reports of several other family literacy programs.

Data were collected in January and February 1997, one to six years after families had terminated the program, by local site coordinators who used forms and procedures developed by the National Center for Family Literacy. Most of the study focused on effects on children; only the impacts of adult literacy education are reported here.

Findings

Sixty-two percent of those for whom attainment of high school certification was an appropriate goal attained high school certification. Fifty percent either obtained a job or obtained a better job. Forty percent enrolled in higher education. Forty-two percent of the 260 former participants who received welfare when they enrolled in the program reduced the amount of public assistance they received.

Strengths

- The sample size is adequate.

Limitations

- The report lacks important information regarding research design and procedures. Most notably, information on sample selection is omitted.

- The programs from which research subjects were drawn were selected because they reportedly met high standards. Thus the sample is not representative.
- The study lacks a comparison or control group.
- Presumably, most findings are based on self-report.

CHAPTER 4: CONCLUSIONS

In accordance with the goals of this study, this chapter addresses four questions: 1. Based on the case analyses of the outcome and impact studies reviewed, how effective is the adult literacy education program in the United States? 2. What are the common conceptual, design, and methodological problems in the studies analyzed? 3. What are the implications of the analysis for adult literacy education policy? 4. What recommendations are warranted?

HOW EFFECTIVE IS THE ADULT LITERACY EDUCATION PROGRAM?

The 23 studies reviewed in-depth for this report were chosen because, with respect to their design, methodologies, and reporting, they were the most credible of the approximately 115 studies initially reviewed. However, as those who have read the case studies can attest, these 23 studies vary considerably in their strengths and limitations and none is definitive to the extent that it alone can put the question of impact to rest.

These studies represent evidence rather than proof of impact, and, like evidence in a trial, their findings must be weighed in order to reach reasonable conclusions. Weighing has two dimensions. The first is the extent to which the various studies converge or diverge with respect to their findings on specific outcome and impact variables. Consensus among studies points towards effectiveness/ineffectiveness while lack of consensus suggests an inconclusive resolution. The second dimension is the credibility of the individual studies. Obviously, when arriving at conclusions, more credible studies must be weighed more heavily than less credible studies.

The conclusions set forth here are deemed to be reasonable inferences from the findings reported in the case studies. They do not represent proof. Indeed it is unlikely that any conceivable study or studies could arrive at certainty. Table 1 presents the data used for this analysis.

Table 1: Results of the Case Studies

STUDY	Emp l	bjob	inc	ced	wel	read	write	math	TLG	GED	slf-c	chld	pgoal	METH
1N	Y			y		y	y	y	y		y			PP-L
2N	?			y		y	y	y			y		y	P
3N	Y		y	y	y				y			y		PP-L
4N	Y				n				n			y		PP-L-C
5S	N		y		n									PP-L-C
6S	Y										y	y	y	PP
7S						y	y	y						PP
8S	Y	y	y	y	y	y	y	y			y	y	y	P
9S	Y		y	y	y	?	?	?			?	y	y	PP-C
10S	Y	y		y		y	y	y			y	y		P
11S											y			P
12S	Y	y		y		y		y		?	y	?	y	PP-C
13S	?		?	y		y	y	y	y	?	?	?	y	P-L
14W									n	y				PP-C
15W	Y		y		y									PP-C
16W	Y			y	y	y	y	y	y	y				PP
17WO														PP
18WO														P
19WO														P
20WO		n				y	y	y			y			P
21WO									y				n	PP
22FL									?	y		n		PP-L-C
23FL	Y	y		y	y					y				P
Total	y=11 ?=2 n=1	y=4 ?=0 n=1	y=5 ?=1 n=0	y=10 ?=0 n=0	y=6 ?=0 n=2	y=9 ?=1 n=0	y=8 ?=1 n=0	y=9 ?=1 n=0	y=4 ?=1 n=2	y=4 ?=2 n=0	y=8 ?=2 n=0	y=6 ?=2 n=0	y=6 ?=0 N=1	

Along the top of the table are the outcome and impact variables addressed by a sufficiently large number of studies to permit inferences:

empl = gains in employment
 bjob = acquisition of a better job
 inc = increased income
 ced = continued education
 wel = termination of or reduction in
 public assistance
 read = self-reported gains in reading
 write = self-reported gains in writing

math = self-reported gains in math
 TLG = tested learning gain
 GED = GED acquisition
 slf-c = self-confidence, self-esteem, or self-
 confidence
 chld = impact on children's education
 pgoal = attainment of learner's personal
 goals

Studies are listed in the order in which they were reviewed in Chapter 3.

Study codes (STUDY): N=national study, S=state, W=welfare, WO=workplace, FL=family literacy.

Methods (METH): P = post measurement only, PP = pre-post, L = longitudinal measurement, C = control or comparison group.

y = study found impact on this variable, ? = inconclusive findings, n = study found no or negative impact, blank cells = variable not studied.

In interpreting Table 1 and the conclusions made from it, three caveats are in order. First, the variables included are those studied by a sufficiently large number of studies to enable reasonable conclusions. However, variable definitions and their units of measure vary among studies. In some studies, for example, learning gain is measured by the CASAS, while in others the TALS or TABE are used. Second, if a given study reported a gain, the gain is listed as positive (y) in the table irrespective of the size of the gain or the quality of the study's methodology. In some cases the gains reported as positive are quite small, and in some cases the limitations of the study render claims of gains suspect. Third, the totals are aggregates of studies conducted at different times and on different populations of adult literacy

learners, welfare clients and employees being examples. Drawing conclusions from such aggregates presumes that doing so is both valid and meaningful.

Employment

Eleven studies found gains in employment, two were inconclusive, and one study found no gain. Most of the studies reporting employment gains lacked comparison or control groups and most were based on self-report data. Lack of a control group is a serious limitation, since employment is highly susceptible to fluctuations in the economy, and, without a control group, it is difficult to infer whether gains were caused by participation in adult literacy or by economic factors. Furthermore, in some cases findings may have been influenced by the bias inherent in self-report.

Two studies, the Washington Workforce Training Study (5S) and the Texas JOBS Program Evaluation (15W), did use matched comparison groups, however. Both used hard economic data on employment rather than self-report and both were of high technical quality. For former adult literacy education participants who enrolled in order to gain employment, the Washington study found negative gains in employment in contrast with the comparison group. In the Texas JOBS Program Evaluation, for which the subjects were JOBS participants and AFDC recipients, employment benefits for those who participated in adult literacy education were found.

Conclusion: In general, it is likely that participants in adult literacy receive gains in employment.

Better Jobs for Those Who Are Employed

Of the five studies that measured job improvement, four reported that participation in adult literacy education helped participants gain better jobs and one study reported no gains in job improvement. Although one of these studies conducted in Ohio (12S) included a comparison group, in all cases the data were based on self-report. In a national study (3N), substantial increases in pay were found for those who were employed at the time of initial data collection, although more than half the respondents believed participation had been of no help in the acquisition of their increased remuneration. The one study that found no job improvement (20WO) reported that neither participants in workplace education nor local partners (employers, teachers, union representatives) believed that participation had resulted in promotion.

Conclusion: In general, participants in adult literacy education believe their jobs improve over time. However, there is insufficient evidence to conclude that participation in adult literacy education causes job improvement.

Increased Earnings

Of the six studies that measured impact on earnings, five found gains and one was inconclusive primarily because of its methodological limitations. Two of the studies reporting gains—the Washington Workforce Study (5S) and the Texas JOBS evaluation (15W)—included matched comparison groups, used hard economic data to measure earnings, and were technically sophisticated. However, the Washington study included only subjects who participated in adult literacy to gain employment, and the gains in earnings found were very small. The Texas study was restricted to JOBS participants.

Conclusion: In general, it is likely that participation in adult literacy education results in earnings gain.

Continued Education

All of the 10 studies that measured continued education found that participation in adult literacy had a positive impact on further education, although one study, the 1993 National Evaluation (3N), was somewhat ambiguous. It found that, while during initial data collection 60 percent of the respondents reported they might attend college and 70 percent said they might attend vocational school, 18 months later the percentage for college attendance had declined to 37 and the percentage for vocational school had declined to 65. Five of the nine studies asked learners whether they planned to enroll in future education while three, which used a pre-post design, asked respondents whether they actually had enrolled. None of the studies included a control or comparison group.

Conclusion: In general, adult literacy education has a positive influence on participants' continued education.

Termination/Reduction of Welfare Dependence

Six studies show a reduction of dependence on welfare for adult literacy education participants, while two do not. Four studies included control or comparison groups. Of these, two showed no change in welfare status and three reported welfare reduction. Of the four studies with comparison groups, two were of high technical quality. Of these two, one, the Washington Workforce Study (5S), reported no reduction. The other, the Texas JOBS Evaluation (15W), did find welfare reduction over time. For one of the studies, the 1973 National Evaluation (3N), the reduction in welfare was very small.

Conclusion: Although the evidence suggests that participants in welfare-sponsored (JOBS Program) adult literacy education do experience a reduction in welfare dependence, the evidence is inconclusive as to whether adult literacy education in general reduces welfare dependence for participants.

Self-Reported Gains in Basic Skills

The overwhelming majority of studies that asked participants whether they had improved in reading (9 studies), writing (8 studies), and math (9 studies) found that learners perceived gains in these subjects. The exception is one study that was rated inconclusive primarily for methodological reasons.

Conclusion: Learners perceive that participation in adult literacy education improves their skills in reading, writing, and mathematics.

Tested Basic Skills Learning Gain

Using tests to measure basic skills gain presents many thorny issues including the appropriateness of tests used, test validity and reliability, attrition of subjects between pre- and posttesting, ceiling and floor effects, and the amount of instruction received between test administrations. These issues are discussed in more detail in the case studies and later in this chapter.

Another issue that plagues inferences on tested learning gain is the lack of a standard in adult literacy education regarding what should be considered an acceptable gain over a given duration of instruction. Without such standards, it is difficult to infer whether reported gains represent poor, good, or outstanding performance.

Five studies found tested basic skills gains, two did not, and one study was ambiguous in its findings. For ABE learners, the National Evaluation of Adult Education Programs (NEAEP) (1N) found small gains on the TABE test after an average of 84 hours of instruction and small gains on the CASAS for ESOL students after 120 hours of instruction. However, as the case analysis notes, there were substantial methodological problems associated with these results. The NEAEP did not include a control or comparison group. The 1973 national evaluation (3N) also showed small gains on the TABE, but almost one fifth of those tested had had 39 or fewer hours of instruction. Like the NEAEP, the 1993 national evaluation did not include a control or comparison group.

Three studies did include control or comparison groups, and two of these, the California GAIN Study (14W) and the National Evaluation of the Even Start Program (22FL), were technically sophisticated. Using the TALS test, the California GAIN Study found no gain on the document and quantitative portions of the test after an average of 251 hours of instruction, although substantial gains were noted for one of the six counties studied in the evaluation. All subjects were welfare recipients and mandated participants in the California GAIN program. For a large sample that received pre- and posttesting on the CASAS test, the National Even Start Evaluation found small gains after about 70 hours of instruction. However, for a smaller sample that included a control group, no significant gains were found.

Conclusion: As measured by tests, the evidence is insufficient to determine whether or not participants in adult literacy education gain in basic skills.

GED Acquisition

Four studies found impacts for GED acquisition and no study found lack of impact. Two studies were rated inconclusive because the manner in which they measured GED acquisition made it difficult to infer the nature of impact. One inconclusive study, an Ohio study (12S), reported that of the 62 percent of the respondents who entered ABE with the goal of obtaining a GED, 40 percent said they attained that goal. The other inconclusive study, a Wisconsin Study (13S), found that while 44 percent of the former participants indicated they had desired to obtain a GED, 27 percent had actually passed the GED tests. The California GAIN Study (14W) and the National Even Start Evaluation (22FL), both of which included comparison groups, found significant impact on GED acquisition.

Conclusion: In general, adult literacy education provides gains in GED acquisition for participants entering at the adult secondary education (ASE) level.

Self-image

In the studies listed, self-image was variously defined as self-concept, self-esteem, and self-image. Of the 10 studies that included this variable, eight showed gains which were usually quite large, two were rated inconclusive primarily because of the way they reported their findings and no study found declines in self-image. With one exception, measures of self-image were based on self-report. That exception was the Tennessee Longitudinal Study (6S), which found small, but statistically significant, gains on the Rosenberg Self-Esteem Scale between initial data collection and follow-up. Yet despite these small measured gains, when the Tennessee subjects were asked whether their feelings about themselves had changed, 70 percent responded affirmatively.

Conclusion: Participation in adult literacy education has a positive impact on learners' self-image.

Impact on Children's Education

The studies reviewed employed several variables to measure impact on children's education, including the extent to which participants helped their children with homework, attendance at PTA meetings, and parent-teacher interaction. Some studies used one variable; others used multiple variables. The assessment of impact, therefore, represents a summation of the findings for each study listed.

Eight studies found that participation in adult literacy had positive effects on children's education, two were inconclusive, and one was negative. Of the two that were inconclusive, one was so primarily for methodological reasons and the other, an Ohio (12S) study that included a comparison group, demonstrated mixed results. Although former participants scored higher than the comparison group on attending parents' meetings, the comparison group scored higher on communication with the schools. Differences between the groups on helping children with homework were not statistically significant. All studies were based on respondent self-report except the one study—the National Even Start Evaluation—that showed negative results.

The National Even Start Evaluation found no significant effects with respect to reading and writing in the home after 18 months for families where parents participated in adult literacy education. Neither did the study find a relationship between children's scores on the Peabody Picture Vocabulary Test (PPVT) and parent's participation in adult literacy education.

Conclusion: According to learners' self-reports, participation in adult literacy education has a positive impact on parents' involvement in their children's education.

Attainment of Personal Goals

Measuring attainment of learners' personal goals raises two interesting and important conceptual issues. First, if participation in adult literacy education should result in new self-understanding and increased aspirations that lead learners to change their goals during the course of instruction, measuring the extent to which initial goals were achieved in a post assessment format is irrelevant.

Second, it is very difficult to measure personal goal attainment with an experimental or matched comparison group design because non-participants who constitute control or matched comparison groups do not have goals for adult literacy education. Consequently, every study that measured personal goal attainment either collected their data from program participants or compared the goal attainment of

participants to former learners who exited the program. Former students included dropouts as well as successful completers.

Yet participation in adult education is generally voluntary. Therefore, positive reports of goal attainment may be artificially inflated by the tendency of those whose goals were not being met to drop out prior to data collection. Similarly, and for the same reason, participants may be artificially more likely to report goal attainment than former learners, if former learner groups include those who dropped out for lack of goal attainment.

Of the seven studies that measured learners' personal goal attainment, six found positive impact and one found little impact. The one study that found little impact (21WO) measured goal attainment as the difference between what subjects said their goals were at program entry and the goals they reported they achieved at follow-up. Thus missing data and changes in goals were recorded as non-attainment.

The 1980 National Evaluation (2N) asked respondents to indicate whether the goals they had set for themselves upon their entry into the program had been met. Only 17 percent reported their goals had not been at least partially met.

Conclusion: Learners perceive that their personal goals are achieved through participation in adult literacy education.

CONCEPTUAL, DESIGN, AND METHODOLOGICAL PROBLEMS

With few exceptions, the studies analyzed for this report were flawed in ways that severely compromised the validity and utility of their findings. At best, public funds have been wasted. At worst, important planning and policy decisions have been made on inaccurate and misleading data. There are at least six causes of the flaws that are inherent in the studies reviewed for this report: inaccurate or incomplete data; over-reliance on self-report data; lack of adequate controls; lack of valid, reliable, and appropriate tests; poor quality research reports; and lack of relevant standards. Clearly, these problems must be avoided in future outcome and impact research if useful knowledge is to result.

Inaccurate and Incomplete Data

An overwhelming majority of the studies, including all those conducted at the national and state levels, collected learner data through adult literacy education programs. Some, most notably the NEAEP (1N) and the Tennessee Longitudinal Study (6S), were open and honest about the data collection problems they experienced. Others were either less candid or simply assumed their data were accurate. Common problems with data collected through programs were: inaccurate learner records, failure to pre- and posttest at specified intervals, administration of inappropriate levels of tests, failure to test, high attrition of subjects between pre and post data collection, programs' withdrawal from the study before data collection was complete, and failure to forward data to researchers in a timely fashion. These problems were most severe for large-scale outcome assessments like the NEAEP that required complete data in order to maintain their protocols for insuring generalized findings.

In designing data collection protocols, researchers must realize that the programmatic context of adult literacy education differs substantially from the more traditional evaluation venues of elementary, secondary, and higher education. Because of open enrollments, it is difficult to pretest learners before they have engaged in instruction, and because of sporadic attendance, fear that testing may alienate learners, and high dropout rates, it is difficult to systematically posttest adequate proportions of those pretested. Many programs lack staff experienced in testing and data collection, and, as a result, inappropriate levels of tests are administered and data collection protocols are violated.

Post-data collection at realistic intervals is a major problem. If actual hours of instruction are employed as the interval criterion, an accurate record of instructional hours received is necessary, and research subjects must be available when posttests are due. Sporadic attendance and dropout render this precision problematic. If larger units of elapsed time, such as weeks, months, or years, are used as the

criterion, there is the problem that the actual hours of instruction learners receive per week or month can vary significantly. Except for GED completion, there typically is no concrete marker for successful completion of adult literacy instruction. Thus it is difficult to assess whether learners who have exited programs are successful completers or program dropouts. Moreover, because adult literacy learners tend to be a geographically mobile population, when collecting data it is extremely difficult to locate those who have terminated programs.

Self-report Data

Most studies relied on self-report rather than objective data for their findings. Lacking pre-data, studies that used a post-only design had little choice. For all studies, collecting self-report data was the obvious option given that hard data on such variables as employment, earnings, welfare reduction, and continued education were simply not available. The exceptions were the welfare and workforce studies that were able to obtain objective data from state welfare and employment records.

The validity of self-report data is a matter of concern for researchers. When at post-data collection learners are asked to remember feelings or events they experienced at program entry, the accuracy of recall is an issue, especially after periods of long duration. More important, there is a normal tendency for respondents to inflate the value of experiences, like adult literacy education, that have high social value and which entail significant sacrifices on the respondents' part. By and large, self-reported gains were higher for the studies reviewed here than were gains measured by objective data. For example, when learners were asked to report their gains in basic skills, positive attribution of gain far outweighed negative attribution. However, when gains in basic skills were measured by tests, the gains found were small or, as with two studies, non-existent. Similarly, when the Tennessee Longitudinal Study (6S) measured gains in self-esteem with the Rosenberg Self-Esteem Scale, the gains were very small. Yet the self-reported gains were very high. Of course, it is quite possible that the tests used were not entirely appropriate or were not sensitive enough to measure the gains learners recognized in themselves.

Self-reported gains in employment tended to be moderate. For the 1973 National Evaluation (3N) they were about 10 percent after 18 months and for the New Jersey Study (8S) the overall gain in employment was 13 percent. However, when gains in employment were measured using actual employment records, they were small in the case of the Texas JOBS Study (15W) and non-existent in the case of the Washington Workforce Study (5S).

Lack of Controls

Many of the variables used in impact studies are influenced significantly by context and time. The state of the economy affects employment. Over time, earnings normally increase as job tenure increases, and dependence on AFDC declines as welfare recipients' children grow older. How, then, is it possible to infer that gains are caused by adult literacy education rather than by unknown contextual or time-related factors?

The most logically defensible way is through the use of an experimental design in which subjects are randomly assigned either to instruction or a control group. Because random assignment insures that participant and control groups are equivalent in all respects except participation, differences in group pre- and post-gains can be attributed solely to the impact of adult literacy education.

Although an experimental design is ideal when inferring causality is a goal, of all the studies reviewed, only the California GAIN Study (14W) and the National Even Start Evaluation (22FL) were able to employ one. As a result, the extent to which adult literacy education causes impact is a very difficult question to answer. In the real world, experimental designs are rare because they are extremely difficult and expensive to employ. Random assignment is frequently impossible, especially on a large scale, and it is often difficult to collect data from control group members unless they are substantially paid. Moreover, extensive attrition from participant groups biases results, and attrition is a common occurrence in adult

literacy education. Experimental designs have their limitations. Although they do permit researchers to infer causality, they generally do not explain how participation produces the positive effects found.

If a true control group cannot be employed, use of a comparison group is a second-best option. However, when comparison groups are used it is vital that the comparison be appropriate. Two studies (5S, 15W) used matched comparison groups in which, through complex procedures, a participant group was compared to a non-participant group that was similar with respect to potent background and demographic variables. Although matched comparison groups provide a useful unit of comparison, they are not equivalent to participant groups in all respects. Indeed, by virtue of volunteering to participate in adult literacy, an act that is protracted and arduous, participants demonstrate a certain level of motivation that does not necessarily apply to comparison group members. Was it participation in adult literacy that caused the social and economic impacts noted in Washington (5S) and Texas (15W) or was it high personal motivation? The answer is not known.

Two studies used comparison groups that were inappropriate. One (10S) used former participants for the comparison, but, because it could not be determined to what extent this group was composed of successful completers as opposed to dropouts, the comparison was flawed. Another study (4N) employed program dropouts as a comparison. However, this comparison too was inappropriate because the participant and comparison groups may well have differed with respect to their motivation and/or ability to participate.

Valid and Reliable Tests

The most commonly used tests for basic skills gain were the TABE and the CASAS tests. Use of these tests, or any other for that matter, raises important issues. The first is the appropriateness of the test for the instruction given. The TABE, for example, is a standardized, general measure of basic reading skills. Although this test may be appropriate for programs that focus on basic skills development, it may be inappropriate for programs that focus on contextualized literacy, such as workplace literacy where the content of instruction is presumably job-specific. Nevertheless, one of the few workplace literacy studies that systematically measured tested learning gain (21WO) used the TABE.

The contradiction that learners generally report large gains in basic skills while tests show much smaller or no gains raises the possibility that the tests available are not sensitive enough to measure important gains perceived by learners. Although the validity and reliability of tests such as the CASAS, TABE, and TALS have been established, for expediency, many studies used only portions of tests, and for portions of tests, validity and reliability are serious issues.

Many tests, the TABE for example, are calibrated into levels. If an improper level of the test is administered, erroneous and misleading results can occur. For example, if a level of the test that is too difficult for the population tested is given, the “bottomed-out” scores that result can only go up. Thus the chance factor, which is inherent in all tests, works in favor of increased scores at posttest, and artificial gains result. If the test level is too easy, artificial declines will result. These “ceiling and floor” effects may have distorted findings for several of the studies reviewed (1N, 3N, 14W, 22FL).

Test administration affects test validity. For timed tests, time specifications must be followed. Tests must be administered at appropriate time intervals. If the interval is too short for reasonable gains to be expected, artificially low gains result, and if the interval is too long, learner attrition often takes an unacceptably high toll.

Poor Quality Reports

Reports of outcome and impact research must contain certain information about design and methodology so that researchers and policy makers can assess the credibility of the research. This critical

information includes basic research design, sampling procedures, data collection procedures, response rates, test validity and reliability, and time intervals between pre- and posttesting. Of the 115 studies reviewed for this report, about 40 percent lacked this vital information and had to be eliminated from further consideration. Since the credibility of their findings could not be assessed, they were essentially worthless.

Lack of Standards

There are two equally important processes involved in credible outcome and impact research. The first requires competent design, execution, and reporting so that valid findings result, while the second entails judgment regarding whether findings represent program success. There are at least two bases for judgment and both are problematic in adult literacy education. The first is comparative judgment in which the gains documented by one study are compared to the gains of similar studies. However, because the outcome and impact studies of adult literacy education vary greatly in design, procedures, and populations, meaningful comparative judgment is confounded. The second is normative judgment. In this case findings are assessed in relation to established standards. However, for adult literacy education, such standards do not yet exist.

This problem is particularly evident in tested learning gain. The problems and issues of testing notwithstanding, how much tested gain should be expected after a given period of instruction? Certainly any gain is not sufficient, for very small gains can hardly be indicative of program success. Because the 23 “most credible” studies analyzed here do not yield a comparative standard, and because there are no normative standards for tested learning gain, we simply do not know.

IMPLICATIONS FOR POLICY

Finding the Answer

The essential task for any outcome and impact study is to assess programmatic effectiveness. Yet as the case studies attest, flaws in study design and execution have severely impeded the ability of most studies to make such an assessment. None of the three national evaluations of the Adult Education Act-sponsored program were able to do so convincingly. To our knowledge, after more than 25 years, only nine credible state studies have been conducted, and all are limited in one way or another.

Although this study has attempted to make reasoned conclusions about program effectiveness from the evidence embodied in the case studies taken together, the data are not sufficient to settle the question of effectiveness at a level of certitude the public deserves. Furthermore, a significant amount of public funds have been wasted on individual studies that were unable to yield valid conclusions. More and better outcome and impact research is needed.

Economic Impact

Increasingly, adult literacy education is being held accountable for its economic outcomes and impacts, and the news from this analysis is basically good. The evidence presented here suggests that, at least in the short-term, adult literacy education does produce employment-related benefits, although the extent to which the jobs acquired are good jobs is still a question. Likewise, the evidence suggests that adult literacy education has a positive short-term impact on earnings. Short-term gains are but the tip of the iceberg, however, and long-term gains need to be investigated.

It is possible, for example, that there are intergenerational effects to participation, as would be the case if increased earnings were invested in the health and education of children. If, 20 to 30 years after

completion of adult literacy education, participants' children were shown to have higher levels of education than the children of non-participants, an important social benefit would have been demonstrated. Similarly, the long-term cumulative effects of adult literacy need to be studied. It may well be that the power of adult literacy education lies not in its function as an end that produces immediate gains but in its function as an enabling means to a wide range of other benefits that, when obtained, yield still more benefits. A hypothetical case in point would be successful learners who go on to further education, subsequently obtain high-level employment, and end up increasing their incomes substantially. Such cumulative gains would not even begin to accrue until five or more years after completion of adult literacy education.

Basic Skills

The findings of this study suggest that while learners generally perceive they have gained in basic skills as a consequence of participation, the evidence from testing is insufficient to infer gain. The implications for policy depend somewhat on one's perspective. According to one perspective, if learners perceive they gain in basic skills, achieve economic benefits, and meet their personal goals, the adult literacy education program is functioning effectively and tested learning gain need not be a major point of concern. Our findings support program effectiveness according to this perspective. An alternative perspective, however, is that adult literacy is first and foremost an educational program that focuses on reading, writing, and computation. It should, therefore, be held accountable primarily for how much its clients learn in these areas, and, since they are an objective measure, tests are an important and valid measure of impact. According to this perspective, there is substantial room for improvement with respect to basic skills.

Clearly, what is meant by basic skills is an important policy consideration. If literacy is defined as a set generalized reading, writing, and computational skills akin to what is taught in elementary education, then standardized test results are a relevant criterion. If, however, literacy is defined as the foundation abilities learners need in order to function in specific contexts—abilities that are context-specific and more wide-ranging than “academic” reading, writing, and mathematics—then standardized tests are anathema and learners' perceptions of their basic skills gains may be the most acceptable criterion for measurement.

Self-image

Of all the evidence presented in this study, the evidence that adult literacy education produces gains in positive self-image (and similar constructs such as self-confidence and self-esteem) is the strongest. From a policy perspective, this is somewhat perplexing, since feeling better about one's self is not a stated goal of the federal adult literacy program. For this reason, perhaps, the question is not whether learners' self-image improves with participation, but whether improved self-image is a lasting effect that promotes impact in the human capital domain and reduces social dependence. On the one hand, it could be that reported gains in self-image are no more than the short-term elation people normally experience when they have completed difficult and protracted tasks. On the other, the effects might be lasting. Although lasting gain has the potential for stimulating successful learners' motivation to succeed in beneficial activities they otherwise might not have attempted, we do not know whether gains in self-image are lasting, and the relationship between enhanced self-image and the attainment of other benefits needs to be established.

Children's Education

In the studies reviewed, impacts on children's education were generally measured by such variables as the extent to which participants read to their children, whether they attended PTA meetings, and the frequency with which they interacted with children's teachers. Based on respondents' self-reports and these measures, it was concluded that adult literacy education has a positive impact on children's education. These variables, however, are essentially surrogates for the long-term effects on children that

would be expected if genuine impact were to be demonstrated, effects that might include children's more positive attitudes toward education, improved school performance, higher secondary school graduation rates, and increased enrollment in higher education. These long-term effects need to be established.

Personal Goal Attainment

As Beder (1991) notes, for many adult literacy education practitioners, that learners obtain their personal goals is the most important objective of instruction. There is a practical as well as philosophical reasons for this. If voluntary learners cannot attain their personal goals, there is no reason for them to attend. On the other hand, policy makers have increasingly stressed the human capital outcomes of adult literacy education, outcomes such as increased employment and income. Those who focus on learner goal attainment note that many learners are not motivated by attainment of jobs and income, and there is evidence to support this assertion (Beder & Valentine, 1987; Mezirow, Darkenwald, & Knox, 1975; Washington State Training and Education Coordinating Board, 1997). Those who stress economic outcomes, however, consider adult literacy to be a workforce training program that must be coordinated with job training and welfare policy.

Given this difference in program expectations, a critical issue is the extent to which the two perspectives are compatible. Can the adult literacy education program achieve human capital objectives while still meeting learners' personal goals? The conclusions of this study, which suggest impact in both arenas, provide evidence in the affirmative.

Accountability, Standards, and Judgments

Much has been said about the need for relevant standards on which to base evaluative judgments of program impact. As noted earlier, the problem is that standards derive from goals and goals derive, at least in part, from conceptions regarding the definition and purposes of adult literacy education. On purposes and definitions there is no consensus. Clearly, to recommend that consensus be achieved is futile, like flying a kite in a tornado. What then should be done? One policy alternative is maintenance of the status quo and the resulting confusion over goals, which has characterized the conduct of adult literacy education since the inception of the federal program. This situation has resulted in a pluralistic system that tolerates ambiguity on the one hand but defies accountability on the other. Another alternative would be the imposition of uniform national standards and measures through legislation or bureaucratic fiat. Although outcome expectations would be clear, the goals currently served by adult literacy education would be necessarily narrowed, most likely in the direction of human capital, and learners who do not have goals that are sanctioned by the system might be under-served. A third alternative is voluntary compliance. In this model, programs would be able to choose from several models and corresponding standards and measures. The goals of the National Institute for Literacy's (NIFL) Equipped For the Future (EFF) program (Stein, 1995) are one alternative. These goals are broad-based and comprehensive and have been generated through systematic qualitative research. The development of EFF performance standards is under way. Models and indicators might be developed for basic literacy skills approaches, workplace literacy, and so on. The issue is whether a voluntary system of outcome accountability would satisfy the current pressures for national accountability evident in Congress and state legislatures.

RECOMMENDATIONS

If the outcomes and impacts of adult literacy education are to be known and accountability is to be demonstrated, a three-part research strategy is recommended. The first is the implementation of a national outcome and impact reporting system that would provide useful outcome information on an ongoing basis. This system should be designed to produce data that is useful for planning as well as for accountability. The second is a comprehensive national longitudinal evaluation that would measure long-term impact at a level of certainty that would be difficult to achieve in a national reporting system. The national assessment should include experimental design and qualitative components. The third is systematic funding and

improvement of state and local outcome studies needed to supplement a national reporting system and a major longitudinal study.

An Effective Outcome and Impact Reporting System

Although, as this report demonstrates, there is a substantial research literature that describes the inputs, processes, and outputs of adult literacy education, the research on program outcomes and impacts has been relatively meager and seriously flawed. As a result, it has been impossible to satisfy adequately the public demand for accountability and the need for valid planning information. How can this situation be changed?

Addressing this question, in 1996 the state directors of adult education recommended the development of a national outcome reporting system for adult literacy education, and a planning study was commissioned by the Office of Vocational and Adult Education (OVAE) and awarded to Pelavin Associates. In the report of this study, Condelli and Kutner (1997) discuss four models for measuring outcomes: comprehensive reporting, the follow-up survey approach, the data sharing/workforce model, and the menu approach.

Comprehensive reporting would entail the expansion and refinement of the current federal reporting system. This system relies on data collected from local programs that are forwarded to the federal government through the states. The accuracy of the current federal reporting system has been questioned (United States General Accounting Office, 1995), and if this strategy were followed, the current reporting system would have to be improved substantially. In the follow-up survey method, comprehensive reporting would be augmented by periodic follow-up surveys conducted by the states. In the data sharing/workforce model, client outcome data from job training, welfare, adult literacy, and other workforce-oriented programs would be pooled into a common database focusing on workforce performance. In the menu approach, programs would be able to choose the measures that were relevant to them from an established list. Each of these approaches has advantages and disadvantages that must be weighed before implementation, and, according to the Pelavin report, a process has been established for doing so.

Condelli and Kutner (1997) discuss several design issues that must be resolved if any of these models is to be effective.

- Uniform operational definitions of “program participant,” participant’s instructional level and outcome and impact variables must be established.
- Data collection intervals must be established and uniformly implemented.
- A viable sampling methodology must be designed and implemented.
- Protocols have to be established for reporting and dealing with missing data.

Although the findings of this research certainly corroborate a need to resolve Condelli and Kutner’s design issues, other equally important issues must be resolved if an adequate outcome and impact reporting system is to result.

Stated bluntly, under current conditions local programs lack the capacity to collect accurate and timely data. The problem is contextual. Open enrollment, part-time staff, sporadic attendance, high attrition rates, and learners’ reluctance to be tested confound systematic data collection. These problems are exacerbated by a weak programmatic infrastructure (Beder, 1996) produced by high staff turnover, reliance on part-time staffing, and low budgets.

If valid and reliable data are to be collected from programs, experience suggests that at least three conditions must be met. 1. Data collectors must be thoroughly trained and adequately compensated; 2. Data collection activities must be carefully monitored; and 3. Data collection protocols must be rigorously followed. This includes collecting data from absent learners and those who have terminated the program.

The reporting system would have to be standardized, carefully administered, and closely monitored.

- Standardization: If an objective of the reporting system is to permit comparisons between states and programs on performance variables studied, the items that measure these variables and the data collection procedures would have to be standardized.
- Administration and monitoring: Standardization would be meaningless unless procedures were established and implemented to insure that standardized protocols were followed. Establishing procedures that were both valid and reliable and within the capacity of state and local programs to implement would be a challenge. Substantial funding and staff development would be required.

A National Longitudinal Assessment

Whichever model suggested by Condelli and Kutner is eventually employed, the task of establishing an accurate reporting system for adult literacy education will be daunting—so much so that even the most optimistic would foresee problems in implementation that will, at least in the short term, limit the credibility of the data that results. To provide the most credible data possible, a national assessment of adult literacy education outcomes and impacts that will address and resolve the problems and limitations of the three previous national evaluations is needed.

- If there is “pay-dirt” in our understanding of outcomes and impacts it probably lies in establishing the long-term intergenerational and cumulative effects of adult literacy education. The best way to do this would be through a longitudinal evaluation in which the same subjects were followed up for a period of no less than five years. Such a study is currently being planned by the National Center for the Study of Adult Learning and Literacy (NCSALL).
- It is absolutely critical that such an assessment be carefully designed, and this task should be assigned to a highly knowledgeable and experienced design team. Because methodological issues are a major concern, the team should include technical experts familiar with the large-scale evaluation of social service programs. Experience tells us, however, that while inadequate methodology has been a problem in outcome and impact research, many assessments have failed primarily for logistical reasons. For this reason, the team should also include researchers and practitioners who are thoroughly familiar with the operational context of adult literacy education. The design team should report prior to the award of a contract to conduct the research, and its recommendations should be incorporated into contract specifications.
- Although it is difficult to fathom a national longitudinal evaluation that employs an experimental design, it may well be possible to conduct controlled studies on a limited number of programs as one component of a national effort.
- Although there have been several excellent qualitative evaluations of adult literacy education (Fingeret, 1985; Fingeret & Danin 1991), we were unable to identify qualitative outcome and impact assessments which met our definition of outcome and impact. If included as a component of a national evaluation, a qualitative assessment might answer critical questions about impact that are difficult or impossible to answer quantitatively: What is the meaning of impact from the perspective of successful learners? Are there important impacts of adult literacy that learners recognize in themselves, but are not amenable to quantitative measurement? How and to what extent do increased self-confidence and self-efficacy enable other positive changes in successful learners’ lives? These are just a few of the possible questions.

Improved State and Local Outcome Studies

Even if an effective national reporting system and a national longitudinal study were to be implemented, it would be necessary to conduct specific outcome assessments at the state and local levels. Such studies would be needed when states wished the depth and detail of analysis necessary to make

complex policy decisions. They would also be necessary to assess the outcomes of special literacy initiatives and idiosyncratic programs such as programs for the handicapped.

To improve state and local studies it is important that the methodological and design issues previously outlined in this report be resolved when studies are planned and conducted. Resolving these issues requires both technical skill and experience with the adult literacy context on the researcher's part. For this reason, the competency of the researcher must be a key consideration when studies are commissioned. Competent studies require adequate funding. Indeed, spending less than what a competent study requires results in worthless data and money wasted.

The most important ingredient in generating excellent outcome research is thoroughness and rigor. In most of the outcome studies reviewed here, the research was designed, the data were gathered and analyzed, the findings were reported, and the investigator was paid. We know something about effectiveness, perhaps, but nothing more, and we are left with the suspicion that the research was flawed in ways that were not revealed. In a thorough study, once initial findings have been generated the researcher explores all possible interpretations of causality. Are the findings really the result of participation in adult literacy or are they quirks of the methodology? Alternative interpretations of the data are explained and limitations are fully accounted for. Are the findings driven or shaped by factors that can be identified by secondary analysis? If so, secondary analysis is conducted to add to and clarify the understanding produced by the research. Design and methodology are explained in detail.

Thorough and rigorous outcome studies can and should contribute to our theoretical and conceptual knowledge of adult literacy education. In this way they can expand knowledge about adult literacy that goes beyond program impact. A case in point is that, despite all the outcome research in adult literacy education conducted thus far, we are left with a contradiction: Learners perceive that they gain substantially in basic skills through participation while data from tests conflict to the extent that it is difficult to conclude whether or not learners gain. Is this because learners recognize important gains in themselves that tests do not measure, or is it because gaining literacy skills is socially desirable and learners inflate their self-reports? If learners are being honest, perhaps the major gains in literacy are indeed contextual and personal. If the tests are right, perhaps the quality of instruction and staff development needs to be seriously examined. Either way one cuts it, the answer has important implications for policy and practice, and the answer is researchable within the parameters of outcome and impact research.

REFERENCES

Studies Analyzed and References Used in the Report

- Becker, W. J., Wesselius, F., & Fallon, R. (1976). *Adult basic education follow-up study 1973-75*. Kenosha, WI: Gateway Technical Institute.
- Beder, H. (1991). *Adult literacy: Implications for policy and practice*. Malabar, FL: Krieger Publishing Co.
- Beder, H. (1996). *The infrastructure of adult literacy education: Implications for policy*. Philadelphia, PA: National Center on Literacy, University of Pennsylvania.
- Beder, H., & Valentine, T. (1987). *Iowa's adult basic education students: Descriptive profiles based on motivations, cognitive ability, and sociodemographic variables*. Des Moines, IA: Iowa State Department of Education.
- Boggs, D. L., Buss, T. F., & Yarnell, S. M. (1979). Adult basic education in Ohio: Program impact evaluation. *Adult Education*, 29(2), 123-140.
- Brizius, J. A., & Campbell, M. D. (1991). *Getting results: A guide for government accountability*. Washington, DC: Council of Governors' Policy Advisors.
- Center for the Study of Human Resources. (1994). *Texas JOBS program evaluation: Final Report*. Austin, TX: Center for the Study of Human Resources, Lyndon B. Johnson School of Public Affairs, University of Texas at Austin.
- Cohen, J., Garet, M., & Condelli, L. (1996). *Reanalysis of data from the National Evaluation of Adult Education Programs: Methods and results*. Washington, DC: Pelavin Research Institute.
- Condelli, L., & Kutner, M. (1997). *Developing a national outcome reporting system for the adult education program*. Washington, DC: U.S. Department of Education, Office of Vocational and Adult Education.
- Darkenwald, G., & Valentine, T. (1984). *Outcomes and impact of adult basic education*. New Brunswick, NJ: Rutgers University, Center for Adult Development.
- Development Associates. (1992). *National Evaluation of Adult Education programs first interim report: Profiles of service providers*. Arlington, VA: Author.
- Development Associates. (1993). *National Evaluation of Adult Education programs second interim report: Profiles of client characteristics*. Arlington, VA: Author.
- Development Associates. (1994). *National Evaluation of Adult Education programs third interim report: Patterns and predictors of client attendance*. Arlington, VA: Author.
- Fingeret, H., & Drennon, C. (1997). *Literacy for life: Adult learners, new practices*. New York: Teachers College Press.
- Fingeret, H. A., & Danin, S. T. (1991). *They really put a hurtin' on my brain: Learning in Literacy Volunteers of New York City*. Raleigh, NC: Literacy South.
- Fingeret, H. A. (1985). *North Carolina adult basic education instructional program evaluation*. Raleigh, NC: North Carolina State Department of Community Colleges.
- Fisk, W. (1994). *Workplace literacy across the three phases of textile manufacturing. National workplace literacy demonstration project. Performance report*. Clemson, SC: Clemson University.

Greenleigh Associates, Inc. (1968). *Participants in the field test of four adult basic education systems: A Follow-up study*. New York: Author.

Gross, A. L., & Feldman, S. (1990). *Evaluation report: The workplace education program of the Central Labor Council and Consortium for Worker Education*. New York: Center for Advanced Study in Education, The Graduate School and University Center, City University of New York.

Hayes, A. E. (1997). *The power of Even Start family literacy: A summary of findings from a follow-up study*. Louisville, KY: The National Center for Family Literacy.

Heath, S.B. (1983). *Ways with words: Life and work in communities and classrooms*. New York: Cambridge University Press.

Jones, P. L., & Petry, J. R. (1980). *Evaluation of adult basic education in Tennessee, 1980*. Memphis TN: Memphis State University, College of Education.

Kent, William P. (1973). *A longitudinal evaluation of the Adult Basic Education Program*. Falls Church, VA: System Development Corporation.

Kirsch, I. S., Jungeblut, A., Jenkins, L., & Kolstad, A. (1993). *Adult literacy in America: First look at the results of the National Adult Literacy Survey*. Washington, DC: National Center for Education Statistics.

Mahaffy, J. E. (1983). *Impact evaluation of adult basic education program outcomes: ABE Evaluation Project final report*. Bozeman, MT: Montana State University, Department of Educational Services.

Martinson, K., & Friedlander, D. (1994). *GAIN: Basic education in a welfare to work program*. New York: Manpower Development Research Corporation.

Merrifield, J. (1998). *Contested ground: Performance and accountability in adult basic education* (NCSALL Reports #1). Boston: National Center for the Study of Adult Learning and Literacy, World Education.

Merrifield, J., Smith, M., Rea, K., & Shriver, T. (1993). *Longitudinal study of adult literacy participants in Tennessee: Year one report*. Knoxville, TN: Center for Literacy Studies, University of Tennessee.

Merrifield, J., Smith, M., Rea, K., & Shriver, T. (1994). *Longitudinal study of adult literacy participants in Tennessee: Year two report*. Knoxville, TN: Center for Literacy Studies, University of Tennessee.

Mezirow, J., Darkenwald, G., & Knox, A. (1975). *Last gamble on education*. Washington, DC: Adult Education Association of the U.S.A.

Mikulecky, L., and Lloyd, P. (1996). *Evaluation of workplace literacy programs: A profile of effective instructional practices*. Philadelphia, PA: University of Pennsylvania, National Center on Adult Literacy.

Mrowicki L., Jones, D., Locsin, T., Olivi, L., & Poindexter, C. (1995). *Workplace literacy in a total quality management environment for the manufacturing and financial services industry. Final performance report*. Des Plaines, IL: The Center-Resources for Education.

Paris, K. A. (1992). *Evaluation of the third year of the Wisconsin Workplace Partnership Training Program, March 1, 1991 through August 31, 1992*. Madison, WI: Center for Education and Work, University of Wisconsin, Madison.

Reder, S., & Wikelund, K. R. (1994). *Steps to Success: Literacy development in a welfare-to-work program*. Portland, OR: Northwest Regional Educational Laboratory.

St. Pierre, R., Swartz, J., Gamse, B., Murray, S., Deck, D., & Nickel, P. (1995). *National evaluation of the Even Start Family Literacy Program: Final report*. Washington, DC: U.S. Department of Education, Office of Policy and Planning.

St. Pierre, R., Swartz, J., Murray, S., Deck, D., & Nickel, P. (1993). *National evaluation of the Even Start Family Literacy Program: Report on effectiveness*. Washington, DC: U.S. Department of Education, Office of Policy and Planning.

Schexnayder, D. T., & Olson, J. A. (1995). *Texas JOBS program evaluation. Second year impacts*. Austin, TX: Center for the Study of Human Resources, Lyndon B. Johnson School of Public Affairs, University of Texas at Austin.

Solorzano, R. (1989a). *An analysis of learner progress from the first reporting cycle of the CALPEP field test: A report to the California State Librarian*. Pasadena, CA: Educational Testing Service.

Solorzano, R. (1989b). *An analysis of learner progress from the second reporting cycle of the CALPEP field test: A report to the California State Librarian*. Pasadena, CA: Educational Testing Service.

Stein, S. (1995). *Equipped for the future: A customer-driven vision for adult literacy and lifelong learning*. Washington, DC: National Institute for Literacy.

United States General Accounting Office. (1995). *Adult education: Measuring program results has been challenging*. Washington D.C.: Author.

Walker, S. M., Ewart, D. M., & Whaples, G. C. (1981). *Perceptions of program impact: ABE/GED in Maryland*. Paper presented at the Lifelong Learning Research Conference, College Park, MD, February 6 and 7, 1981.

Washington State Training and Education Coordinating Board. (1997). *Workforce training results net impact and benefit cost evaluation*. [Meeting NO. 52, March 25, 1997]. Author.

Young, M. B., Fitzgerald, N., & Morgan, M. A. (1994b). *National Evaluation of Adult Education: Executive Summary*. Arlington, VA: Development Associates.

Young, M. B., Fitzgerald, N., & Morgan, M. A. (1994a). *National Evaluation of Adult Education Programs fourth interim report: Learner outcomes and program results*. Arlington, VA: Development Associates.

Young, M., Hipps, J., Hanberry, G., Hopstock, P., & Golsmat, R. (1980). *An assessment of the state-administered program of the Adult Education Act: Final report*. Arlington, VA: Development Associates.

Studies Abstracted and Included in the Appendix

Aker, G., Jahns, I., & Schroeder, W. (1968). *Evaluation of an adult basic education program in a southern rural community*. Tallahassee, FL: Florida State University, Department of Adult Education. (ERIC Reproduction Services No. ED388780).

Alamprese, J. (1987). *Investing in change: Competency-based adult education in California*. Sacramento, CA: California State Department of Education.

Alamprese, J. (1993). *Systematizing adult education: Final report of the Connecticut Adult Performance Program (CAPP)*. Washington, DC: The COSMOS Corporation.

Alpena Community College. (1993). *Alpena Community College Workplace Partnership Project*. MI: Author. (ERIC Reproduction Service No. ED 356345)

Appalachian Adult Basic Education Demonstration Center. (1970). *Demonstration, developmental and research project for programs, materials, facilities and educational technology for undereducated adults. Long range follow-up of adult education participants. West Virginia state module. Final report series.* Morehead, KY: Morehead State University. (ERIC Reproduction Service No. ED054411)

Askov, E., Maclay, C., & Bixler, B. (1992). An intergenerational study of the impact of computer-assisted instruction with low-literate parents. In T. Sticht, M. Beeler, & B. McDonald (Eds.), *Intergenerational transfer of cognitive skills: Volume I: programs, policy, and research issues.* Norwood, NJ: Ablex Publishing Corp.

Association for Community Based Education. (1989). *ACBE evaluations of community based literacy programs: 1988-1989.* Washington, DC: Author.

Berkovitz, L. (1994). *Learning with East Aurora families. Project Evaluation.* Des Plaines, IL: Adult Learning Resource Center. (ERIC Reproduction Service No. ED 373187)

Boyter-Escalona, M. (1995). *Enhancing workers' skills for the workplace and for life. Worker education program final report.* Chicago, IL: Northeastern Illinois University.

Bramnick, B. (1981). *Assessment of adult basic education program impact.* Philadelphia, PA: School District of Philadelphia, Division of Adult Education.

Brown, E. (1990). *Evaluation of the R.O.A.D. to Success Program.* University Park, PA: Institute for the Study of Adult Literacy.

Claus, R., & Quimer, B. (1995). *Saginaw Adult Basic Education: Process and product evaluation report 1994/95.* MI: Department of Evaluation Services. (ERIC Reproduction Services No. ED388780)

Connecticut Business and Industry Association. (1992). *Connecticut Business and Industry Association's Statewide Workplace Literacy project: Final performance report.* Hartford, CT: Author.

Coplin, J. (1994). *Weck Ed. Weck Educational Development Program. Final Performance Report.* Durham, NC: Weck & Durham Community College.

Coston, C., & Dagard, C. (1995). *Champaign National Workplace Literacy Project. Final evaluation report.* Urbana, IL: Champaign Board of Education.

Cox, L., & Lane, C. (1981). *Evaluation of the adaptation of the personalized system of instruction to nontraditional adult learners. Final report. Volume III: PSI meets ABE.* Durham, NC: Research Triangle Institute. (ERIC Reproduction Service No. ED 208181)

Darkenwald, G. (1983). *Project F.I.S.T.: Functional in-service training. Submission prepared for the joint dissemination review panel.* Edison, NJ: Middlesex County College.

Darling, S. (1982). *Jefferson County adult reading program. Submission prepared for the Joint Dissemination Review Panel.* Louisville, KY: Jefferson County Adult Reading Program, Jefferson County Board of Education.

Darling, S., & Hayes, A. (1989). *Breaking the cycle of illiteracy: the Kenan Family Literacy program model.* Final report 1988-89. Louisville, KY: The National Center for Family Literacy.

Darling, S., & Hayes, A. (1996). *The power of family literacy.* Louisville, KY: The National Center for Family Literacy.

Educon Diversified, Inc. (1995). *Investing in Minnesota's future: An analysis of economic, societal, and personal impacts of literacy and adult basic education programs in Minnesota.* Belle Plain, MN: Author.

Fingeret, H., & Danin, S. (1991). *They really put a hurtin' on my brain*. Raleigh, NC: Literacy South.

Hayes, A. (1993). *Hawaii Family Literacy Pilot Project: Summary report 1990-93 and interim follow-up*. Louisville, KY: The National Center for Family Literacy.

Hudson, P., & Gretes, J. (1994). *New horizons: A national workplace literacy program. Final report and external evaluation impact study*. Conway, SC: Horry-Georgetown Technical College.

Koen, S., & Musumeci, M. (1984). *BES Adult Literacy Project*. Norwalk, CT: MATRICES Consulting Group.

Lafayette School Corporation. (1991). *Lafayette Adult Reading Academy (LARA) and St. Elizabeth Hospital employee literacy program. Final program report.* Lafayette, IN: Lafayette School Corporation and St. Elizabeth Hospital.

Lashof, J. (1992). *BEST: A learner-centered workplace literacy partnership of the Vermont Institute for Reliance and General Electric Aircraft Engines.* Rutland, VT: Vermont Institute for Self-Reliance.

Lipiec, J. (1993). *Verifying applied literacy skills (VALS) in ABE Programs.* Schnecksville, PA: Lehigh County Community College.

Lipscomb, J., & Cronin, C. (1994). *Tools for the 21st century. Southwest Alabama cooperative literacy project. Final evaluation.* Mobile, AL: University of South Alabama.

Mikulecky, L., & Lloyd, P. (1993). *The impact of workplace literacy programs: A new model for evaluating the impact of workplace literacy programs.* Bloomington, IN: Indiana University, School of Education.

Mott Community College. (1993). *Workplace literacy project. Computer aided instruction in basic workplace skills, May 1, 1991-October 31, 1992. Performance report. Executive summary.* Flint, MI: Author. (ERIC Reproduction Service No. ED 378308)

National Center for Family Literacy. (1992) *Dissemination process submission.* Louisville, KY: Author.

New York City Board of Education. (1985). *Adult basic education final report 1984-85. OEA evaluation report.* Brooklyn, NY: Author. (ERIC Document Reproduction Service No. ED 278793)

Orangeburg-Calhoun Technical College.(1995). *VISIONS for Greater Employment Opportunities. Final report.* Orangeburg, SC: Author. (ERIC Reproduction Service No. ED 379511)

Pacific Training and Technical Assistance Corp. (1974). *Evaluation of the community based Right to Read program.* Berkeley CA: Author.

Ramussen, Connie. (1992). *Workplace basic skills in the metal casting industry for world class process and technology.* Alexander City, AL: Central Alabama Community College.

Scheier, E., & Senter, D. (1969). *Evaluation of Learning 100: An adult basic education project in Bedford-Stuyvesant, 1967-68.* Huntington, NY: Educational Development Labs, Inc.

Seaman, D., Denyer, J. (1991). *Evaluation of a workplace literacy program: A cooperative effort between Houston Lighting and Power Company and North Harris County College.* College Station, TX: Texas A &M University.

Siefer, N., & Latkiewicz, J. (1992). *Third project evaluation report—summative for Salt Lake Community College.* Salt Lake City, UT: Salt Lake Community College. (ERIC Reproduction Service No. ED 355373)

Sherron, R. (1986). *A longitudinal evaluation of adult basic education in Virginia: 1970-1985.* Richmond VA: Virginia State Department of Education.

Smith, M., & Dalheim, Z. (1990). *Project READ. A study of twenty disabled adults.* Pittsfield, MA: Pittsfield Public Schools. (ERIC Reproduction Service No. ED 328726)

Snoddy, J. (1990). *Macomb Reading Partners research report.* Clemens, MI: Macomb County Library. (ERIC Reproduction Service No. ED 380560)

Turner, R., & Stockdill, S. (1987). *The technology for literacy project evaluation. A special report from the Donor Review Board.* St. Paul, MN: Saint Paul Foundation. (ERIC Reproduction Service No. ED 295028)

Wasson, J. (1965). *Evaluation of an eight week adult education program*. St. Paul: MN: Ramsey County Welfare Dept., Work and Training Program. (ERIC Reproduction Service No. ED021194)

White, G. (1990). *A comparative study of adult education Indianapolis/Richmond. Third party evaluation final report*. Indianapolis, IN: Indiana Opportunities Industrialization Center of America State Council.



National Center for the Study of Adult Learning and Literacy

The Mission of NCSALL

The National Center for the Study of Adult Learning and Literacy (NCSALL) will pursue basic and applied research in the field of adult basic education, build partnerships between researchers and practitioners, disseminate research and best practices to practitioners, scholars and policymakers, and work with the field to develop a comprehensive research agenda.

NCSALL is a collaborative effort between the Harvard Graduate School of Education and World Education. The Center for Literacy Studies at the University of Tennessee, Rutgers University, and Portland State University are NCSALL's partners. NCSALL is funded by the U.S. Department of Education through its Office of Educational Research and Improvement (OERI) and OERI's National Institute for Postsecondary Education, Libraries, and Lifelong Learning.

NCSALL's Research Projects

The goal of NCSALL's research is to provide information that is used to improve practice in programs that offer adult basic education, English to Speakers of Other Languages, and adult secondary education services. In pursuit of this goal, NCSALL has undertaken research projects in four areas: (1) learner motivation, (2) classroom practice and the teaching classroom practice and the teaching/learning interaction, (3) staff development, and (4) assessment.

Dissemination Initiatives

NCSALL's dissemination initiative focuses on ensuring that the results of research reach practitioners, administrators, policymakers, and scholars of adult education. NCSALL publishes a quarterly magazine entitled *Focus on Basics*; an annual scholarly review of major issues, current research and best practices entitled *Review of Adult Learning and Literacy*; and periodic research reports and articles entitled *NCSALL Reports*. In addition, NCSALL sponsors the Practitioner Dissemination and Research Network, designed to link practitioners and researchers and to help practitioners apply findings from research in their classrooms and programs. NCSALL also has a web site:

<http://gseweb.harvard.edu/~ncsall>

For a more information about NCSALL, please contact:

John Comings, Director
NCSALL
Harvard Graduate School of Education
101 Nichols House, Appian Way
Cambridge, MA 02138
(617) 495-4843
ncsall@gseweb.harvard.edu



National Center for the Study of Adult Learning and Literacy

**THE OUTCOMES AND IMPACTS
OF ADULT LITERACY EDUCATION
IN THE UNITED STATES
APPENDIX A: ABSTRACTS OF STUDIES REVIEWED**

by

**Patsy Medina
Rutgers University, Graduate School of Education**

**NCSALL Reports #6A
January 1999**

The National Center for the Study of Adult Learning and Literacy
Harvard Graduate School of Education
101 Nichols House, Appian Way
Cambridge, MA 02138

The work reported herein is supported by the Educational Research and Development Centers Program, Award Number R309B60002, as administered by the Office of Educational Research and Improvement/National Institute on Postsecondary Education, Libraries, and Lifelong Learning, U.S. Department of Education, through contract to Harvard University. However, the contents do not necessarily represent the positions or policies of the National Institute on Postsecondary Education, Libraries, and Lifelong Learning; the Office of Educational Research and Improvement; or the U.S. Department of Education, and you should not assume endorsement by the Federal Government.

Aker, G., Jahns, I., & Schroeder, W. (1968). *Evaluation of an adult basic education program in a southern rural community*. Tallahassee, FL: Florida State University, Department of Adult Education. (ERIC Document Reproduction Service No. ED 388 780)

Purpose: This evaluation was the second phase of an evaluative study of an adult basic education program for seasonal farm workers. Phase I used observations and a case-study approach to make recommendations about student screening, curriculum, instructional staff, and program expansion.

Variables: Outcome variables were grade-level progression and student satisfaction. Quality of learning was measured by a Likert scale on a questionnaire. Sources of greatest satisfaction were measured by indicating "specific subject matter," "learning in general," "prepare for a job," "relate to other people," "miscellaneous," and "no response." Other variables included the rate of student dropout, appraisals of instructional staff, and the relationship between program success and the characteristics of students and instructional staff.

Population: A random sample yielded 305 students ($N = 305$). Fifty-three percent were male. Most of the students were between 40 and 59 years old (56%). The mean pretest grade-level distribution was 2.6. A 33% representative sampling procedure assured a relative distribution of students in rural and urban centers.

Design: This study was a pre-post comparison.

Data Collection: The data were collected via standardized tests, questionnaires, and interview schedules. Existing student records were also used.

Instrumentation: The standardized test used to measure reading gains was not mentioned. Quality of learning was measured by a Likert scale.

Findings: All students who persisted in the program made some type of reading gain. The distribution range at pretest was 0.0-8.1. The posttest range was 1.1-11.4. The mean increased from 2.6 to 4.9; the median from 2.7 to 4.5. A majority gained from 1.0 to 3.9 grade levels during the course of the program. Students who received the lower pretest scores were more likely to achieve at the higher levels. A *t* test by program site revealed that each center secured a mean gain which could have been due to chance less than one time in a hundred.

Fifty-six percent of the students indicated that their learning resulted from participation in the program and that “very much” learning had taken place. The retention rate of students was 72.4%. Age was negatively associated with dropout. The older the student, the less the likelihood of dropping out.

Strengths: The research design was competent. Researchers examined the relationship between characteristics of students and staff and program success.

Weaknesses: The standardized reading test was not identified. Questionnaires seemed poorly designed and their methodology was not described. Program success was measured by only two criteria, grade-level gain and rate of dropout. There was no control group.

Alamprese, J. (1987). *Investing in change: Competency-based adult education in California*. Sacramento, CA: California State Department of Education.

Purpose: This study had two purposes: (a) to determine how 306 agency personnel had interpreted California’s competency-based adult education (CBAE) mandate, and (b) to ascertain the effects of the mandate on agency management, classroom functioning, and student outcomes. Only the portion dealing with student outcomes is abstracted here.

Variables: The variables were tested learning gain and retention.

Population: The population was adult literacy and ESL students in California. For tested learning gain, the population was 3,950 in 1983-84 and 6,413 in 1984-85.

Design: The design was pre- and posttesting for learning gain. Program and state records were used for retention.

Data Collection: Students were posttested after 100 hours of instruction. A sample of programs identified as high-implementing on the CBAE mandate was compared to a sample of low-implementing programs.

Instrumentation: The CASAS achievement test was used.

Findings: In 1983-84, the average gain on the CASAS was 6.8. In 1984-85, the gain for high-implementing agencies was 5.0; it was 3.5 for low-implementing agencies. Retention rates in 1983-84 were 50% ($N = 446$ classes), 59% in 1983-85 ($N = 561$ classes), and 49% in 1984-85 ($N = 984$ classes). At 100 hours, the

retention rate for high-implementing programs was 66.5% versus 47.4% for low-implementing programs.

Strengths: The sample was a respectable size.

Weaknesses: There were few outcome variables and no control group.

Alamprese, J. (1993). *Systematizing adult education: Final report of the Connecticut Adult Performance Program (CAPP)*. Washington, DC: COSMOS Corp.

Purpose: This document reported the formative evaluation of the Connecticut Adult Performance Program (CAPP), a five-year initiative to improve the state's adult literacy education. CAPP was a comprehensive system that included program management, staff development, etc. The evaluation was designed to collect information on the areas of program operation that required technical assistance and training; assess the impact of CAPP on program management and operations; and determine the effectiveness of the Bureau of Adult Education and Training in conducting the CAPP initiative. Although much of the evaluation and report focused on management issues, a portion focused on learner outcomes. That portion is reported here.

Variables: Measurements included tested growth in basic skills and learners' attainment of their individual goals.

Population: The sample consisted of adult literacy education students enrolled in a life-skills program adapted from CASAS. Students were pre- and posttested from 1987 through 1991. Numbers varied by year and by the level of instruction. In 1987-1988, there were 482 valid cases for reading and 197 for listening. In 1990-91, there were 1,068 valid cases for reading and 1,879 for listening. Students' attainment of personal goals was also measured through a survey. The total population was 1,271 for 1987-88; 1,771 for 1988-89; 2,714 for 1989-90; and 11,001 for 1990-91.

Design: The design was a longitudinal pre- and posttesting comparison for learning gain.

Data Collection: A multi-faceted data collection process was designed which included development of monitoring instruments, site observations, analysis of program and student assessment information and participation in state-sponsored meetings.

Instrumentation: The test, known as the CAPP Survey Achievement Pre and Post, was adapted from the CASAS test. The test measured gains in reading and listening. Data on the reliability and validity of the test were not reported. Data on learners' achievement of their personal goals were collected using a survey instrument, CAPP Goal Attainment. A yearly survey was used to measure attainment of personal learning goals.

Findings: Levels *a* and *b* gain scores were highest for reading in 1987 (9.9 for level *a*; 7.5 for level *b*), and declined in 1988 and 1989. In 1990 scores increased to 7.2 for level *a* and 6.1 for level *b*. Gain scores for level *c* ranged from 4.1 in 1987 to 4.6 in 1990. For listening, gain scores for level *a* in 1987 were the highest (10.1) and declined to the five-point range in subsequent years. Gains for level *b* were also highest in 1987 and declined in future years. Scores for level *c* were highest in 1988 and declined to about the 1.5 range thereafter. Although tests of statistical significance were not reported, the substantial sample size would suggest that the gains were significantly different.

Data from the survey showed yearly retention figures of 63% for 1987, 59% for 1988, 66% for 1989, and 42% for 1991. In 1990, 15% of all students (both retained and no longer active) achieved their personal goals and left the program, 14% left the program because of barriers, and 30% left the program for unknown reasons.

Strengths: The sample size was relatively large. Yearly data were collected over four years.

Weaknesses: Fine points of design and methodology were not reported. There was no control or comparison group. The test for learning gain was unique; hence, gain can not be compared with data from other programs. The time period between pre- and posttesting was not reported.

Alpena Community College. (1993). *Alpena Community College Workplace Partnership Project*. MI: Author. (ERIC Document Reproduction Service No. ED 356 345) Alpena: MI.

Purpose: This project was funded by the National Workplace Literacy Program from May 31, 1991 to January 31, 1993. The partnership included the college, Besser Company, Baker Company, and Fletcher Paper Company. The work-site classes included reading, mathematics, and communications. Class content was based on the needs of adult learners. Several of the mathematics courses were rather advanced (e.g., trigonometry and algebra). Classes were held at the work

site and consisted of 16 contact hours. At one site the classes were on-the-clock; at two sites they were off-the-clock.

Variables: Reading gains and attitudes regarding the classes (including some measures on how the classes impacted the participants) were measured.

Population: Two hundred and thirty adults participated in the classes. The majority (176) worked for the Besser Company.

Design: The design was pre-post comparison and evaluated by an independent consultant.

Data Collection: Project staff collected the data. However, for a variety of reasons, they discontinued administering the tests halfway through the project. The pre- and posttest data were analyzed using *t* tests for related samples. Each subject area was analyzed separately.

Instrumentation: The Educational Testing Service's Tests of Applied Literacy Skills (TALS) were used to measure reading and math gains. Attitudes were measured quantitatively by a Likert scale in a questionnaire and qualitatively by an instrument with open-ended questions.

Findings: Analyses of the pre- and posttest data of reading, math, algebra, and communication scores showed no significant increases, although the posttest means were higher. Learners' self-esteem increased. All responded positively to the statement, "You feel more confident about your learning abilities because of this class." Qualitatively, the most common positive response involved increase in self-confidence.

Strengths: The data collected generated quantitative and qualitative data.

Weaknesses: Data were incomplete. The collection process was haphazard. Data did not seem to be collected from all participants. Class time of 16 hours was too short to show significant gains. There was no control group. The sample size for each class was small.

Appalachian Adult Basic Education Demonstration Center. (1970).
Demonstration, developmental and research project for programs, materials, facilities and educational technology for undereducated adults. Long range follow-up of adult education participants: West Virginia state module. Final report series. Morehead, KY: Morehead State University. (ERIC Document Reproduction Service No. ED 054 411)

Purpose: The Long Range Follow-Up Survey (LRFUS) had two main objectives: (a) to follow a sample of ABE graduates for three years and (b) provide a three-year supportive program to reinforce the ABE experience. Inherent in both objectives was the intent to create another model (rather than standardized reading tests) for assessing impact. The second year of implementation is described here.

Variables: LRFUS sought to measure the relevance and impact of ABE on the economic, family, educational, and community lives of the former students.

Population: Eighty-five former ABE participants were drawn from rural and urban areas of three counties in West Virginia. By the study's second year, nine people were no longer part of the sample. Group I consisted of 30 men who had taken part in the original field study and were enrolled in compulsory Title V and WIN programs. Group IV consisted of 25 men who were assigned a combination of ABE and diesel mechanic class. Groups II and III were 30 female former volunteers from Title III ABE. Most of the students from all four groups were Level III ABE participants.

Design: This follow-up study utilized a case-study design.

Data Collection: The sample population was divided into four groups. Each group was followed by a different teacher-counselor. Originally the design called for the LRFUS subjects to be compared with a matched control group, but the researchers decided against it.

Instrumentation: There was no instrumentation. However, in supporting the LRFUS subjects (one of the objectives), the researchers used a combination of newsletters, picnics, and personal contacts. The study used a 15-item questionnaire to gather follow-up data.

Findings: This follow up was a description of the second year of a three-year study. No comprehensive analysis of results was provided.

Strengths: This study had no real strengths other than the objectives of this study, which seemed solid.

Weaknesses: Numerous case studies were provided with no analyses. There was no control group.

Askov, E., Maclay, C., & Bixler, B. (1992). An intergenerational study of the impact of computer-assisted instruction with low-literate parents. In T. Sticht, M. Beeler, & B. McDonald (Eds.), *Intergenerational transfer of cognitive skills: Volume I: programs, policy, and research issues* (pp. 149-158). Norwood, NJ: Ablex Publishing Corp.

Purpose: Computer-assisted instruction (CAI) courseware was developed by Penn State for parents of Chapter I children receiving compensatory education. The courseware used a whole-word approach. Its goal was expanded word recognition for nonreaders. A research study measured the impact of the courseware on parents and their children.

Variables: Reading achievement was the sole variable studied.

Population: Parents had to score below a fourth-grade level to be part of the sample. Fifty-two completed the required 20 hours of instruction and had pre- and posttest data. A control group included 24 subjects.

Design: This study employed an experimental design. Pre- and posttest scores were analyzed using analysis of variance for repeated measures. Courseware was used about 80% of the time during instruction.

Data Collection: Teachers administered the battery of tests upon completion of entrance interviews and after 20 hours of instruction.

Instrumentation: The Slosson Oral Reading Test (SORT), two sections of the Baltimore County Design (BCD), and the Bader Reading and Language Inventory were used.

Findings: The scores of the experimental group rose from 3.26 to 3.93 on the SORT, from 18.38 to 23.33 on the BCD-E2 (words in isolation), from 20.13 to 22.67 on the BCD-E3 (words in functional settings), and from 2.33 to 3.63 on the Bader Inventory. Significant differences between the experimental and control groups were found in all four measures (alpha level at .05).

Strengths: There was a control group.

Weaknesses: The sample size was small. Teachers administered the tests to their students. Only one variable was measured.

Association for Community Based Education. (1989). *ACBE evaluations of community based literacy programs: 1988-1989*. Washington, D.C.: Author.

Purpose: “To expand knowledge about these literacy programs, produce credible evidence of their impacts, identify exemplary models and methods.”

Variables: A convenience sample of nine community-based adult literacy programs was selected. Each program chose the variables for study from a list of ten. Possible variables included skills in reading, writing, math, and oral language (for ESL); reading and writing activities outside the classroom; fostering children’s intellectual activities; community involvement and contributions; self-esteem; self-determination; and retention.

Population: Data were collected from all students in each selected program. The sample size varied by program and the variable studied (in most cases the size was less than 30).

Design: Programs were rewarded with a \$5,000 mini-grant for participation. The design was longitudinal. Baseline data were collected no later than the new learners’ third session. Follow-up data were collected after about eight weeks and then again after about 30 weeks. At time two, five of the nine programs failed to have 65% of the baseline sample present for measurement. One program was dropped at time three for lack of a usable sample size.

Data Collection: Data were collected at the program by trained staff and generally took between one and two hours.

Instrumentation: The Adult Basic Learning Examination (ABLE) was used to measure reading and math; the California Achievement Test (CAT) to measure writing; and the Basic English Skills Test (BEST) to measure English proficiency for ESL learners. The Association for Community-Based Education (ACBE) developed a survey instrument to measure reading and writing activities outside the classroom, children’s intellectual and academic development, and community activities and contributions. The Culture-Free Self Esteem Inventory was used to measure self-esteem and an ACBE-developed instrument was used to measure self-determination. The same versions of the instruments were administered at the two follow-up sessions.

Findings: Six programs selected reading skills as a measured variable. Three of the six programs showed statistically significant gains. Five programs measured writing gains. Only one program showed statistically significant gains at the first follow up, which were not sustained at the second follow up. Two programs measured math gain. No program showed significant gains in math between the first and third data collections. Two programs measured oral English. One program showed a statistically significant gain over seven weeks; the other did not.

Learners' reading activities outside class were measured by eight programs. A "ceiling effect" on the baseline data collection precluded meaningful conclusions. Learners' writing activities outside of class were measured for five programs. Although there was a ceiling affect, it was less pronounced than for reading. Three of five programs showed significant gains.

Only one program measured children's intellectual and academic development. The moderate gains were not statistically significant. Community activities and contributions were measured at two programs. There were significant gains between time one and two, but the gains dissipated at time three for one program and the numbers were insufficient at the other.

Self-esteem was measured in all nine programs. Two thirds of the participants demonstrated low self-esteem at time one. Seven programs had statistically significant gains. Self-determination was measured at seven programs. Three programs had significant gains.

In six programs, 9% of the participants dropped out after two sessions, another 7% dropped out between six and eleven weeks, and an additional 31% left after five months.

Strengths: The study was carefully designed, and the methods and procedures were clearly explained. The measures and instrumentation were solid.

Weaknesses: There were no control or comparison groups. The small sample sizes yielded nonsignificant results for gains that would be significant in larger sample sizes. Because programs were permitted to chose the variables studied in their programs, some variables were studied in only a small number of programs. There was substantial attrition of subjects in the follow-up data collections.

Berkovitz, L. (1994). *Learning with East Aurora families. Project evaluation.* Des Plaines, IL: Adult Learning Resource Center. (ERIC Reproduction Service No. ED 373 187)

Purpose: This evaluation assessed Learning with Aurora Families (LEAF), a one-year family literacy program developed and implemented by Waubensee Community College. LEAF sought to recruit a minimum of 50 families who had four- and five-year-old children enrolled in at-risk programs. LEAF's objectives included increasing the basic and parenting skills of the adults as well as the readiness skills of the children.

Variables: The outcome variables included the impact of LEAF on the adults' basic skills, academic progress, career plans, and the evolvement of their goals (personal, educational, economic, and parental). It also ascertained changes in parental behaviors.

Population: LEAF included a total of 162 participants from East Aurora. Fifty-two participants were four- and five-year-olds, 51 were parents or other significant adults, and 59 were siblings. The 51 adults included ABE and ESOL students.

Design: LEAF operated in four-week sessions, each divided into five components: (a) adult education, (b) preschool or kindergarten classes, (c) parenting-life skills classes, (d) parent-child and family times, and (d) home visitations. The evaluation was pre-post comparison.

Data Collection: Project staff collected the data. Tests and questionnaires were administered before intervention and at the conclusion of the project. Data were analyzed by an independent evaluator.

Instrumentation: Reading gains were assessed by administering the Test of Adult Basic Education (TABE). A Likert scale was used to measure parenting goals. A family literacy questionnaire had three components: (a) a nine-item chart on which parents indicated the frequency of their parenting and readiness behavior with their children; (b) a nine-item chart to ascertain the adults' opinions about their children's education; and (c) four open-ended questions to measure project outcomes. The impact of parental behaviors and improvement in family relationships were measured qualitatively by home observations as well from the questionnaires.

Findings: The average attendance of basic skills students who began LEAF in November was 83 hours out of a possible 132 (62% attendance rate). Those who

began in January averaged 83 hours out of a possible 108 (65% attendance rate). Thirteen adults of the original 29 (42%) who were given the pretest also took the posttest. The reading gain average was .86. One participant took and passed the GED test, three were ready to take it, and one enrolled in a community college. Although positive impact on parenting behaviors was indicated, the data from the questionnaires were not adequately summarized. Pretests of the attitude questionnaire about parents' participation in the education of their children were misplaced. However, information from the posttests indicated that parents still felt that teachers were better qualified than they were to instruct their children. Furthermore, they seemed to view their children--but not themselves--as learners.

Strengths: LEAF was a thoughtfully designed program which considered the whole family. Some of the questionnaires were well designed, yielding both quantitative and qualitative data.

Weaknesses: The sample size was small and no control group was used. Analysis of the outcomes was inadequate.

Boyter-Escalona, M. (1995). *Enhancing workers' skills for the workplace and for life. Worker education program final report. Chicago, IL: Northeastern Illinois University.*

Purpose: From May 1993 through March 1995, Northeastern Illinois University, the Amalgamated Clothing and Textile Workers' Union, and local businesses cooperatively implemented a workplace literacy project. The goal was to provide 500 workers with literacy skills needed in the workplace through a partnership between small businesses, community organizations, a labor union, and a university. The project had two components: long-term classes in basic skills and job-specific workshops. The project had four levels of program outcomes: worker, company, union, and university. Only worker outcomes will be discussed in this abstract.

Variables: Studied were math achievement, English vocabulary and general communication skills, self-esteem, participation in union-related activities, enhanced worked performance, and opportunities for promotion.

Population: Four hundred and sixty-five employees participated in the long-term classes. The majority were immigrants from Latin American countries, with significant numbers from Asia. The short-term workshops had 271 participants. The majority were U.S. citizens.

Design: Although the report was primarily descriptive, it included a pre-post comparison component.

Data Collection: Data were collected by project staff throughout the project. Standardized tests were administered at intake, midpoint, and upon exit. An outside evaluator provided a summative evaluation.

Instrumentation: The TABE measured math achievement; the BEST Oral and Literacy Test measured English communication and writing skills; and individual learning plans assessed personal goals. Other instruments used to gather quantitative data were customized work texts, holistic writing samples, attendance records, and company and union records. Qualitative data were collected through student evaluations, progress reports generated by teachers, anecdotal comments, and open-ended interviews.

Findings: The average gain in English oral skills was 5.6 points for ESL students. The average English writing skills gain was 1.0. Those who scored highly on the BEST Oral test were given the BEST Literacy to measure higher-order language skills, averaging a gain of 5.2 points. Twenty students were promoted at their job sites.

Strengths: The project was well designed. Evaluation methods were comprehensive.

Weaknesses: There was no control group. Descriptive analysis of the data was weak.

Bramnick, B. (1981). *Assessment of adult basic education program impact*. Philadelphia, PA: School District of Philadelphia, Division of Adult Education.

Purpose: The impact of ABE was assessed in a sample of classes in 1977-78.

Variables: The variables were self-reported gain in reading, math, writing and consumer skills; enrollment and retention; voting behavior; community participation; and employment.

Population: The sample was selected from 17 classes that met for eight hours per week at ten sites in Philadelphia. This population included 1,216 students; 25% were randomly selected for the study, excluding students with less than 12 hours of instruction. The sample size was 267 students.

Design: The design was post only self-report.

Data Collection: Subjects were surveyed by telephone. The survey included 93 questions. Approximately one third of the sample was reached; the others could not be contacted. The final number was 89. On average 2.6 calls were needed to reach the subject and the interview lasted an average of 10 minutes. The time period between enrollment in the program and data collection was not reported.

Instrumentation: Checklist questionnaires were developed specifically for this project.

Findings: Seventy-eight percent of the students reported improved skills in reading, 81% in math, 80% in writing, and 50% in consumer skills. Fifty-eight percent reported being enrolled in courses other than ABE and 79% planned future enrollment. Eight-four percent were registered to vote, with 77% voting in 1978 and 84% in 1979. Fifty-four percent increased community participation, 47% increased social involvement, 70% felt more confident talking, 33% used the library more frequently, 67% read more books, and 52% read more magazines. Thirty-five percent of those who had children reported increased attendance in PTA meetings, 35% reported increased interaction with their children's teachers, and 70% reported that they helped their children more with homework. Thirty-six percent of the students were employed full or part time. Of the employed, 81% reported that ABE helped them to get a job, 19% reported a promotion, and 44% reported increased job security or pay.

Strengths: The range of impact variables was wide. A detailed report of methods and procedures was provided.

Weaknesses: The sample size was low. There was no control or comparison group. Data were exclusively self-reported. Amount of instruction between enrollment and data collection was not reported.

Brown, E. (1990). *Evaluation of the R.O.A.D. to Success program*. University Park, PA: Institute for the Study of Adult Literacy.

Purpose: R.O.A.D (Real Opportunities for Advancement and Development) to Success was a pilot project developed through a partnership of the Pennsylvania Department of Education, the American Federation of State, County and Municipal Employees, and the Institute for the Study of Adult Literacy. The goal was to improve the reading skills of transportation workers to help them pass the

commercial driver's license (CDL) test. The major components of R.O.A.D. to Success were interactive computer courseware and print-based materials. The evaluation ran from June 1989 to January 1990. Participant outcome variables were measured, and attention was given to the adequacy of curriculum materials and the process by which they were developed and used.

Variables: The outcome variables were reading gains, gains and passing scores on the CDL test, and attainment of the CDL. Several variables related to participant changes in attitudes regarding the program, materials, CDL test, and being a student were also measured.

Population: The treatment group consisted of workers from the Pennsylvania Department of Transportation (PennDOT) who had previously failed the CDL exam. Sixty-eight workers attended the classes and 58 completed 100 hours of instruction. The average age of the students was 41, with an average reading grade of 5.5 and an average of 11 years of formal schooling. The sample was selected in the following manner:

- All PennDOT workers from four counties were administered the Quick Assessment Test (QAT).
- Those who scored less than 80 were given the TABE Locator.
- Those scoring less than 23 (8.6 grade level or lower) were eligible for the program.
- Of 77 eligible workers, 68 enrolled.
- They were given Criterion References Tests (CRT), which measured basic educational skills in context.

A comparison group was chosen in a similar manner for a control group of 10. The comparison group had access to the CDL manual and workbook but not the R.O.A.D. materials or classes.

Design: This study had an experimental design with a comparison group. A chi-square test was computed to measure significant differences in workers' attitudes.

Data Collection: The QAT was administered by Institute staff. The TABE Locator was administered by the Pennsylvania Department of Education and the TABE Survey by project counselors. The tests and survey were administered before and after the program.

Instrumentation: The QAT, CRT, and TABE were given to measure basic skill levels. An attitude survey was also administered. Various instruments (e.g.,

observation forms, teacher response forms) were developed to collect formative data.

Findings: The average CRT score increased from 70.9 to 79.8, a significant difference. However, change between treatment and comparison groups was not significant. On the QAT, the average scores for the treatment group increased significantly from 33.4 to 49.7. A significant difference also appeared between the scores of the treatment and comparison groups. On the TABE, scores increased significantly at the .001 level. Average scores increased from 643 to 700. No posttest data on the TABE were available for the comparison group. Although shifts were apparent in workers' attitudes toward the intended objectives of the program, most questions showed no significant differences.

Strengths: This project was adequately designed and evaluated. It had a comparison group. Various instruments and methods were used to collect data. The data were well analyzed. Scores between tests were correlated.

Weaknesses: Only a few variables were studied. The sample size was low. The comparison group size was significantly smaller than the sample size.

Claus, R., & Quimer, B. (1995). *Saginaw adult basic education: Process and product evaluation report 1994/95*. Saginaw, MI: Department of Evaluation Services. (ERIC Document Reproduction Service No. ED 392 978)

Purpose: This evaluation of the Saginaw City School System's ABE programs studied the process of instruction and products (outcomes) of that instruction during 1994-95. This abstract describes only student achievement and goal attainment data.

Variables: The outcomes component focused on student enrollment, recruitment, dropout prevention, student achievements, objective assignment, and attainment.

Population: The 1,096 students enrolled during the 1994-95 school year were classified as handicapped, ESL, incarcerated, senior citizen, and regular. Of the latter classification, 629 regular ABE students attended at least 12 hours of instruction. They were assigned individual objectives.

Design: This study was pre-post comparison. Data from the 1994-95 school year were compared to data from 1992-93 and 1993-94, and focused primarily on the assignment and completion of student objectives.

Data Collection: Data that described student performance on individually assigned goals were recorded by teachers on the Adult Basic Education Reporting Form. Data on basic skills achievement were collected by the program supervisors and sent to the evaluators. Data were also collected from teacher observations.

Instrumentation: Reading gains were measured by the TABE. It was not clear at what point after 12 hours of instruction the posttests were administered.

Findings: Eighty-seven percent of students who received at least 12 hours of instruction attained their goals, with 64.8% attaining mastery on over 75% of their objectives. (These were objective outcomes for all categories of students.) Four hundred and eight of the regular ABE students (64.9%) completed at least 75% of their objectives. Compared longitudinally, 449 (74.0%) students attained 75% of their goals in 1992-93 and 1993-94. One hundred and eighty-eight (22.7%) students who received at least 12 hours of instruction realized one grade-unit gain.

Strengths: There were few strengths in the outcomes component of this study. However, some of the observational data provided a good indication of what actually occurred in the classroom.

Weaknesses: Attrition was high. There were no rigorous and standardized procedures to attain data on reading gains. There was no control group. The objectives assigned and attained were not delineated nor categorized in the data analyses. Therefore, this data can not be used to make informed conclusions by others.

Connecticut Business and Industry Association. (1992). *Connecticut Business and Industry Association's Statewide Workplace Literacy Project: Final performance report*. Hartford, CT: Author.

Purpose: This report focused primarily on the goals, objectives, and activities of the workplace literacy project. The project centered on nine companies that had identified literacy needs as a result of implementing TQM processes. Only the outcome data are reported here.

Variables: Expected wage gain within six months of completing training and tested learning gain were studied.

Population: The sample consisted of participants in workplace education in nine Connecticut companies.

Design: Companies were asked whether they expected to increase the wages of successful learners and, if so, by how much. Seven of the nine companies administered posttests to determine learning gain.

Data Collection: How the tests were administered was not described. The time interval between pre- and posttesting ranged from 12 weeks to 36 weeks.

Instrumentation: There was little information on the tests. They measured different competencies for each company and were scored in different ways.

Findings: Only five companies supplied data on expected wage increases, of which two provided reliable data. At one site the expected increase was .52, at the other it was .33. Coordinators in neither company were able to stipulate that the gain was necessarily attributable to participation in the literacy program. Gain scores at five companies were computed for 185 participants (29% of the total). Forty-nine percent did not gain at all, 15% gained in the low range (1-20 points), 22% gained in the middle range (21-40 points), and 15% gained in the high range (41-56 points).

Strengths: The one strength was the study's workplace context.

Weaknesses: The sample size was small. There was no established validity or reliability of the tests and the tests were idiosyncratic. There was no comparison group. The description of methods and procedures was meager. The time interval between pre- and posttesting was considerable.

Coplin, J. (1994). *Weck Educational Development Program. Final performance report.* Durham, NC: Weck & Durham Community College.

Purpose: This final report described a workplace literacy program conducted at Weck, Inc. for employees of its manufacturing and warehouse divisions. The job-related curriculum was developed from a task analysis.

Variables: Variables studied were GED completion and learner satisfaction.

Population: There was a sample of 81. Twenty-eight participants lacked high school diplomas.

Design: The descriptive design was post only.

Data Collection: Students were asked about their satisfaction levels. Informal interviews were conducted with company staff. Surveys and written comments were analyzed.

Instrumentation: A 19-item satisfaction survey was used.

Findings: Of the original 36 participants who started in 1991, 11 received their GEDs. Interviews with managers and program staff indicated that the program received support. Students' written comments suggested that satisfaction was high.

Strengths: None.

Weaknesses: There was no formal methodology or procedures. This study measured students' satisfaction only. Results of the survey were not reported.

Coston, C., & Dagard, C. (1995). *Champaign County National Workplace Literacy Project. Final evaluation report.* Urbana, IL: Champaign Board of Education.

Purpose: The Champaign County Workplace Literacy Project was managed through the Champaign Board of Education. Four local companies, major employers in the county, participated in the project. The project had two overall purposes: (a) to provide training to employees based on their needs, and (b) to design and implement a curriculum that targeted job-specific skills to prepare employees for participating in TQM/SPC processes. The project implemented three cycles of classes, each lasting 15 weeks.

Variables: Variables measured were prose literacy (understanding and using information contained in various textual materials) and document literacy (processing information found in documents), mathematics achievement, knowledge and ability to participate in TQM/SPC processes, and the effects of increased worker productivity.

Population: All participants were employees of four local companies. They were selected from various job classes, including maintenance, technical, manufacturing, and administration. Two hundred and fifty-five employees participated in the classes. Standardized pre- and posttest prose achievement data were collected from 161 individuals, document achievement data were collected from 171, and math achievement data from 166.

Design: An external evaluator used the Triphase Evaluation process based on Stuffelbean's decision-making model. The three phases evaluated were input, process, and outcome. The analysis involved a comparison of matched pairs of the pre- and posttest results and a percentage analysis of the observed change during treatment.

Data Collection: Project staff collected data from most sites. The project evaluator analyzed the statistics. No posttest data were collected from two sites. The use of the ABLE was discontinued after the first cycle. Problems arose in collecting interview data from some sites and during some cycles.

Instrumentation: Instruments included participant and supervisor surveys, supervisor interview forms, and two standardized tests (TALS and ABLE).

Findings: Data were inconclusive on the extent to which this project impacted workers' levels of productivity and their understanding and use of SPC and TQM. The aggregate result of prose literacy for all participants demonstrated a statistically significant growth during all cycles, although achievement in document literacy showed no significant growth. In fact, participants experienced a decrease in the average scores on the latter. Most participants experienced statistically significant increases in their math scores.

Strengths: Evaluation methods were comprehensive.

Weaknesses: There was no control group. Data collection procedures were not rigorous. Much of the data were incomplete. The standardized test did not seem to measure what was taught.

Cox, L., & Lane, C. (1981). *Evaluation of the adaptation of the Personalized System of Instruction to nontraditional adult learners. Final report. Volume III: PSI meets ABE.* Durham, NC: Research Triangle Institute. (ERIC Document Reproduction Service No. ED 208 181)

Purpose: The Personalized System of Instruction (PSI) was a self-paced, mastery-oriented system that stressed the use of printed instructional material, immediate feedback, and peer proctors. This method had been successful in teaching college students. The study sought to gauge the effectiveness of PSI on adults in settings other than colleges and universities. The main objective was to answer the question, "Is further investigation of the use of PSI for nontraditional adult learners likely to be worthwhile?" The study was also designed to identify salient areas for further investigation. Some outcome data were generated.

Variables: Outcome variables were reading and mathematics achievement; changes in self-concept, self-dependence, and self-assertiveness; and evidence of increase in life-coping skills.

Population: All participants were ABE students. They were enrolled in classes meeting twice a week for two- or three-hour sessions. One class was situated in a housing authority site, one was at a community school setting, and one in a prison. Most of the students were Black and approximately 90% were female, except for the prison inmates who were all males. Students ranged in age from 18 to 50, but most were 30 to 50 years old.

Design: The research design was a case-study approach. Although the primary emphasis was on collecting observational data from three sites during a six-month implementation, the researchers also collected impact data. They collected data from two other sites for comparison.

Data Collection: To insure baseline data, students in the three research sites and two comparison sites were given the ABLE and self-concept scale within the first four weeks of class. Classroom teachers administered the instruments. Pre- and posttest ABLE data were recorded for ten PSI students and five from the comparison groups. Twelve PSI students and nine comparison group students had pre- and posttest scores on the self-concept scale.

Classroom observations were to be conducted primarily by the classroom teacher, while researchers planned to visit each PSI site once a week. Observational data were summarized by site. Data collected were the numbers of units mastered, pre- and posttest achievement and self- concept scores, and attendance records.

Problems occurred with the interview process. Interviewers spent a lot of time "tracking down" interviewees. There were 319 incomplete interviews: 70% were due to the inability to locate interviewee, 11% were unwilling to participate, and 7% provided very little useful information due to physical, mental, or emotional disabilities.

Instrumentation: Researchers developed several forms to collect observational data. The ABLE was used to collect reading and math achievement data. Self-esteem data were obtained by the Tennessee Self-Concept Scale.

Findings: The PSI students showed average gains of almost one grade level in reading and more than one grade level in math. No such gains were found in the comparison groups. However, the latter group's entry-level scores were much lower than the PSI groups; thus, they did not represent the same population. Self-concept pretest scores were almost identical for the two groups. PSI students showed a slight increase at posttest while the comparison groups showed a slight decrease.

Strengths: There was a comparison group.

Weaknesses: No real conclusions could be reached regarding PSI effect on self-concept because the sample size was so small. The comparison group had a much smaller size than the treatment group and was not representative of the treatment group in all variables measured. The research design and data collection were weak.

Darkenwald, G. (1983). *Project F.I.S.T.: Functional In-Service Training. Submission prepared for the Joint Dissemination Review Panel. Edison, NJ: Middlesex County College.*

Purpose: Project F.I.S.T. was a volunteer-based administrative and instructional delivery system aimed at meeting the special needs of low-level readers (0-4). F.I.S.T. evaluated its program from 1981 through 1983 to apply for validation by the Joint Dissemination Review Panel (JDRP). It measured tutor and student outcomes. Only the latter is discussed in this abstract.

Variables: Student outcomes measured were reading achievement and change in reading behaviors.

Population: The sample size for both the initial and replication group was 26. All the ABE students lived in New Jersey. The mean age for both groups was 32.5; mean grade completed was 9.2. Fifty-three percent were women. Half were from minority backgrounds (nearly all Black) and 68% were employed.

Design: A single group pre-post design with replication was employed.

Data Collection: Data were collected by project staff under the supervision of an external evaluator who also performed the statistical analyses. The standardized reading test was administered at entry and at four-month intervals.

Instrumentation: The Woodcock Reading Mastery Test was used to measure reading gains. A questionnaire was administered in both studies to gather data on the application of reading skills and other positive life changes due to participation in F.I.S.T.

Findings: In the original study, both the initial and replication groups showed statistically significant gains on the Woodcock. The gains for the replication group were similar to those of another JDRP-approved program, the Jefferson County Adult Reading Project (JCARP). The initial group performed considerably better than the replication and JCARP groups.

All but one of the learners indicated positive life changes and general improvement in reading skills. All 21 items on the questionnaire showed gains. Especially significant was that one fifth of those looking for jobs became employed, nearly three fifths started reading newspapers for the first time, and about a third of the parents began helping their children with schoolwork.

Strengths: The replication design was rigorous. Methodology was well described. Results were compared to another nationally validated program.

Weaknesses: There were low numbers and no control group. The data collection methods for the questionnaire were not well described and it was unclear how it was developed or by whom. Some data were self-reported.

Darling, S. (1982). *Jefferson County Adult Reading Program. Submission prepared for the Joint Dissemination Review Panel*. Louisville, KY: Jefferson County Adult Reading Program, Jefferson County Board of Education.

Purpose: The Jefferson County Adult Reading Program (JCARP) was initiated in 1978 to develop and field test a model adult literacy program. Field testing in 1978-79 provided information to help redefine the model for the 1979-80 funding year. Two years of field testing showed that teachers' certification had little impact on students' progress in reading; materials should be of high interest to adults but designed for those reading below a sixth-grade level; and student reading gains were related to increased hours of instruction. Essential to the subsequent model was the idea that "student-staff interaction would result in consistent attendance and high retention." As a consequence, JCARP integrated a counseling construct into most phases of its program, including student

recruitment, staff training, instructional design, and evaluation. This study was submitted to the Joint Dissemination Review Panel and described data from the field test years as well as the 1980-81 model program.

Variables: The outcome variables were reading gains and attrition rates. Hours of instruction completed were also studied, along with the achievement of student goals (such as employment and education) in 1980-81.

Population: The initial field-test year (1978-79) included 293 students who tested below 4.0 grade equivalent on the SORT. The following year studied 508 students at 26 locations reading up to a 6.0 grade equivalent. During the 1980-81 model year, 591 students enrolled. Of the 1,392 students enrolled, 54% were male and 46% were female; 59.8% were Black and 36% White. Other groups enrolled were categorized as Hispanic, Asian or Pacific Islander, American Indian, or Other. Thirty-nine percent of the students were employed. The median age was 27. Students were considered enrolled once they took a pretest. Separation from a program occurred after one month of nonattendance or when students indicated their withdrawal. Demographic distributions of the JCARP students were similar to other ABE students from Kentucky.

JCARP's attrition rates in 1980-81 were compared with comparable programs in Kentucky. Five hundred and ninety-one learners enrolled in JCARP that year and 706 in 10 comparable programs.

Design: JCARP used an experimental design.

Data Collection: Demographic and attrition data were collected from all 10 ABE centers. Data to compare reading gains were collected from five programs located in urban environments comparable to Jefferson County.

Instrumentation: The TABE (levels E and M) was used to measure reading gains during the comparison year. TABE raw scores were converted to grade equivalent scores. JCARP had a 6.0 grade-equivalent level as a maximum for enrollment, so the SORT was used as a screening instrument. Comparison groups had no such guidelines and screening tests were not used.

During the two field-test years, JCARP used different standardized reading tests to measure reading gains. In 1978-79, the Woodcock Reading Mastery Test was utilized and in 1979-80, the Adult Basic Learning Examination (ABLE).

Findings: At 22%, JCARP participants experienced a lower attrition rate than comparable programs. Among the comparison groups, the highest attrition rate was 80% and the lowest was 52%. A chi-square analysis was significant at the .001 probability level.

JCARP students made significant gains in reading ability. The average gain was 1.53. A *t* test of these results was significant at the .001 probability level. The mean hours of instruction was 82.

The gains of the JCARP participants were compared with those from five other learning centers (173 matched subjects). JCARP students showed a .70 greater equivalent gain than did comparison group participants.

A *t* test on hours of instruction was not statistically significant.

Two hundred and eighty-five students had employment-related goals; 139 achieved them. Eighty-eight JCARP participants of the 325 who listed high school equivalency as a goal made sufficient progress to enroll in GED programs. Furthermore, 62 students enrolled in other education or training programs and 175 students said they planned to continue their education.

Strengths: The study's experimental design included a comparison group. The design was rigorous so it ensured statistical reliability and generalizability. This research included data from a two-year pilot study.

Weaknesses: Although common practice in Kentucky, TABE standardized testing procedures were violated because no time limitations were imposed. Furthermore, comparison data were used only during 1980-81. Because teachers tested their own students in the JCARP program and the comparison programs, tester bias could be a factor making the findings less reliable.

Darling, S., & Hayes, A. (1989). *Breaking the cycle of illiteracy: The Kenan Family Literacy Program model. Final report 1988-89*. Louisville, KY: National Center for Family Literacy.

Purpose: This is the 1988-89 final report of the Kenan Trust Family Literacy Project implemented by the National Center for Family Literacy (NCFL). Seven pilot sites--three in Kentucky and four in North Carolina--implemented the Kenan model. It required basic skills instruction for children and adults in the program. The major goal of the project was to "break the cycle of illiteracy" by integrating early childhood education with basic education and instruction in parenting skills for the parents.

Variables: NCFL and its researchers developed summary profiles of types of adults who enrolled in their programs. The following variables described the differences among groups: (a) degree of acceptance of social norms, (b) commitment to change, (c) hopefulness of change in self and personal conditions, (d) confidence of change, (e) personal capability to learn and change, and (f) supportiveness of environment.

Achievement variables were improvement of parenting skills, increases in home literacy, math and reading achievement, GED attainment, and pursuit of employment or further education.

Population: The sample size for the pilot sites was 114 adults and 288 children. The average age of the parent was 26.6 years. The families averaged 2.7 children.

Design: Data were descriptive with a pre-post component. No systematic research design was evident.

Data Collection: The sites used various procedures and instrumentation to collect data. The report was not clear in describing which procedures were implemented nor when and how they were used.

Instrumentation: Instruments included tests developed by teachers, case studies of families, anecdotal and descriptive records, the TABE, interviews, surveys, and observations.

Findings: The following “types” of parents received GED certificates: Type A--3, Type B--4, Type C--1, Type D--1, Type E--2, Type F--2, and Type G--2. Fifty-nine percent gained over one grade level and 29% gained over two grade levels on the TABE Reading test. Seventy-one percent gained one grade level, 48% gained over two grade levels, and 24% gained three grade levels on the TABE Math test. In addition, 80% gained over one grade level, 45% gained over two grade levels, and 45% gained over three grade levels on the TABE Language test.

Strengths: Comprehensive variables were measured.

Weaknesses: There was no comparison group. There were no details on methodology. The report failed to provide details on context. Except for tested learning gain, most data appeared to be self-reported.

Darling, S., & Hayes, A. (1996). *The power of family literacy*. Louisville, KY: National Center for Family Literacy.

Purpose: Beginning with the 1991-92 school year and continuing through 1994-95, NCFL implemented the Toyota Families for Learning Program and the Knight Family Education Program at 80 sites in 19 cities. This report summarized the research findings during those funding years as well as the outcomes of follow-up studies on families from one to six years after enrollment.

Variables: The outcome variables measured for the adults were retention rates, language and math gains, GED attainment, self-esteem, confidence in parenting, and amount of home literacy activities. The follow-up studies addressed obtaining and keeping employment, continued enrollment in education or training programs, reduction of dependence on public assistance, and parental involvement in their children's school.

Population: Almost 2,500 families participated during the four funding years. It was not clear how many adults participated. The follow-up studies involved 200 adults and 315 children who had attended the original Kenan model sites, the Hawaii Pilot Family Literacy program sites, or those sponsored by NCFL in Florida, New York, and Arizona. Adults in the NCFL programs received a median of 140 hours of instruction over a 24-week period.

Design: The data were primarily descriptive. A systematic research design was not evident.

Data Collection: Follow-up data were collected by telephone or face-to-face interviews. Other procedures were not described.

Instrumentation: Academic achievement instrumentation included CASAS and the TABE. The Rosenberg Self-Esteem Questionnaire measured general self-esteem. It was not clear how the other variables were measured.

Findings: In the ongoing studies, adults at the lowest levels of literacy at enrollment gained an average of 4.4 CASAS scale points. Twenty-eight percent made gains of one or more CASAS levels. The average gain on the TABE was 1.3 grade levels. Descriptive data also indicated that parenting skills, parent-child activities, and adult self-esteem improved.

Of the 200 parents included in the follow-up studies, 51% attained the GED, 43% were employed (compared with 14% before enrollment), 13% enrolled in higher education or training programs, 11% were continuing with ABE, and 23% of those on public assistance at enrollment had ceased to collect benefits.

Strengths: Comprehensive variables were measured. Much of the data were compared to literacy programs that focused solely on adults.

Weaknesses: There was no control group. The report failed to provide details on context or methodology. Other than tested learning gain, most data appeared to be self-reported.

Educon Diversified, Inc. (1995). *Investing in Minnesota's future: An analysis of economic, societal, and personal impacts of literacy and adult basic education programs in Minnesota.* Belle Plain, MN: Author.

Purpose: The purpose of this report was to inform the governor, the Minnesota legislature, and state taxpayers on the outcomes of adult basic education.

Variables: The variables primarily concerned data on economic impact.

Population: This report wove an argument for impact using data from a variety of sources. The Minnesota Department of Education supplied most of the statistics (presumably the same data reported to OVAE each year). The sample size varied with the statistic presented.

Design: This purely descriptive report used data from the state, the popular press, NALS, and other sources to make its case. Some anecdotal testimonials were included.

Data Collection: These procedures were not described.

Instrumentation: The instrumentation was not described.

Findings: Between 1960 and 1994, 147,000 persons received their GED. Most of the data were used to demonstrate the need for adult literacy education rather than to portray outcomes.

Strengths: The report presented a wealth of useful descriptive information.

Weaknesses: Although the report claimed to be an outcome assessment, it was actually more of a public relations document. No data on outcomes were actually collected; all sources were secondary. Most of the statistical data were presented to argue need rather than to demonstrate impact. The methods and procedures used to collect the data were not described.

Fingeret, H., & Danin, S. (1991). *They really put a hurtin' on my brain: Learning in Literacy Volunteers of New York City*. Durham, NC: Literacy South. (ERIC Document Reproduction Service No. ED 332 026)

Purpose: This research study evaluated an educational program, Literacy Volunteers of New York City (LVNYC), where tutors provided small group instruction to adults with low literacy skills. The major question addressed was the impact of Literacy Volunteers of New York City on its students.

Variables: Change was examined in three areas: (a) literacy skills; (b) self-concept, attitudes, and beliefs related to literacy development; and (c) involvement in literacy tasks outside the program. A quantitative piece analyzed students' standardized test scores and writing samples. Writing variables included sub-skills such as the number of words used in general and per sentence, content organization, vocabulary, sentence and paragraph structure, and mechanics.

Population: Twenty tutors and 16 LVNYC staff members were interviewed. Sixty-five students were interviewed in focus groups and another 37 by telephone. Ten of the focus group students were interviewed individually several times as case studies. The reading scores of 114 students and 66 writing samples from 33 students were analyzed.

Design: This was a qualitative case study, although test data were analyzed at 50-hour intervals.

Data Collection: Fieldwork was conducted and data collected over seven months in 1990. Open-ended focus group interviews, individual interviews, and observations were the primary methods used to collect data. Coding the qualitative data involved using a computer text-analysis program specifically developed for the project. The data to analyze students' writing samples were collected at various points. The first writing sample was collected during the first month of participation. Standardized reading test data were collected at 50-hour intervals.

Instrumentation: A consultant generated a rubric modeled on various writing assessment tools to analyze the writing samples. A reading vocabulary inventory was given as a locator when students enrolled in LVNYC. Those who scored below 3.0 were administered the Test of Adult Basic Education (TABE). Those who scored below 3.0 were given the Adult Basic Learning Examination (ABLE).

Findings: Writing Samples: The relationship between cumulative length of time in the program and change in writing scores was examined. Thirty-three paired writing samples representing three groups were analyzed. The mean length of time in the program for each group was 5.38 months, 9.43 months, and 16.67 months, respectively. The difference score average for each group was 9.23, 6.75, and 7.5, respectively. Although the researchers concluded that rapid gains in writing scores occurred in the initial stages of instruction and then reached a plateau, they were not sure what these findings meant. They hypothesized that slower writing gains for students over time may signify changes in content from personal stories to varied subject matter. No claims were made about the statistical significance of these findings.

Reading Gains: The researchers hypothesized that the more instructional hours a student received, the greater the gain in reading grade level. Average grade-level score at entry for 114 students was 1.518. Average posttest score was 2.4. Average grade-level gain was .9 years. Average length of time in the program was 229 hours. The gains were not found to be statistically significant. The researchers noted that “the results of the Analysis of Covariance (ANCOVA) reveal no statistically significant effects of differences in grade level gains between the pre and posttest scores based on length of time in the programs, grade level at entry into the program, or the relationship between the length of time in the program and entry grade level score.”

To assess qualitative impact of the program on LVNYC’s students, the researchers presented an impact framework that viewed literacy as “practices” (i.e., something that one participates in doing rather than being a characteristic of the person). They concluded that literacy is “situational” and distinguished between private and public literacy practices. Public and private literacy practices occurred inside and outside LVNYC.

LVNYC students assessed their reading and writing skill development by offering examples of what they could do during their tutoring sessions. They also assessed their progress by noting their progression through books of increasing difficulty, and gauging how their reading speed and comprehension had improved. The positive impact of LVNYC on some of its students was seen in the high level of involvement they had in extracurricular activities offered by the program, including participation in tutor and student orientation and training, state and national conferences, a publishing program, and student committees.

The impact of new “private” literacy practices outside the program was most evident in activities that had a high level of predictability (e.g., paying bills).

Interpersonal relationships also changed, such as students doing tasks independently or involving new people to support their reading and writing development. Four environments--jobs, banks, public transportation, and grocery stores--were consistently mentioned by students as places where they engaged in public literacy practices in new ways. The researchers also found that students' beliefs and attitudes toward reading and writing tasks changed as did their self-image as learners.

Strengths: A rigorous design included qualitative and quantitative data. The framework generated by the qualitative data--that literacy is practiced situationally--is important to the field.

Weaknesses: The small sample size for some of the variables examined make it difficult to conclude that the results were representative of the program as a whole. There was no control group.

Hayes, A. (1993). *Hawaii Family Literacy Pilot Project: Summary report 1990-93 and interim follow-up*. Louisville, KY: National Center for Family Literacy.

Purpose: In 1989, three sites in Hawaii adopted and adapted the Kenan Family Literacy Program model. The report evaluated the project after a three-year period. This abstract will describe only the impact upon adults.

Variables: Measurement of academic, social, and personal characteristics of adults was planned, but very little pre- or posttest data were obtained.

Population: One hundred adults initially enrolled in one of the three sites. Seventy-three remained until the end of the first year.

Design: The design was descriptive.

Data Collection: Two sites implemented the Hawaii Competency-Based High School Program and the third a GED-enrichment program. Little description of collection procedures was included.

Findings: Thirty four of the 41 parents who did not have a diploma at enrollment received GED certification. Of the adults who had pre- and posttest achievement data (the numbers were not stated), all improved approximately one grade level during the time enrolled. At one site, 29 or 30 students attended a special

summer/transition program at a local community college. There were few changes in employment.

Strengths: There were no real strengths.

Weaknesses: There was no control or comparison group. There were few details on methodology. The data on adults were self-reported.

Hudson, P., & Gretes, J. (1994). *New Horizons: A national workplace literacy program. Final report and external evaluation impact study.* Conway, SC: Horry-Georgetown Technical College.

Purpose: New Horizons was a workplace literacy project in which employees of Georgetown Steel Corporation attended classes provided by Horry-Georgetown Technical College. This project had eight objectives. Among them were to (a) conduct a comprehensive needs assessment of employees and develop competencies bases on those needs, (b) design course curriculum around those competencies, and (c) develop an individual education plan (IEP) for each employee who participated in workplace literacy classes.

Variables: The variables were reading, writing, and math gains; self-confidence as readers, writers, and in mathematics; self-esteem; and on-the-job problem solving, time management, and productivity.

Population: Fifty-three percent ($N = 454$) of Georgetown Steel employees participated in at least one of the classes. Their mean age was 40 years old. Most participants were male. Two hundred and ninety-four employees were White and 159 were Black. The majority of participants had been employed by the company for more than six years. However, only 57 participants had pre- and posttests in math, 61 in reading, and 19 in writing. It is not clear how many completed the surveys to gather data for the other variables.

Design: There was a pre-post comparison design.

Data Collection: Besides pre- and posttesting, interviews were conducted with participants and their supervisors. Researchers also analyzed IEPs. Supervisors were interviewed each quarter using information from the survey.

Instrumentation: The Bader Reading Inventory was used to place those who read above 4.0 in reading and writing classes. Those who scored lower were placed in

an emergent reading class. Customized tests were used to measure reading, writing, and math gains.

A project-developed survey was used to gather information from supervisors. It consisted of open-ended questions. A 12-item participant survey was also created, with a five-point scale ranging from “strongly agree” to “strongly disagree.”

Findings: Participants had a 39% increase in math, a 22% increase in reading, and a 15% increase in writing. In addition, 91% of the participants “strongly agreed” or “agreed” that their communication skills had increased, 77% believed the program had increased their productivity, and 93% felt their self-esteem had improved.

Strengths: Many variables were measured.

Weaknesses: The sample size seemed small, although the true sample size was not made clear. There was no control group. Much of the data were self-reported. Data collection methods were not well described.

Koen, S. & Musumeci, M. (1984). *BES Adult Literacy Project*. Norwalk, CT: MATRICES Consulting Group.

Purpose: The BES Adult Literacy Project was the classroom instruction component of Bronx Educational Services. This research study was its submission to the Joint Dissemination Review Panel and described student outcome data from 1981 to 1984.

Variables: The outcome variables were reading gains and attrition rates.

Population: There were seven cohorts of students. All adults were in the nonreader or beginner reader category. Ninety students had matched pre- and posttest scores, which represented 52% of the entire BES enrollment during the three years of program evaluation. The majority of the students (81%) were Black; 16% were Hispanic. Fifty-nine percent were males. Three quarters of the students were employed; the majority earned less than \$12,000 a year.

Design: A quasi-experimental research design, the Cohort Design, was used to assess program impact. It combined longitudinal and cross-sectional approaches into one research plan. *T* tests were computed for the entire student sample and for individual cohorts. An independent *t* test was also used to determine the

difference in reading achievement between students exposed to the BES program and those who were not.

Data Collection: Students were tested at the beginning and the end of each instructional cycle (100 hours). TABE tests were administered by BES staff. The project evaluator spot-checked scoring for accuracy.

Instrumentation: The TABE (levels E and M) was used to measure reading gains during the comparison year. TABE raw scores were converted to scaled scores and grade equivalent scores. Scores were transferred to data rosters provided to external evaluators.

Findings: Reading gains on the TABE were statistically significant for all seven cohorts. Cross-sectional comparisons demonstrated on all TABE reading scales that BES students outperformed comparable groups of students not exposed to the program.

Strengths: The study used a quasi-experimental design which included a control group. The design was rigorous. Reading gains were measured over time.

Weaknesses: Very few variables were measured. The uniqueness of the program makes it difficult to duplicate.

Lafayette School Corporation. (1991). *Lafayette Adult Reading Academy (LARA) and St. Elizabeth Hospital Employee Literacy Program. Final program report.* Lafayette, IN: Author & St. Elizabeth Hospital.

Purpose: This literacy project sought to increase the usefulness of health-care materials by reducing their volume and to improve the basic skills of hospital employees.

Variables: Outcome variables measured were literacy, verbal communication, and problem solving around particular work tasks or strands, such as terminal cleaning, safety instruction, and disaster codes.

Population: Two hundred and thirty-six participants, primarily white females employed at St. Elizabeth Hospital, participated in the project. The majority had been employed at the hospital for five years or less. They were drawn from various departments.

Design: The report was primarily descriptive. This was not a research study, but a literacy project in which nine training strands were developed.

Data Collection: Employee reactions were elicited through surveys, interviews, and testimonials. Achievement gains were measured by pre- and posttests of job-specific skills. Change in participant behavior on the job was measured by observation, checklists, and rating scales. After a comprehensive needs analysis, new types of health-care forms were created to measure progress. It was not clear who developed them or how.

Instrumentation: All instruments were generated by the project.

Findings: All participants tested higher on targeted skills.

Strengths: Variables measured were job specific. Useful forms were developed for the hospital.

Weaknesses: There was no comparison group. Data were exclusively self-reported. It was not clear how variables were measured. Instrumentation was developed entirely by the project and not well described.

Lashof, J. (1992). *BEST: A learner-centered workplace literacy partnership of the Vermont Institute for Reliance and General Electric Aircraft Engines*. Rutland, VT: Vermont Institute for Self-Reliance.

Purpose: From April 1990 to March 1992, the Vermont Institute for Self-Reliance conducted context-based literacy instruction for employees of two General Electric Aircraft Engine plants and two Burlington Electric Department plants. Responsive Text (RT), a computer-augmented reading environment, was one of the primary modes of instruction. One of the project's goals was to evaluate the extent to which RT provided an effective tool for developing employees' literacy skills. The project also sought to increase the literacy levels of employees and evaluate the extent to which these increases improved their work performance.

Variables: Instructors' perceptions of improvement in employees' literacy skills were measured. The variables included comprehension of expository, work, and narrative materials; decoding ability; vocabulary knowledge; interest in reading and writing; participation in group activities; ability to use and interest in RT; and writing ability.

Supervisors' perceptions of improvement in employees' literacy skills were measured. Variables included changes in the employee's ability to read and write job-related materials, and the frequency he/she reads and writes at work. Participants' perceptions of these categories were also measured.

Participants' perceptions of changes in their work performance were measured. The variables included leadership in the workplace, initiative in the workplace, knowledge of her/his job, self-confidence in the workplace, absenteeism, quality of work, and level of responsibility. Supervisors' perceptions of these categories were measured.

Population: It is not clear how many workers actually participated, because the numbers reported are a duplicated count. In other words, participants were counted more than once.

Design: This was an evaluative report. There were pre-post comparison components and a portfolio assessment.

Data Collection: Qualitative and quantitative data were collected. Multiple sources of data were used to triangulate consistent patterns of changes.

Instrumentation: Portfolio assessment procedures were used to measure outcomes. The instruments included pre- and post-interview forms, reading assessments, writing samples, a scale of teachers' perceptions of improvement, a participant self-evaluation form, a supervisor's response form, and a responsive text evaluation form. All instruments were developed by the project. They drew on the holistic assessment approach developed by Susan Lytle and incorporated a portion of the California Adult Learner's Progress Evaluation Program.

Findings: Participants increased in each of the four measures of literacy performance. In all but two cases, the mean differences were statistically significant. Instructors and supervisors perceived participants as making gains in the four areas of literacy performance as did the participants themselves. Participants and supervisors also perceived major changes in work performance.

Strengths: Assessment instruments were very comprehensive. Multiple sources of data were used to triangulate consistent patterns of changes.

Weaknesses: Much of the data were perceptions reported by instructors and supervisors. There was no control group. The number of employees served is reported as a duplicated count. It is, therefore, not clear what the actual n is.

Lipiec, J. (1993). *Verifying applied literacy skills (VALS) in ABE Programs*. Schnecksville, PA: Lehigh County Community College.

Purpose: This study compared the effectiveness of a traditional GED curriculum with one based on applied literacy skills. It also sought to determine the validity of the Educational Testing Service's Tests of Applied Literacy Skills (TALS) as a predictor for GED performance.

Variables: Learner outcome variables were GED attainment and performance, reading and math achievement, retention, and meeting stated goals.

Population: Originally 70 students enrolled in one of the two classes, but only 14 students from each class finished the program and had complete pre- and posttest data. Thirty-five students were reached by telephone and interviewed three months after completing the program.

Design: This study had an experimental design with a control group. Pre- and posttests were compared using a *t* test on mean scores.

Instrumentation: The TALS and ABLE were used for achievement data. Results of GED practice tests were also analyzed. A seven-question follow-up survey was used.

Data Collection: Pre- and posttests were administered by project staff at the beginning and conclusion of the classes. GED practice tests were administered during class. Follow-up interviews were conducted with learners three months after they had completed the program. Retention information was collected by attendance sheets signed by students at each session. The research staff of the college conducted the statistical analyses.

Findings: Approximately 40% of the students from each group completed the program. This result did not support the original hypothesis regarding retention. In fact, the control group students averaged 1.5 more sessions than the experimental group. The average age of those who completed the GED program was 39 years, compared to 27 years for those who dropped out.

Both groups showed significant gains on the TALS and ABLE. However, the experimental group gain was more significant in traditional basic skills while the control group's gain was more significant in applied basic skills. Again, this result countered one of the original hypothesis of the study. When the data from the two

groups were compared, no significant score differences appeared on the ABLE, TALS, or GED practice tests.

The follow-up interviews revealed that the amount of progress made by experimental and control groups in meeting personal goals was essentially the same.

Strengths: There was a control group. Appropriate variables were measured. Follow up was conducted. The researchers asked questions which are important to the field. Local data were compared to national data.

Weaknesses: There were small sample sizes in each group and substantial attrition. The data collection methods could have been better described.

Lipscomb, J., & Cronin, C. (1994). *Tools for the 21st century. Southwest Alabama Cooperative Literacy Project. Final evaluation. Mobile, AL: University of South Alabama.*

Purpose: Seven manufacturing plants and the University of South Alabama developed a workplace literacy program, the Southwest Alabama Cooperative Literacy Project. The primary objective was to increase productivity by increasing the literacy skills of employees. The project lasted 18 months in 1993-94.

Variables: Reading and math gains were the two variables studied.

Population: Twenty-six employees participated in one or more sessions of reading classes and 35 in one or more of the math classes. The majority were males between the ages of 35 and 60. They held diverse positions at their manufacturing plants. All scored below seventh grade on the ABLE.

Design: Data were primarily descriptive. There was a pre-post component.

Data Collection: Data collection methods were not adequately described.

Instrumentation: The ABLE was used to measure reading and math achievement.

Findings: There were 20 pre- and posttests. The average gain was 4.2 points on the ABLE, Form C.

Strengths: There were no strengths.

Weaknesses: Sample size was small. It lacked a control group. Data were primarily anecdotal. No correlation was made between literacy gains and job performance. This was not a formal study.

Mikulecky, L., & Lloyd, P. (1993). *The impact of workplace literacy programs: A new model for evaluating the impact of workplace literacy programs.* Bloomington, IN: Indiana University, School of Education.

Purpose: Because many evaluations of workplace literacy programs have been superficial, NCAL funded a project in 1990 to develop and pilot a model for evaluating the impact of these programs. This study produced data on the impact of programs at two sites. An ancillary goal was to refine the model for use at other sites.

Population: Participants were employees enrolled in workplace literacy programs at two companies with very distinct work sites. Each company had education programs which were recognized as effective by state and federal agencies.

Delco Chassis, Rochester, NY

Delco Chassis was a unionized electrical motor manufacturing plant with over 3,600 employees. Education programs were operated jointly by union and management with support from state and regional agencies. All learners in this study were enrolled in three different types of programs. (a) The technical preparation class had 14 participants. This class was a six-week, full-time course designed to prepare them for further training. A control group of 12-15 employees had enrolled but not yet begun the class. (b) The GED preparation class had 15 participants. No control group was described. (c) The ESL class had 15 participants. A control sample contained five.

Cumberland Hardwoods, Sparta, TN

Cumberland Hardwoods was a nonunion rural plant. Because the company had a successful education program for several years, many of the employees had already passed through the available training courses prior to the study. Consequently, the samples at this site were quite small. For example, the GED course had so few students that insufficient data were obtained. Furthermore, no control groups could be formed. The population sample studied ($N = 21$) were employees taking part in a course, Communication and Collaboration, designed to help them work cooperatively as self-directed teams. Participants were members of two such teams.

Variables: Learner literacy variables included changes in beliefs (e.g., self esteem); changes in practices (e.g., reading and writing at home and work); changes in reported reading process and reading comprehension; and changes in plans (e.g., educational). Other measures were employer objectives such as employee attendance and ratings of supervisors. The researchers attempted to measure some family literacy variables, but the numbers were too small for statistical analysis.

Design: Originally this study planned a “complete” experimental design, but small sample sizes precluded that possibility. As a result, some of the data were analyzed using statistical comparisons of quantitative information. Other quantitative, as well as some qualitative, data were analyzed by categories that emerged from interviews and open-ended questions.

Data Collection: Interview, test, and classroom questionnaire data were collected before and after each course or at appropriate intervals during the class semester. Interviews were conducted by the researcher on a one-to-one basis. The classroom teachers administered the questionnaire and cloze test during the class period. Research personnel and on-site coordinators conducted class observations. Notations of class activities were documented on a timed basis.

Instrumentation: The researchers used a combination of approaches to develop instruments, including literature reviews, surveys of indices of productivity at work, and discussions with plant supervisors to develop anchored rating scales. Susan Lytle’s conceptual framework for changes in adult literacy was used to develop interview and questionnaire items. These items were designed to reflect local use. Interview protocol forms, questionnaires, cloze tests, and classroom observations sheets were developed. Learners’ reading abilities and literacy practices at work and outside were measured by a questionnaire that used Likert scales, fill-in-the-blank items, and anchored rating scales. Cloze tests were used to gauge workplace reading comprehension. The family literacy interview protocol consisted of open-ended questions. An ESL competency checklist was also devised. Supervisor productivity rating forms were sometimes left with the supervisors and at other times completed with a researcher.

Findings: Major results on the impact of employee participation in workplace literacy programs, excluding findings from the ESL class, are summarized below. Changes in variables were judged statistically significant when compared to the control group.

Changes in beliefs: The technical preparation group showed a statistically significant improvement from pre- to post-interview when asked, "How literate do you consider yourself to be?" Self-rating and holistic scores were statistically significant at the $p < .02$ and $p < .01$ levels. The written questionnaire also contained a question about literacy beliefs. The latter data did not depict significant results, suggesting an advantage of the interview technique over the questionnaire method. Participants did not significantly change their views about what constitutes a literate person.

Changes in practices: The Cumberland samples showed a significant gain in the task of "asking questions at work." Reading and writing practices away from work were significant for the technical preparation group. There was also a significant gain in the range of reading for GED students.

Except for the technical preparation group (because it was a full-time class and participants did not work while attending), all groups showed pre-post gains in work-related reading and writing. Participant responses were measured by a count of items measured and holistically--the frequency, breadth, and difficulty of reading. However, learner self-ratings showed very little gains. Nonetheless, they showed significant increases ($p < .05$) on two items, talking in meetings and having one's ideas discussed in meetings. The Cumberland Communication and Collaboration class also showed gains in working cooperatively with others. Changes in literacy activities away from work were also measured. The technical preparation class showed statistically significant differences and the GED class showed gains. The GED class also showed a significantly wider range of reading.

Changes in process and ability: Changes in reading process and reading comprehension were measured by interview questions and cloze tests constructed from plant reading materials. The Delco technical preparation class made statistically significant gains on the cloze test. Others did not, perhaps because the reading passages were too difficult. Significant gains were demonstrated on some of the interview items.

When participants were asked to read a passage and describe what processes they used to read it, the posttest answers were deemed statistically significant for the Delco GED groups ($p < .005$). Their answers covered reading strategies (e.g., skimming) and topics of interest (e.g., products manufactured by competitors).

Changes in plans: When asked about their plans for the future in one, five, and ten years and to relate reading to those plans, the technical preparation students

showed significant improvements for one- and five-year plans ($p < .02$ and $p < .05$).

The other items measured were either not outcome variables or could not be accurately measured due to small numbers. Supervisor ratings at the Cumberland plant showed statistically significant gains.

Strengths: Some population samples had control groups. The study generated quantitative and qualitative data. The design and measurement instruments were solid. The design included various modes to gather data. Comprehensive variables were measured in several ways. The study evaluated various ways to collect data from workplace literacy programs.

Weaknesses: Some of the results were self-reported. Not all population samples had control groups. The sample sizes were small. Only data from two workplace literacy programs were included. It is questionable whether the Cumberland site should have been used at all, given that most employees had already passed through the program. It is difficult to understand how data gathered from just these two sites would effectively help develop and refine a research model for use at other sites.

Mott Community College. (1993). *Workplace Literacy Project. Computer aided instruction in basic workplace skills, May 1, 1991-October 31, 1992. Performance report. Executive summary.* Flint, MI: Author. (ERIC Document Reproduction Service No. ED 378 308)

Purpose: The Mott Community College (MCC) Workplace Literacy Project was designed to provide basic workplace skills to employees from Lapeer and Genesee Counties. Individualized training plans were developed for each employee. They were then placed in classes customized to the needs of the workplace and individualized to their assessed skills. The classes offered were basic mathematics, beginning algebra, intermediate algebra, college algebra, analytical geometry/calculus I, basic sentence skills, basic writing, English composition, reading improvement, and reading-study skills development.

Variables: Variables measured were reading, writing, and math for work-related activities, critical-thinking and problem-solving skills in workplace situations, and interpersonal and communication skills in workplace interactions.

Population: Five hundred and sixty-six employees from 20 companies were identified for participation. It was not clear how many of these employees actually

enrolled in classes or how many completed classes. Furthermore, many classes had duplicate counts. Thirty-nine percent were between the ages of 20 and 30; 35% were between 21 and 40; 17% were between 41 and 50; and 7% were over 51 years of age. Sixty-four percent were males. Only 6% had not received a GED or high school diploma. Eighty-six percent were White, 9% Black, and 3% were Hispanic or Native American.

Design: The study had a pre-post design. Training was formatively evaluated at the midpoint of each training sequence and summatively evaluated at the end of each training sequence. The partnerships between MCC and the 20 work sites had a substantial variability of participation. For example, two partners had only one employee in the program, while another had 86. The average participation per partner was 28.3.

Data Collection: Once employees were identified, they completed a battery of assessment tests. Supervisors were asked to fill out a five-point scale survey to rate employees.

Instrumentation: Instruments used were MCC's placement examination, the TABE, and the Myers-Briggs Personality Test. Only 214 participants were given placement tests.

Findings: Narrative descriptions of outcome data were almost nonexistent, although statistical tables were presented. *T* tests for paired samples were provided. Performance in each academic area improved significantly, with math showing the greatest difference and vocabulary and comprehension showing the least.

Supervisor ratings indicated that all partners considered the goal of helping employees advance in their careers at least moderately well fulfilled. Most also felt it increased productivity. Supervisors representing 70% of participants noted that 26.4% had extreme improvement in their self-esteem, 36% had extreme improvement in worker moral, and 27.8% had extreme improvement in attitude toward continuing their education.

Strengths: There were no real strengths.

Weaknesses: Very few of the participants had real outcome data. There was substantial duplicated counting of participants. The data generated were poorly described. Some of the data reported were subjective views of supervisors. It was

not clear what some of the placement instruments were measuring. There was substantial variability of employee participation in the 20 companies.

National Center for Family Literacy. (1992) *Dissemination process submission*. Louisville, KY: Author.

Purpose: This submission was a proposal for a U.S. DOE dissemination grant which included some evaluative data on family literacy programs using the Kenan model. Adult literacy education was a required component of all Kenan model programs.

Variables: The National Center for Family Literacy (NCFL) had a comprehensive evaluation plan which included outcome measures for children as well as adults. The evaluation included tested learning gain, public assistance, employment, and further education.

Population: The study population varied according to the program site. All subjects were engaged in adult literacy education. The Kenan Family Literacy Model Programs had seven sites with 122 students; sites in Hawaii had 53 students; a site in Salem, Oregon had 18 students; and one in Mesa, Arizona had 50.

Design: Except for tested learning gain, most of the data were descriptive. There was a pre-post follow up, but the time between baseline and follow up varied by program. No control or comparison group was used.

Data Collection: Data collection procedures were not explained in detail. The length of time between baseline and follow up was an average of 22 weeks of attendance at the Kenan model programs, two years at the Hawaiian sites, and one year at Salem, Oregon. The length of time was not reported for the Mesa, Arizona site.

Instrumentation: An intake survey was used. Follow up and tested learning gain were measured by the TABE or CASAS.

Findings: Kenan Family Literacy Model Programs: Grade-level gains were reported for reading, mathematics, and language. In reading, 41% gained less than one year, 30% gained one year, and 29% gained two to three years. In math, 20% gained less than one year, 25% gained one to two years, 24% gained two to three years, and 24% gained more than three years. In language, 20% gained less than

one year, 25% gained between one and two years, 10% gained two to three years, and 45% gained more than 3 years.

Fourteen learners passed the GED; 41% were in some form of higher or continuing education; 66% were enrolled in educational programs, had definite plans for further education, or were employed; and 35% were employed (versus less than 10% at enrollment). In addition, 41% were not receiving public assistance, at enrollment 34% were not.

Hawaiian sites: At the Hawaiian Competency-Based High School Program students could take either the GED or earn a school district diploma. Of the 29 students enrolled in the first two years who did not have a diploma, all but four received high school certification. At one site, all adults ($n = 22$) who completed the year enrolled in summer programs at the community college.

Salem, Oregon: Seven of the 18 passed the GED. The average CASAS gain in reading was 13.5 scale points and the average CASAS gain in math was 16.5 scale points. Five students obtained a driver's license, 13 registered to vote for the first time, four enrolled in regular programs at the community college, and four entered work for the first time.

Mesa, Arizona: Of the 50 adults in the program, seven entered work for the first time, eight continued in community college ABE after their children entered kindergarten, six obtained their GED, and 20 remained in the program.

Strengths: Comprehensive variables were measured.

Weaknesses: The report did not mention details on methodology or context. Most data appeared to be self-reported. There was no control group.

New York City Board of Education. (1985) *Adult basic education final report 1984-85. OEA evaluation report*. Brooklyn, NY: Author. (ERIC Document Reproduction Service No. ED 278 793)

Purpose: The Adult Basic Education/High School Equivalency (ABE/HSE) Services Program of the Board of Education's Office of Adult and Continuing Education provided services to adults at over 300 sites. Its primary goal was to help program participants attain a higher level of literacy and enroll in other programs to further their education. This report compared variables measured in 1983-84 and 1984-85.

Variables: Variables studied were reading and math gains for basic education (BE) students, and English language oral proficiency for ESOL students. Changes in course levels (e.g., from BE1 To BE2) and changes in student employment patterns were measured for both groups.

Population: The sample of students consisted of those who enrolled in both the 1983-84 and 1984-85 program years. The sample of 6,355 students was 33% of the total enrolled during 1983-84. Course-level advancement data were available for 1,302 BE students. Math scores were available for only 345 BE students and reading scores for 1,153. HSE program math scores were available for 109 students and reading scores for 299. Employment status data were available for 5,466 students.

Design: Evaluators conducted secondary analyses on data compiled from program reports. A number of students and teachers were interviewed to gather qualitative data.

Data Collection: Evaluators collected achievement test data from program files and reports.

Instrumentation: The JOHN test was used to measure ESL oral proficiency. The TABE measured reading and math gains.

Findings: Of the 406 BE students who entered at level one in 1983-84, 42% progressed one level, 4% two levels, and 1% three levels. Of the 439 BE students who entered at Level Two in 1983-84, 37% progressed one level, 43% two levels, and 16% three levels. The average grade equivalent score in math for BE students was 5.4 in 1983-84 and 5.6 in 1984-85. The difference in students' grade-equivalent math scores was not educationally meaningful. The average grade-equivalent score in reading for BE students was 4.8 in 1983-84 and 6.0 in 1984-85. The comparison in students' grade-equivalent math scores showed a moderate improvement in test scores. HSE program scores showed decline from one year to the next. In 1983-84 the average math score was 6.7, in 1984-85 it was 5.9. Average reading scores dropped from 8.3 to 7.7.

Thirteen percent of the 2,518 students who were unemployed and available for work in 1983-84 were employed in 1984-85. Of the 1,157 students who had been unemployed and unavailable for work, 9% became employed by the 1984-85 school year.

Strengths: The sample was large and a huge amount of data was analyzed.

Weaknesses: The attrition rate was high. Very few variables were measured. Much of the data were incomplete.

Orangeburg-Calhoun Technical College. (1995). *VISIONS for Greater Employment Opportunities. Final report.* Orangeburg, SC: Author. (ERIC Document Reproduction Service No. ED 379 511)

Purpose: The VISIONS project was a workplace literacy program conducted in partnership with Orangeburg-Calhoun Technical College. Participants were drawn from three companies. The project had five objectives, including the design and development of job-specific literacy audits and curriculum. Another objective was to provide instruction to 250 employees which would result in new or continued employment, career advancement, or increased productivity. This project also sought to increase the basic skills competencies of 90% of the participants.

Variables: Outcome variables were competencies in speaking, listening, reasoning, and solving problems.

Population: Two hundred and ninety-four adults participated in the workplace classes. More than 90% were between the ages of 26 and 65; 37% were White, 61% Black, and 66% were female. Sixty-two percent of the participants were married and 85% had attained their high school diploma or GED before taking part in the classes. All were employed by the Regional Medical Center of Orangeburg and Calhoun Counties; United Technologies, an automobile assembly plant; or Holman, a cement and mining operation. The latter two were located in very rural areas.

Design: The design was pre-post comparison. Upon completion of a task analysis, a competency-based curriculum was developed for each work site.

Data Collection: Data were collected by project staff.

Instrumentation: No standardized tests were used. All instruments that measured gains were competency-based and developed by project staff (primarily the people teaching the courses). All other documentation to collect data were created by project staff. The forms to measure participant outcomes included supervisor pre-post employee ratings and student interview or student evaluation of instruction.

Findings: Ninety-nine percent of the participants attained a gain. There was a 57% average gain in better job attitude and a 58% gain in job knowledge. As reported

by supervisors, participants showed a 50% gain in productivity. Self-assessment surveys indicated that instruction helped workers perform their jobs better, with ratings increasing from 72% in the first cycle to 91% in the fourth cycle.

Strengths: Useful materials and courses were developed for the three participating companies.

Weaknesses: Some participants were double counted. The lack of standardized testing makes it difficult to generalize this approach and its results. Teachers administered the pre- and posttests that they had developed. There was no control group.

Pacific Training and Technical Assistance Corp. (1974). *Evaluation of the community based Right to Read program*. Berkeley CA: Author.

Purpose: This study was a U.S. DOE-sponsored evaluation of the Right to Read program. The evaluation was quite comprehensive, looking at program effects as well as learner outcomes. The evaluation focused on 13 programs serving in-school youth and 11 programs serving out-of-school adults. Data on the adult projects are reported here.

Variables: The variables studied were tested learning gains, students characteristics, and program characteristics.

Population: The study population consisted of Right to Read participants in 11 randomly selected sites. The cumulative enrollment was 1,198 students. Six hundred and forty participants were continuees and 322 received pre- and posttests (a completion rate of .50).

Design: The design was pre-post comparison. Participants were administered reading tests and gains were measured against student, teacher, and program characteristics.

Data Collection: Pretests were administered in November and December 1973. Project staff monitored the testing. Posttests were administered in May 1974. Student and staff data were collected bimonthly until May 1974.

Instrumentation: The Reading Everyday Activities in Life (REAL) test was used to measure learning gain. Learners' attitudes toward reading were measured with an attitude scale developed by the researchers. Student data were collected through intake and termination forms. Other instruments were attendance data

forms, a staff questionnaire, an on-site observation guide, and a project director's interview guide.

Findings: Significant gains in reading occurred. Average gains were 6.15 in the four- to six- month period. There were no significant differences between programs using paid tutors and those using volunteer tutors. The termination rate was 30.9%. Those who could read in their native language gained more than those who could not. Nonnative speakers who were fluent in English did better than those who were not fluent. Those who perceived their participation to be volunteer did better than those who did not.

Strengths: The sample size was substantial. The design was well conceived and executed. Comprehensive variables were measured. It was a well-documented report.

Weaknesses: There was no control group.

Payne, B. (1989). *Project F.I.S.T.: Functional In-Service Training. Submission for program recertification.* Edison, NJ: Middlesex County College.

Purpose: This was the application for Joint Dissemination Review Panel recertification for Project F.I.S.T., a volunteer-based administrative and instructional delivery system aimed at meeting the needs of low level readers. Student and tutor outcomes were measured. Solely the latter is discussed in this abstract.

Variables: Student outcomes measured were reading achievement and change in reading behaviors.

Population: Fourteen (14) were from the home site and four from an adoption site in Maine (n=14). The replication group (second data set) consisted of five students from the home site and 48 from the Literacy Assistance Center (LAC) database.

Design: A single group pre-post design with replication was employed.

Data Collection: Data were collected by project staff. The standardized reading test was administered at entry level and at four-month intervals.

Instrumentation: The SORT and TABE were the standardized tests used. A questionnaire was also administered.

Findings: Students made significant progress on the SORT. They averaged eight months of improvement in four months. In addition, when compared to the data from LAC (of 110 students and 3.0-4.9 grade level), the average gain for F.I.S.T. students on the TABE was 1.1 after four months of instruction versus the same gain after 200 hours for the LAC sample.

Strengths: Results were compared with those of a nationally validated program and the LAC database.

Weaknesses: There was a low n. The sample of students was described in a confusing manner. Three different standardized reading tests were used.

Ramussen, C. (1992). *Workplace basic skills in the metal casting industry for world class process and technology*. Alexander City, AL: Central Alabama Community College.

Purpose: Between March 1991 and August 1992, Central Alabama Community College and Robinson Foundry, Inc. established a workplace basic skills project for the metal casting industry. This project sought to measure seven employee objectives and four industry objectives. The two institutions wanted to establish a correlation between literacy gains and productivity measures.

Variables: Employee outcome variables were literacy achievement, knowledge and performance of industrial process, and GED completion.

Population: Three hundred and three workers enrolled; 216 completed all 16 hours of instruction and had pre-post data.

Design: The design was descriptive with a pre-post component.

Data Collection: Exit surveys were administered during the second and third reporting periods. Data collection methods were not described.

Instrumentation: The TABE was used to measure general literacy. Instructors developed pre- and posttests to measure knowledge of industrial process and concepts of management. Performance of the industrial process was measured by surveys that the instructors generated based on a supervisor rating form developed by a workplace literacy consultant.

Findings: A correlation between literacy gains and productivity measures was not established. No one attained a GED. The TABE scores were difficult to interpret;

however, reported reading gains were two to three grade levels. Seventy to 85% of participants perceived the project to have helped them with reading, writing, computation, and problem solving; and 70% said it had helped job performance. Job attitude improved by 71% and job knowledge by 62%. Retention was 71%.

Strengths: There were no strengths.

Weaknesses: The treatment period, 16 hours, was short. The researchers did not provide samples or descriptions of instructor-developed tests. Much of the data were self-reported.

Scheier, E., & Senter, D. (1969). *Evaluation of Learning 100: An adult basic education project in Bedford-Stuyvesant, 1967-68*. Huntington, NY: Educational Development Laboratories, Inc.

Purpose: L-100 was a multimedia communications skills system developed by Educational Development Laboratories for use with undereducated adults and youth. This study sought to compare its effectiveness in an undereducated adult population from the inner city with a more traditional reading program.

Variables: Reading achievement and hours of instruction attended were measured.

Population: Forty-nine students enrolled in the two L-100 classes. A control group of 47 enrolled in two classes that used traditional reading curricula, such as SRA Kits and the Lippincott Reading for Meaning Series. Only 37 of the experimental group and 26 of the control group had pre-post data. Sixty percent of the students in the experimental group had completed 4-7 years of schooling, while 62% of the control group had completed 9-12 years of schooling. All the members of the experimental group were employed, as were 73% of the control group. Most of the students from both groups were women. Sixty-two percent of the experimental group were between 35 and 49 years of age; 65% of the control group were between 20 and 34 years of age.

Design: This study had an experimental design. The analysis of covariance statistical technique was used. Qualitative data were classified and summarized.

Data Collection: The Metropolitan Achievement Test (MAT) was administered after 10 hours of instruction, after 100 hours of instruction, and at 200 hours of instruction. Research staff administered questionnaires to teachers and interviewed the teachers.

Instrumentation: The MAT was used to measure reading achievement. Subjective data were collected by questionnaires completed by and interviews given to the teaching staff.

Findings: Twelve (24%) of the experimental group and 21 (44%) of the control group left the program without completing the prescribed 200 hours. No correlation was made between retention rates and method of instruction. The average hours of instruction for the experimental group was 157 and 170 hours for the control group. Again, no correlation was made between average hours of instruction and method of instruction. The experimental group scored significantly higher on the MAT than did the control group.

Strengths: There was a control group. There was a more than adequate treatment period. Statistical adjustments were made for pretest differences. Variables were correlated statistically.

Weaknesses: There were small sample sizes in each group. The experimental and control groups did not seem comparable in several areas. Few variables were measured. The reading instrument was not normed for adults.

Seaman, D., & Denyer, J. (1991). *Evaluation of a workplace literacy program: A cooperative effort between Houston Lighting and Power Company and North Harris County College*. College Station, TX: Texas A&M University.

Purpose: Between 1990 and 1991, Houston Lighting and Power Company, in conjunction with North Harris County College, developed a workplace literacy project to prepare the company's employees for the commercial driver's license (CDL) test.

Variables: Variables were reading achievement, passing the CDL test, and desire to further one's learning.

Population: Originally the project meant to serve 200 employees in the pilot and regular classes. This goal was not met due to lack of release time by the employer. The report did not make clear how many employees actually participated. The target pool was originally 1,200 employees, of which 1,190 were assessed by one of two instruments and 500 participated in a one-day class on "test-taking skills."

Design: The report was primarily descriptive with a pre-post component.

Data Collection: It was not clear who collected data or how it was collected.

Instrumentation: Instruments used were the CDL Basic Skills Assessment and the TABE.

Findings: Of those who enrolled in the regular class, 99% passed the CDL test.

Strengths: Independent evaluators assessed the project's pilot class and the main instructional program that followed. Evaluation was based on Steele's (1990) four-component model.

Weaknesses: There was no control or comparison group. The sample size was not clear. Data collection methods were not clearly described.

Sherron, R. (1986). *A longitudinal evaluation of adult basic education in Virginia: 1970-1985*. Richmond VA: Virginia State Department of Education.

Purpose: This research was actually a cohort study, although it was called a longitudinal evaluation. Similar instrumentation was used in 1970, 1978, and 1985, and the results were compared. Most of the data described program and teacher variables. Some data on student impact were presented.

Variables: For the three cohort years, these variables were compared: income, employment, reasons for nonemployment, residence moves, post-ABE training, attendance, and learning gain.

Population: There was a 10% random sample of students. The actual sample size was not given.

Design: The design was a descriptive cohort study with cohorts in 1970, 1978, and 1985.

Data Collection: The methods and procedures of survey administration were not described.

Instrumentation: The test used for tested leaning gain was not described. Other data were collected through survey.

Findings: Students' income increased over time as did students' family income. It was not reported whether the data were adjusted for inflation. Full-time employment decreased over time. Reasons for nonemployment remained stable over time. Over time, more students did not change their residence within a five-year period. More students expected to attend a college or university over time; preferences for other types of post-ABE training did not change. In 1970,

significant gains in both reading and math were achieved; in 1978, gains in reading were not significant, but gains in math were. In 1985, gains in reading were not significant, but gains in math were. Neither the test used nor the time interval between pre- and posttesting were described.

Strengths: There were time series data and comprehensive variables.

Weaknesses: Methods and procedures were not adequately explained. This was a cohort study, not a longitudinal study. Without a control group, gains over time can not be explained by participation in ABE.

Siefer, N., & Latkiewicz, J. (1992). *Third project evaluation report--summative for Salt Lake Community College*. Salt Lake City, UT: Salt Lake Community College. (ERIC Document Reproduction Service No. ED 355 373)

Purpose: Salt Lake Community College, the National Semiconductor Corporation, and Natter Manufacturing, Inc. collaborated in implementing a workplace literacy project. This project had eight objectives, including conducting a literacy audit, performing task analysis for selected jobs, designing and implementing curriculum, and evaluating and disseminating the results. It was originally designed to serve 225 employees.

Variables: Variables measured were impact on employees' workplace literacy levels, workplace technological skills, workplace communication skills, participation in team meetings, and job advancement.

Population: Three hundred and twenty-five employees from the National Semiconductor Corporation enrolled in Pre-Statistical Process Control (SPC), ESL, math, and technical reading and writing classes. Thirty-five employees from Natter Manufacturing attended classes in math, communication, and ESL. Most of them enrolled in the ESL class and many of them enrolled in more than one class.

Design: There was a pre-post design. The evaluation plan for this project generated soft and hard data.

Data Collection: Project staff collected the data. Procedures were not made clear.

Instrumentation: During the project it was difficult to pretest for the course objectives and project personnel did not want to rely on generic skill tests. Checklists, supervisor ratings (four- item scale), personal interviews, and

anecdotal summaries were used to collect data and assess the impact of the project.

Findings: Natter Manufacturing: Of the 35 enrolled in refresher math, 26% showed gain to the 80% mastery level. Of the 19 enrolled in the workplace communication class, 37% showed gain to the 80% mastery level. Five people (71% of the class) showed gain to the 80% mastery level in technical reading and writing.

National Semiconductor: Of the 137 who attended the refresher math course, 82% showed some gain, and 63% achieved mastery level. Of the 101 participants who took both pre- and posttests, 86% showed gain and 53% achieved mastery level.

Self-reported data indicated that some employees were reading the newspaper more frequently, were more involved with their children's homework, and had increased self-esteem.

Strengths: This evaluation study had no real strengths.

Weaknesses: Many classes met for so few hours that true impact could not be measured or attributed to participation in the classes. Learner count was duplicated in many cases.

Smith, M., & Dalheim, Z. (1990). Project READ. A study of twenty reading disabled adults. Pittsfield, MA: Pittsfield Public Schools. (ERIC Document Reproduction Service No. ED 328 726)

Purpose: Project READ's teaching objective was to improve the reading levels of adults diagnosed with learning disabilities. Its three research objectives were to (a) investigate which of three teaching approaches was most effective with this population, (b) study what is effective assessment, and (c) measure the impact of reading difficulty and subsequent remediation on an adult's self-esteem.

Variables: The outcome variables were reading gains, the impact of instruction on self-esteem, and changes in students' lives.

Population: Initially 25 students were selected for the project. Their ages ranged from 20 to 67 years. Eighty percent were men. Most were employed. Seventeen had school-aged children. All scored below fifth grade on a battery of tests. All had intelligence in the normal range and English was everyone's first language. Five dropped out before completing the program ($N = 20$).

Design: The project had pre- and posttesting over time and a case-study design.

Data Collection: Project READ took place and data were collected from 1986 to 1990. Data were collected at the beginning of the project and at its conclusion. However, it is not clear at exactly what other intervals collection occurred.

Instrumentation: The study used a variety of instruments. Standardized tests included the Wechsler Adult Intelligence Scale to measure intelligence and discrepancy between reading accuracy and oral comprehension. Subtests from the WAIS were used as auditory and visual memory measurements. The Wide Range Achievement Test was used to measure word recognition, spelling and writing, and the Woodcock Reading Battery to assess word attack skills. Silent reading was measured by the Stanford Diagnostic Reading Test or the TABE; oral reading was assessed by Gary Oral Paragraphs and Gilmore Oral Paragraphs. Self-esteem and other psychological variables were measured by the Depression Adjective Checklist, the Gordon Personal Profile Inventory, and the IPAT Anxiety Test. Project staff created intake and completion interview instruments. In-depth case studies were kept on many of the students.

Findings: Although the samples were small, this study produced many results because so many types of comparisons were made and reading strategies assessed. Consequently, there were too many outcomes to delineate in this abstract. The major findings are noted below.

Reading gains were measured by grade-level gain and the number of lessons required to attain it. The average grade-level gain for the CC group was one year and three months after 210 lessons; the RFS grade-level gain was three years and three months after 203 lessons. Compared to the average gain in public school, the CC group was 1.5 times the average and the RFS group 3.1 times the average. Oral and silent comprehension were a key area of difference between the CC and RFS groups. The CC group gained eight months while the RFS groups had a gain of four years. This was a rate gain of 4.6 times the average.

The RFS and EQ groups were compared for reading gains on untimed oral tests. The EQ group gained three years and six months after 184 lessons. The RFS groups gained three years and nine months after 216 lessons. When compared to public school averages, the EQ group was 4.0 times the average and the RFS group was 3.3 times the average. Again, the greatest differences were found in comprehension. The EQ groups had a rate gain of 5.1 versus 3.6 for RFS. In silent reading the two groups were almost identical with nearly a 4.0 gain for both.

The groups showed greater gains in comprehension than in phonetic and word structure decoding, even though the curriculum stressed these latter skills. The researchers concluded that students with learning disabilities obtain sufficient phonological understanding and have enough practice in the use of phonological and word structure clues that it frees them to attend to comprehension. This conclusion ran counter to the accepted beliefs about phonics instruction.

More than 50% of the students indicated that their plans for the future were different than when they enrolled. Seventy percent of participants noted that they had higher self-esteem and 60% had job changes which were upwardly mobile.

Strengths: This study had an adequate design. Three rigorously selected samples were compared to one another. Many variables were analyzed, compared, and correlated. The data collected generated quantitative and qualitative information. This type of study can help literacy practitioners in the classroom.

Weaknesses: There were too many assessment instruments. There were low sample sizes. Description of data collection methods was weak. Too many reading skill variables were analyzed and some seemed redundant. Because of the way it was structured, Project READ and this study would be difficult to replicate.

Snoddy, J. (1990). *Macomb Reading Partners research report*. Clemens, MI: Macomb County Library. (ERIC Document Reproduction Service No. ED 380 560)

Purpose: This study evaluated Macomb Reading Partners (MRP), a one-to-one volunteer tutoring program of the Macomb Literacy Project. The primary objective of the study was to examine the reported motivations of adults with low levels of literacy who seek entry into a literacy program. It also sought to investigate the conclusions of Aslanian and Brickell (1980), who found that adults in transition or experiencing a trigger event (such as divorce, being laid off, finding a new job, or illness) tend to seek new learning experiences. The two major research questions were: (a) What are the motivations reported by low-literate adults for entering a volunteer literacy program; and (b) What are the levels of literacy of the adults in this study? The study also focused on the employment patterns of participants and their coping or concealing strategies in a literate world.

Variables: The outcome variables were gains in reading grade levels and the application of literacy skills in daily lives. The study also summarized participant responses to the questions: "Do you feel better about yourself?" and "Why or in

what ways do you feel better about yourself?" The responses were inductively coded and grouped into the categories of improved literacy, enhanced self-confidence or self-esteem, personal accomplishment, and enhanced job opportunities or performance.

Population: Selected participants had to have been enrolled in the literacy program for at least six months and have no mental, emotional, auditory, visual, or neurological impairments. Students' grade-level reading equivalent ranged from primer to fifth grade. Seventy-three percent were men and the mean age was 38. Seventeen percent had completed grades 6-8, 40% grades 9-11, and 43% completed high school. Most were employed and approximately 60% worked in the automotive field.

The sample was selected from the list of participants who remained after the first two criteria were met. Subjects were then selected randomly and contacted by telephone about their availability and willingness to participate. This procedure continued until a sample of 30 was selected. The researchers believed this sample was representative of all participants in the program.

Design: The design was pre-post comparison. Quantitative data were reported and analyzed by using frequency counts, measures of central tendency, and other descriptive comparisons. The data generated from the selected sample were compared with the general population of MRP.

Data Collection: This study lasted approximately one year. Planning began in October 1989, data collection ended in July 1990, and data analysis was completed by September 1990. Program participant files were used to collect information about experiences with tutoring and to validate information gained during interviews. Interviews were conducted by the researchers at the beginning of the study and at its completion. Interviews generally lasted an hour. Researchers took extensive notes throughout the interviews (they were not taped). Responses were inductively coded and grouped into categories. A comparison group was not interviewed.

All students at MRP were tested at six-month intervals, a procedure that continued throughout the study. Initially students were tested by a trained member of the staff. Thereafter, the tutor administered the test. Program staff scored and recorded the results.

Instrumentation: The Word Opposites Test of the Botel Reading Inventory was used to measure reading gains. Life skills and self-esteem were explored by an

interview protocol devised by the researchers based on a study by Darkenwald and Valentine (1984). Four pilot interviews were conducted and the protocol was revised slightly.

Findings: The average reading gain per month was .204. In general, the results showed that the lower the grade-level equivalency of an adult at entry the greater the gain per month. Furthermore, the older student, aged 40-58, had higher per month achievements than younger participants. The average monthly gain was significantly higher for those tutored for 10 months or less than for those tutored 13 months or more. Qualitative data indicated that many of the students had changed to more effective reading strategies.

All participants answered affirmatively when asked, "Since you started this program, have you used your reading skills to do something you couldn't do before, or to do it better?"

Study subjects were asked to name some of those things. All the answers were delineated. The most common were reading a menu, mail, Bible, newspaper, and magazine article. Voting was also mentioned fairly often. Ninety-seven percent of the respondents reported feeling better about themselves. They also mentioned most often that feeling better was due to improved literacy and enhanced self-confidence or self-esteem.

Strengths: In general, the research design was adequate. Quantitative data were measured against a comparison group. There was a solid interview protocol and rigorous analyses of qualitative data. Researchers based their study on and compared their results to previous research.

Weaknesses: The sample was very small. Given the test used, true reading-gain results could be questionable (e.g., no reading in context). Some of the data were self-reported.

Turner, R., & Stockdill, S. (Eds.). (1987). *The Technology for Literacy Project evaluation. A special report from the Donor Review Board.* St. Paul, MN: Saint Paul Foundation. (ERIC Document Reproduction Service No. ED 295 028)

Purpose: The Technology for Literacy Center (TLC) implemented a project designed to investigate and execute new, innovative ways to teach literacy skills to adults living in the greater St. Paul area. It had several components: direct service, staff development, research, and incentives grants to existing literacy organizations. TLC developed a competency-based program built on a mastery

system of instruction. The project implemented formative and summative evaluation plans. This abstract describes the student impact results of the summative evaluation.

Variables: Impact variables were reading and mathematics achievement. Attrition rates and reactions to the TLC approach were also measured.

Population: This study focused on 165 learners who enrolled between September and November of 1986. They were considered the summative group. Most of them were English-speaking adults who lived in the St. Paul tri-county area. Eight-one percent (81%) of the learners read below the eighth grade level at intake and 41% of them had not previously enrolled in a literacy program. Fifty-one percent were White, 28% Black, 8% Hispanic, 5% Asian, and 5% Native American. All were voluntary students. Only 50 were still in attendance when achievement gains were measured.

Design: Formative and summative evaluation plans were implemented. The outcomes of the summative group were compared to a control group of other TLC enrollees. They were also compared to learners in local and nationally recognized programs. All of the achievement data gathered were reported quantitatively. Focus group data were analyzed qualitatively.

Data Collection: A formative evaluator, employed by the supporting foundation, designed a system for data collection, and collected and analyzed most of the data. Data were regularly collected from student files, which were updated with each student visit. Data were aggregated quarterly. Students were given the TABE at enrollment and in April for everyone who had attended at least 12 hours of instruction.

Focus groups were used to collect qualitative data. Three were conducted by the primary summative evaluator during TLC's regular hours. The interviews followed a six-question protocol. A total of 20 people took part in the focus groups.

Instrumentation: Instruments used to collect data were standardized achievement tests, group interviews, exit interviews, student questionnaires, learner activity logs, and case studies. The TABE was the standardized achievement test. Criterion-referenced tests were also developed to document progress on the computer-based modules.

Findings: The outcomes of the summative group were as follows (numbers were reported as means):

The composite reading gain was .75 grade level for students who tested between 0 and 4 on the pretest ($n = 15$). The gain in comprehension was .98 grade level. The vocabulary gain was .6 grade level. This group averaged 47.2 hours of instruction. It was calculated that 63.2 hours were needed to make a full grade-level change in reading, 48.2 hours were needed for a full grade-level change in comprehension, and 78.7 hours were needed for such a change in vocabulary.

The composite reading gain was .85 grade level for students who tested between 5 and 12 on the pretest ($n = 23$). The gain in comprehension was 1.13 grade level. The vocabulary gain was .79 grade level. This group averaged 44.5 hours of instruction. It was calculated that 52.5 hours were needed to make a full grade-level change in reading, 39.6 hours were needed for a full grade-level change in comprehension, and 56.6 hours were needed for a full grade-level change in vocabulary. In sum, the impact of TLC upon the summative group was positive for reading comprehension and math, but not for reading vocabulary. The control group showed no such gain in comprehension.

The researchers also provided a comprehensive analysis of the impact of hours of instruction in relation to achievement gains. TLC's results were also compared to the three validated NDN literacy programs. The researchers concluded that TLC's reading gain results were "in the ballpark" with the three programs. However, the same conclusion was not reached for math gains. JCARP's students showed 13.7 hours per grade-level change on the ABLE compared to TLC's 45.4 hours.

Qualitative data indicated that TLC participants felt they were learning and many of the learners described the differences it had made in their lives--reading to their children, keeping pace with their children, enhanced self-esteem, and ability to be self-motivated.

Strengths: This was a solid experimental study with a strong summative evaluation. Analysis of data was very comprehensive and well described. There was a control group.

Weaknesses: The sample size was small for measuring impact gains. Hours of instruction were limited for many of the participants. Attrition rates were high.

Wasson, J. (1965). *Evaluation of an eight week adult education program*. St. Paul, MN: Ramsey County Welfare Department, Work and Training Program. (ERIC Document Reproduction Service No. ED 021 194)

Purpose: The Ramsey County Welfare Department and the St. Paul Public Schools conducted an eight-week summer adult education program for public assistance recipients. The clients were referred to a work experience program, a work training program, or an education program. The purpose of the study was to describe and analyze the outcomes of the educational program, and relate those outcomes to student demographic and psychometric characteristics.

Variables: The variables measured were reading and math gains, changes in work values, and student perception to the program. Changes in counselor description of clients and counselor perception of program were also measured.

Population: The population of 34 males and 19 females were clients of the Ramsey County Welfare Department's Work and Experience Project. All had an average IQ. All the males and one female were married. Three females were single and 15 divorced. The males ranged in age from 19 to 52 years, with a mean of 32.0. The females ranged from 23 to 46 years. Forty two of the participants were White and 10 were Black. Forty-one percent of the males and 79% of the females had been unemployed for more than 52 weeks. Highest grades completed were the following:

<u>Highest Grade Completed</u>	<u>Males</u>	<u>Females</u>
6th grade or less	14.7%	
7th grade or less	5.9%	
8th grade or less	5.9%	10.5%
9th grade or less	26.5%	31.6%
10th grade or less	29.4%	21.0%
11th grade or less	17.6%	21.0%
12th grade or less	2.9%	15.8%

Design: The study had a pre-post comparison design.

Data Collection: Reading, math, and psychometric tests were administered by a trained psychometrist in groups of 8 to 12 during two half-days of testing. Reading and math instruments were administered three times during the program period: at intake, during the first week of class, and during the last week of class.

Instrumentation: The Gates Reading Survey was administered to measure reading speed, vocabulary, and comprehension. The Jastak Wide Range Achievement Test measured computational ability. Both instruments yielded raw and grade scores. A scale to measure work-value changes was administered. It was a 53-item measure of work attitude developed by Dr. Arthur Bradely and the Minneapolis Veterans' Administration Hospital. The Army General Classification Test measured intellectual capacity. Percentiles were transformed to IQ scores. The Minnesota Paper Formboard was used to test spatial reasoning. The Kruder Preference Record, Vocational, Form CH was administered to measure vocational preference.

Findings: Reading speed increased from 7.9 to 11.2 for the thirty participants who had valid pre- and posttest reading scores. Reading vocabulary increased from 9.3 to 10.3 and reading comprehension from 9.5 to 10.5. The average grade-level increase of math scores was 1.92. Seventeen people took the GED test. Six passed and 11 failed. The data suggested that those who scored lower than 9.0 at intake were not able to pass the GED after eight weeks of instruction. All who read 10.0 and better at intake passed the GED test.

Upon completion of the course, 36% were either going directly to work or into an employment training course. These training programs included janitorial, dietary assistant, nurse's aide, laundry, and orderly courses. Thirty-eight percent chose to enter educational programs.

There were no significant changes in the students' responses to the Kruder work-value scale.

Strengths: Valid assessment instruments were used. Variables were correlated. The research design and data collection methods were adequate.

Weaknesses: There was a small sample size. There was no control group.

White, G. (1990). *A comparative study of adult education Indianapolis/Richmond. Third party evaluation final report. Indianapolis, IN: Indiana Opportunities Industrialization Center of America State Council.*

Purpose: This study sought to compare the effectiveness of traditional classroom instruction to CAI for adults enrolled in ABE classes. Its purpose was to measure the effectiveness of the methods in raising the levels of basic skills after 80 hours of instruction. Effectiveness of instruction was also correlated with the various learner styles of the students. The CAI site was located in Indianapolis, IN; the traditional site in Richmond, VA.

Variables: Outcome variables measured and compared were reading and math achievement, retention rates, and self-esteem.

Population: One hundred and forty-nine students enrolled, but only 50 met the criteria for comparison purposes. The majority were women ranging in age from 18 to 67 years. The mean entry reading level on the TABE was the equivalent of 7.5.

Design: The design was pre-post comparison.

Data Collection: Methods included classroom observation, student survey and interviews, reviews of instructional records and materials, joint site meetings, and interviews with project administrators. Data was collected upon entry, at 30-day intervals, and upon exit over a period of 18 months.

Instrumentation: The ABLE was used.

Findings: The average gain for CAI students was 2.6 grades and 1.84 for traditional students.

Strengths: Effects of methods of instruction were compared. Effectiveness of methods of instruction was correlated with various student learning styles. Anecdotal information regarding workplace literacy programs was comprehensive.

Weaknesses: Most data were self-reported. There was no clear research design.