EVALUATING THE OUTCOMES OF LITERACY TRAINING: A FEASIBILITY STUDY

Cheryl L. Madill
Department of Health Studies and Gerontology
University of Waterloo

Anita M. Myers
Department of Health Studies and Gerontology
University of Waterloo

Abstract: Challenges inherent in evaluating literacy training programs include strict confidentiality policies, the reluctance to "test" students, irregular meeting schedules, and the limited attention span of low-literate students. Guided by interviews and focus groups, this project explored the feasibility of tutor-based assessments in three domains: cognitive skills, functional literacy, and literacy-related self-confidence. Over a period of one and a half years, the assessment battery was returned by 38 of 67 tutors from two literacy programs—one highly structured and curriculum based, the other learner centered. Level of formal education and program placement were found to be important indicators of functional performance and confidence. Length of time the student had been in the program, meanwhile, was unrelated to functional performance and inversely related to ratings of self-confidence. Recommendations for further evaluation of literacy programs are based on our findings regarding utility of the various measures, feasibility for administration, and acceptability to students, tutors, and program administrators.

Résumé: Les défis à relever face à l'évaluation des programmes d'alphabétisation comprennent les politiques de confidentialité, l'aversion de mettre les étudiants à l'épreuve, les irrégularités des horaires de rencontre et la capacité d'attention limitée des étudiants analphabets. Guidé par des groupes de consultation et des entrevues, ce projet explore la valeur concrète d'une évaluation par tuteur dans les trois domaines suivants: habiletés cognitives, alphabétisation fonctionnelle et les aspects de la confiance en soi reliés à l'alphabétisation. Suivant une période d'un an et demie, les séries d'évaluations ont été retournées par 38 des 67 tuteurs des deux programmes d'alphabétisation (un structuré et planifié et l'autre centré sur l'apprentissage). Le nom-

bre d'années d'étude formelle et le placement au programme ont été suggérés comme d'importants indicateurs de performance fonctionnelle et de confiance. La durée de la participation au programme par contre n'a pas été reliée à la performance fonctionnelle mais inversement reliée aux mesures de confiance en soi. Les recommandations pour de futures évaluations de tel programmes sont basées sur les résultats concernants l'utilité des différentes mesures, la valeur concrète de l'administration, et l'acceptabilité aux étudiants, tuteurs et administrateurs des programmes.

Based on two national surveys, an estimated 16–24% of Canadians are functionally illiterate (Creative Research Group, 1987; Montigny, Kelly, & Jones, 1991). This figure drops only slightly when immigrants are excluded. Literacy has been identified as the most important factor in population health (Breen, 1992; Perrin, 1989), affecting workplace and household safety and the individual's ability to access the health care system. More broadly, literacy is directly linked to employment opportunities and overall standard of living. The growth in grassroots literacy training programs and the emergence of a literacy branch in the Ministry of Education and the Ontario Training and Adjustment Board (OTAB) attest to the recognition of this problem.

As in many areas of social programming, evaluation has not been an explicit criteria for either initial or continued funding, nor have credible evaluation strategies been used to guide program planning (Myers & Gonda, 1987). A review of the field (Padak & Padak, 1991) and a recent ethnographic study (Fingeret et al., 1994) both strongly recommended the development of systematic evaluation protocols for literacy training programs.

Although the Ministry of Advanced Education, Training, and Technology in B.C. has developed a set of "evaluation guides" for literacy training (Selman, 1991; Thomas, 1991), student satisfaction was the only criteria specified for evaluating program performance. Client satisfaction surveys can provide valuable "process" information, but it can be argued that consumers are unable to judge the competency of service delivery. Questions regarding perceived program benefits are also problematic, as some clients may inflate ratings as justification for continued participation. Persons who perceive that their needs are not being met by the program tend to drop out and are unlikely to be included in client satisfaction surveys. Most importantly, such questionnaires are typically anonymous, precluding linkages with other sources of information—client characteristics,

participation rates, and indicators of progress. Thus, client satisfaction measures are inadequate to assess program outcomes, much less the relative effectiveness of different programmatic approaches.

CHALLENGES IN LITERACY EVALUATION

Varying definitions of literacy, different training approaches, restricted access to students, and the reluctance to "test" literacy students are some of the challenges inherent in evaluating literacy programs.

Definitions

The most common definition of literacy—used by both the United Nations Education, Scientific, and Cultural Organization (UNESCO) and Statistics Canada—is "having at least a grade nine education" (Ontario Public Health Association, 1989). Such a cut-off point presents several problems, including fluctuating standards across time and locations, as well as a failure to consider self-education. The Southam News Survey found that 13% of their sample of 2,400 Canadians were "false literates" (i.e., had achieved a grade nine education but could not perform several literacy tasks to a criterion level) and 9% were "false illiterates" (Creative Research Group, 1987).

Whereas "basic" literacy is seen as a person's overall ability to read and write, "functional" literacy focuses on "the information-processing skills necessary to use the printed material commonly encountered at work, at home and in the community" (Montigny et al., 1991, p. 14). A number of other skills in addition to reading and writing skills—such as numeracy literacy and computer literacy—have also been identified (Open Learning Network, 1994). Operationally defining and measuring each of these literacy abilities presents similar challenges. For instance, should a cut-off point be used to separate "literates" from "illiterates," or should literacy be viewed as a continuum of skills, abilities, and experiences?

Training Approaches

Literacy training can be characterized by two ends of the spectrum—programs that are highly structured and "curriculum based" versus those that are "learner centered" (Sticht, 1990). The two programs chosen for this study—Laubach Literacy and Core Literacy—represent these differing approaches.

Laubach Literacy is an international organization whose programs offer a highly structured, phonics-based curriculum (Meyer & Keefe, 1988). As part of a 12-hour workshop, volunteer tutors are strongly encouraged to strictly adhere to the progressive set of four reading skill books (and accompanying check-ups) and to introduce supplementary materials only once the curriculum has been mastered. An advanced Challenger Series for reading is also provided, as are skill books in spelling and mathematics.

Similar to Laubach, a local grassroots program—Core Literacy—provides training for volunteer tutors. Students and tutors also typically meet on a one-to-one basis (although a small number of group sessions are offered in both programs). In contrast, Core does not have a set curriculum and encourages a number of training approaches guided by each student's personal objectives (such as completing an application for a driver's license).

Some programs—such as Laubach—offer both basic literacy training and ESL. Our focus groups with tutors revealed a number of differences in both characteristics and expectations of these two types of students. Students in basic programs tend to be Canadian born and educated and fairly homogeneous, in contrast to ESL students. A decision was made to focus on basic literacy training for this study. The stigma of "illiteracy" is particularly salient for students in basic programs, fostering a highly protective attitude on the part of both administrators and tutors. This attitude, as we discovered, has important implications for the evaluation of such programs, restricting direct access to both student records and students themselves.

Assessment Strategies

It has been argued that literacy assessment should address three components: cognitive (basic skills); enactive (transference of skills to daily situations); and affective (perceived achievements and confidence) (Finlay & Harrison, 1992).

Cognitively based reading and comprehension skills have typically been measured using standardized tests such as the Adult Basic Learning Exam (ABLE) (Karlsen & Gardner, 1986), the Canadian Adult Achievement Test (CAAT) (Psychological Corporation, 1986), the Slossan Oral Reading Test (Jackson, 1990), and the Wide Range Achievement Test (WRAT) (Metz, 1989). These tests are either norm-referenced or criterion-referenced. Numerous criticisms of their use

with adult literacy students (Lichtman, 1974; Sticht, 1990) include their producing test anxiety (reminiscent of formal schooling), containing patronizing content, and making score interpretation in grade-level equivalency difficult. The material is often too difficult for lower-level-literacy students, producing a floor effect in scores and precluding discrimination for placement purposes. Most importantly, such tests are not appropriate for assessing individual progress if the program's curriculum does not correspond with the test content (Sticht, 1990).

Enactively based tests of functional literacy focus on tasks and situations encountered in daily life, such as following a bus schedule and interpreting a prescription label. Functional assessments appear to be better received by students than traditional reading tests (Librande & Jones, 1988). Functional literacy tests developed for use in Canada include the two survey instruments developed for Statistics Canada (Montigny et al., 1991) and Southam News (Creative Research Group, 1987), and the Ontario Test of Adult Functional Literacy—the OTAFL (Jones & Librande, 1987). Of the three, the OTAFL has the most extensive psychometric evidence. There is some overlap between the instruments. For instance, the Southam survey used items derived from the U.S. National Assessment of Education Progress (NAEP) (Kirsch, 1985). The Statistics Canada survey, meanwhile, used both NAEP and OTAFL items.

Affectively oriented indicators reflecting gains in perceived achievement and self-confidence are viewed as critical measures in such programs particularly given the stigma associated with illiteracy (Finlay & Harrison, 1992). Tested against such global constructs as self-esteem and locus of control, situation-specific self-efficacy has consistently emerged as a predictor of program adherence and as an important outcome in and of itself (Strecher, 1986). Based on Bandura's (1982) theoretical framework, response or outcome expectancies (such as learning to read in order to get a better job) are important to initially motivate a person to enroll in a given program. Self-efficacy expectancies (or self-confidence), meanwhile, influence whether an individual will attempt a given task, as well as the degree of effort or persistence expended in the face of obstacles. Most importantly, self-efficacy is amenable to change. Performance accomplishments, as well as positive feedback and reinforcement from others (such as tutors), exert powerful influences on self-efficacy. To date, only two studies have empirically examined self-efficacy in the literacy domain. Shell and Murphy (1989) found that self-efficacy was significantly correlated with reading achievement

in college students. Gorman (1981) measured self-efficacy in adult literacy; however, no justification was provided for item selection, and the study was limited to tutors' ratings of their students.

The present study investigates the feasibility of assessing all three of the above indicators for tracking the progress of adult students in basic literacy programs. We wanted to compare student and tutor ratings, as well as the degree of correspondence between cognitive, enactive and affective measures. We hypothesized that students at higher levels in their program or in the program for longer durations should perform better on all three sets of measures. As suggested by Jones (1989) and consistent with Bandura's (1992) theory, we further predicted that the frequency with which a task is routinely performed would be related both to level of confidence in performing the task in question and to successful performance.

METHODS

The strict confidentiality policies of both literacy programs prohibited direct access to student records and the students themselves. From the Laubach councils, we were able to examine 50 randomly selected intake forms (names removed) that the administrator completes via interview to explore each prospective student's personal goals. A content analysis revealed that approximately 40% of incoming students mentioned such general goals as "improving reading and writing," and the same proportion mentioned vocational goals such as "getting a better job." The most common personal goals were "reading to my children" or "helping my children with homework" (19%) and "getting a driver's license" (13%). Otherwise, goals were highly individualized—each mentioned by only one or two people. Interestingly, several students alluded to the theme of self-worth in their goal statements (e.g., "bettering myself").

Interviews with program administrators and a number of focus groups with tutors were integral to the development of the evaluation protocol in several stages: instrument selection, pilot testing, and procedure selection. Although the participation of service deliverers has been advocated for evaluation in general, with literacy programs it is crucial, as direct assessment of students by outsiders is not permitted. In the months of preparation for this study, the senior author also went through the weekend tutor training workshop as a participant observer.

Instrumentation

> Given the reluctance on the part of both administrators and tutors
> to "test" their students, acceptability of the chosen "progress assess-
> ment" tools was a key consideration. We explored the possibility of
> using one of the standardized adult reading tests with our focus
> groups of tutors from both programs. This strategy was rejected by
> the tutors for reasons similar to those cited by Sticht (1990). Laubach
> routinely uses a series of curriculum-based "check-ups," which were
> used as the cognitive assessment component for tutors' students.
> Unfortunately, a comparable assessment tool did not exist for the
> non-curriculum-based Core program.

> The OTAFL was selected as the enactive tool based on its demon-
> strated psychometric properties (Jones & Librande, 1987) and tutor
> acceptability. The OTAFL consists of a series of three progressively
> more difficult sets of functional tasks commonly encountered in daily
> life (such as reading a prescription label and filling out an applica-
> tion form).

> A new instrument—the Literacy Efficacy Scale (LES)—was devel-
> oped to address the affective component operationalized as self-effi-
> cacy. Item generation began by taking the goals from the sample of
> student intake forms as well as all 100 items from the OTAFL. Each
> item was classified into one of 28 specific tasks (such as reading a
> medication label, reading a map, writing a cheque). Additional items
> representing tasks of lower- and higher-order difficulty were taken
> from the instruments developed by Gorman (1981) and Shell and
> Murphy (1989), respectively. Following Bandura's (1982) protocol,
> confidence in ability to do each task was rated on a continuum from
> 0% (no confidence) to 100% (total confidence). Students were also
> asked to rate each item from 0 = *not at all important* to 5 = *very
> important* (that they should be able to do the task), and from 1 =
> *never*, 2 = *sometimes*, to 3 = *often* (the frequency they actually do the
> task). The identical scale and rating formats were used to elicit tu-
> tor's perceptions of their students. Following tutor feedback, removal
> of redundancy, and item rewordings, 45 items were selected for the
> final LES scale and grouped into six subscales.

Procedure

> A list of tutor names was obtained from Core Literacy and three
> Laubach councils in Ontario. These three councils—located in an

urban, a rural, and a northern centre—were sampled to increase geographical representativeness. All tutors on the lists were sent a letter introducing the study and were telephoned approximately a week later. Eligible (currently tutoring an adult student in basic reading and writing) and consenting tutors were mailed a package that included the Laubach check-up exercises (for Laubach students), the OTAFL, the LES (for both student and tutor ratings), a background questionnaire, and an acceptability rating form. An instruction sheet for tutor administration and a telephone number to call collect for assistance were also included in the package. Finally, a certificate was provided for the tutor to sign and present to the student upon completion of the questionnaires. A postage-paid self-addressed envelope was included for returning the materials.

Tutors who had been called at least six times over approximately two weeks (at different times of the day and on different days of the week) were considered unreachable. Tutors who had not mailed in the completed packages within approximately six weeks were called approximately every three weeks until they either returned the package or withdrew from the study. These reminder phone calls also provided an opportunity for tutors to ask questions or report concerns. Notes were taken regarding verification of administration procedures and comments on the measures themselves.

As the study progressed, several tutors commented that they felt uncomfortable administering Laubach check-ups for skill books their students had not yet completed. A small telephone survey of 10 tutors from the Kitchener-Waterloo Literacy Council (who had received their packages first) revealed that the majority were concerned about adversely affecting student self-esteem by prematurely administering check-ups. (Normally, check-ups are administered only once the student completes the corresponding skill book). As no discriminative information could be obtained unless all levels of the skill book check-ups were administered, and to avoid adversely affecting participation, this instrument was subsequently dropped from the questionnaire package.

FINDINGS

Response Rate

In total, 124 tutor names were provided by the literacy organizations. Of these, 13 (10.5%) could not be reached, either because of

incorrect information in the literacy programs' records or because they were not at home at any of the six calling times. Among the 111 (89.5%) of tutors who were contacted, 67 (60.4%) were eligible to participate.

Tutors were ineligible if they were not currently tutoring or were tutoring a student in only math (versus reading). Of the eligible tutors, 9 (13.4%) refused to participate, citing such reasons as concerns about assessing their student and lack of interest. Among the 58 tutors who were mailed questionnaire packages, 38 (65.5%) returned those packages. The average time to complete the assessment package was three months (ranging from two weeks to seven months). A detailed breakdown of actual completion times for each component is provided later in this article.

Sample Profile

The average age of the tutors in the sample was 45 (range 17–82), compared to 41 for the students (range 21–70). Seventy-one percent of the tutors versus 39% of the students were female. Analysis of the tutor-student matchings revealed that all female students were matched with a female tutor, and the excess of male students meant that many male students were paired with a female tutor. Half of the tutors had a university education, 42% had completed high school, and the remainder had completed grade 10. Among students who reported their education level (n=35), 51% had less than a grade 9 education, 29% had completed some high school, and 20% had completed high school. The average age at which the students left school was 16.

Characteristics of Laubach students and tutors were compared to those obtained in a national survey conducted for Laubach Canada in 1990 (Sullivan, White, & Batstone, 1990). Student and tutor characteristics were similar, as shown in Table 1, with the exception that the proportion of our students who had been in their literacy program for more than one year was more than double that in the national survey.

Among 14 students who were administered at least one of the Laubach check-ups, average scores for exercises corresponding to Skill Books 1, 2, 3, and 4 were 97%, 78%, 83%, and 62%, respectively. Students scored well above the 75% criterion score on the check-ups that corresponded to the skill books their tutors reported

Table 1
Comparison of Laubach Students and Tutors: Current Study versus
Sullivan et al.'s (1990) National Survey

Characteristics			Current Study	Sullivan et al.
Students	Average age		41 (22–70)	35–44
	Gender (% male)		58	60
	Education (% less than grade 9)		58	63
	Percent in the program for the specified time	< 6 months	13	40
		6 months–1 year	19	30
		> 1 year	68	30
	Percent in each skill book level	1	4	13
		2	48	27
		3	30	38
		4	19	22
Tutors	Average age		47 (18–82)	45–54
	Gender (% male)		27	20
	Average time tutoring (years)		4 (<1–14)	3 (<1–19)
	Average number of students tutored		2 (1–15)	1–15

they had previously completed (average scores = 97%, range 86–100). For the five students who attempted exercises beyond their current skill book level, the average score was 59% (range 38–78). Two of the five students scored at the criterion level (i.e., at least 75%) on a higher-level check-up, but scores were relatively low (75% and 78%). Given that only a few attempted the more advanced exercises, tutor classification of current skill book level was used in the analyses regarding program level.

Potential scores on each of the three levels of the OTAFL are 24 (level A), 31 (level B), and 45 (level C). Sample means for the students in this study were 21.2 (range 9–24), 25.7 (range 11–30), and 37.8 (range 25–44), respectively. Of the 38 students in the sample, 36 completed level A, 32 level B, and 19 level C. All those completing higher levels had first completed the lower levels. The OTAFL scores were not normally distributed; thus nonparametric tests were used for analyses involving these scores.

Cronbach's alpha was computed for the newly developed LES scale. Both student and tutor confidence ratings were internally consistent for the scale as a whole (alpha=.96 and .98, respectively) and for

each of the six subscales (range .76–.96). As expected, average confidence scores across seven of the items (purposefully chosen from Shell and Murphy [1989] at a higher level of difficulty) were significantly lower than for the remaining 38 items (mean=41 vs. 63, p<.0001).

Total confidence scores on the LES were calculated by averaging scores across all 45 items, resulting in scores ranging from 0 to 100. Twenty-two students (58%) provided confidence data for all 45 items, and two students provided no data. Among the 14 students who provided incomplete data, an average of eight items were missed (range 1–37). There did not appear to be a particular item that was consistently missed. In order to make the most use of the available data, scores were also calculated for those students who were missing less than 10% of the items (i.e., students who were missing data for four or fewer items) by averaging confidence scores across the items they did answer. This technique resulted in confidence data for a total of 29 of the 38 students. Similar procedures were used to calculate average frequency and importance scores. Average frequency scores were calculated for 29 students (21 provided complete data), and 30 students provided sufficient data to calculate average importance scores (21 provided complete data).

The average overall confidence score by students was 61 (standard deviation=27.4, range 10–97); tutor ratings of their student's confidence level averaged 56 (standard deviation=29.4, range 10–96). Overall and subscale confidence mean scores were tested for normality using a Shapiro-Wilk statistic. There was evidence to reject the hypothesis of normality for the average confidence scores (p=0.01); therefore, nonparametric tests were used in analyses involving LES scores.

Tutors' average confidence scores were subtracted from those of their students, and differences were analyzed using a Wilcoxon signed rank test. On average, students' overall confidence scores were approximately 3% higher than their tutors' scores, but this difference was not significant (p=0.6).

Performance Indicators

Program level, based on tutor classification (for Laubach students only), was divided into two levels of achievement—those who were working on Skill Books 1 and 2 (n=14) versus those who were working

on Skill Books 3 or 4 or the Challenger series (n=17). Of these 33 students, 2 could not be classified into a skill book level because their tutors did not adhere to the Laubach curriculum.

Thirty of these Laubach students attempted at least the first level of the OTAFL assessment. Students who completed only the first or second levels of the OTAFL (n=15) were compared to students who completed all three levels (n=15). Students in the higher skill book levels were more likely than the lower-level students to complete the third OTAFL level (χ_1^2=3.39, p=0.07). Thus, mainly higher-level students (as defined by skill book level) completed the higher levels of the OTAFL, which may account for the unexpectedly high average OTAFL scores for all three levels.

Students at the higher skill book levels also scored higher than lower-level students on each of the three OTAFL levels (22 vs. 19, 27 vs. 23, and 39 vs. 34, respectively), and these differences were marginally significant for levels A (p=0.06) and C (p=0.1) when compared using a Wilcoxon two-sample test.

Table 2 illustrates student LES confidence scores by skill book level. Laubach students at the higher skill book levels also reported higher average total confidence scores on the LES compared to students at lower levels (67 vs. 47), and this difference was marginally significant when compared using a Wilcoxon two-sample test (p=0.07).

Table 2
LES Confidence Scores for Laubach Students by Skill Book Level

Self-efficacy subscale	Skill Books 1 & 2 mean (SD)	Books 3, 4 & Challenger mean (SD)
General literacy	66 (15.3)	75 (11.7)
Commercially related items	58 (26.7)	69 (30.0)
Transportation	45 (29.5)	64 (27.0)
Access to services	40 (27.5)	58 (30 .0)
Work/school-related	30 (27.2)	50 (27.3)
Family/social-related	35 (28.9)	65 (30.9)
Total	47 (25.9)	67 (26.1)

Length of time in the program was used as a further indicator of progress for both Laubach and Core students. In this study, students were divided into two groups based on whether they had spent less than one year in their program (n=13) or had been participating in the program for one year or more (n=23) (two students had missing data). Surprisingly, overall average confidence scores were higher for students who had been in the program for less than one year (71 vs. 54). The LES scores of these two groups were compared using a Wilcoxon two-sample test; however, the difference was not significant. Average OTAFL scores for the two groups were similar, and no significant differences were found.

Formal education was another indicator related to performance. Students were divided into two groups: those who had less than a grade 9 education (n=18) and those who had completed at least grade 9 (n=17) (three students had missing data). The group with the higher level of education rated their overall confidence higher than students with a lower level of education (67 vs. 51), and this difference was marginally significant (p=0.08). The OTAFL scores of the two groups did not differ significantly, although the average scores of students with at least a grade 9 education were higher on each of the three levels.

Finally, gender differences were explored as a possible indicator of performance. Overall, LES confidence scores were significantly higher for female students (z=2.21, p=.03). A greater percentage of females (71%) than males (41%) completed all three levels of the OTAFL, and females performed significantly better on levels A (z=1.86, p=.06) and B (z=1.92, p=.06). No gender differences emerged with respect to program level (for Laubach students), length of time in the program, or level of formal education.

Correspondence in Enactive and Affective Performance

Students were divided into two groups depending on whether they had completed all or some of the OTAFL levels (sample sizes were too small to look at completion of each OTAFL level separately). A Wilcoxon two-sample test was used to compare the groups' average confidence scores. Average confidence scores were higher for the group that had completed all OTAFL levels (69 vs. 52), but the difference was not significant.

As described previously, 28 of the 45 self-efficacy items corresponded to one or more of the 100 OTAFL questions (for example, the self-

efficacy task "to read a road map" corresponded to two questions in level A of the OTAFL test). Students' average confidence scores across these 28 self-efficacy items were calculated, and then students were split into two groups at the median score of 73 (only students who had rated their confidence for all 28 items, and who had completed all OTAFL tasks, were included, n=11). The OTAFL scores of the two groups were compared using a Wilcoxon two-sample test. Students with average confidence scores above the median scored significantly better (p=.05) on the OTAFL (mean 93 of 100 correct) than students with confidence scores below the median (mean 73 items correct). The same relationship was found between confidence scores and task frequency (see Figure 1). Students whose task frequency scores were above the median rated their confidence significantly higher than those with lower task frequency scores (p=0.002). Reported task frequency and performance on the corresponding OTAFL questions were also significantly related (p=0.05).
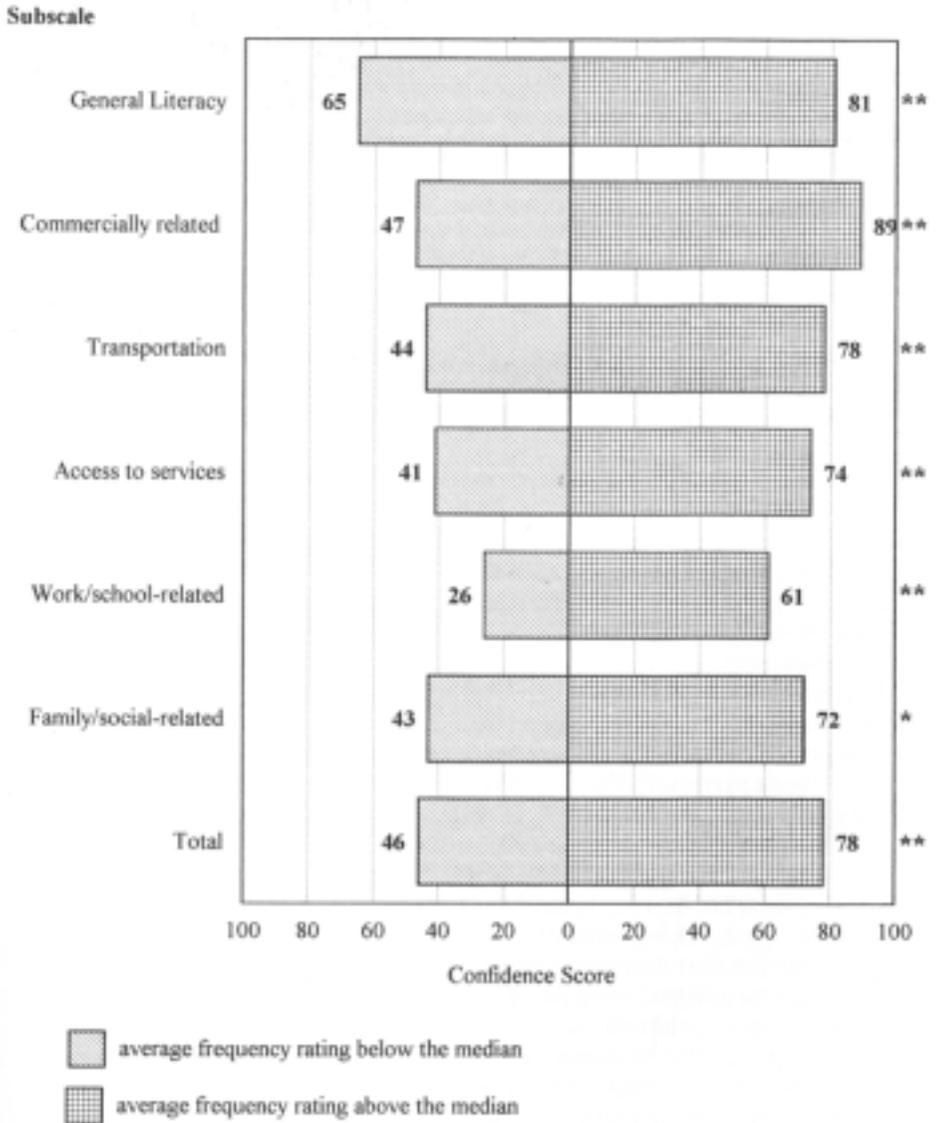
Acceptability Ratings

As part of the assessment package, tutors and students were asked to rate the acceptability of the instruments on a scale from 0 (not acceptable) to 5 (very acceptable). Eighty-eight percent of Laubach tutors and 77% of students rated the Laubach check-up exercises as "very acceptable" (a rating of 4 or 5). Tutors and students were also very receptive to the OTAFL: 87% of tutors and 74% of students rated this assessment as "very acceptable." The LES received somewhat lower acceptability ratings: 79% of tutors and 55% of students rated the LES as "very acceptable." A further 21% of tutors and 37% of students rated this tool as "acceptable," and only two students indicated that it was "not acceptable" (rating of 0 or 1).

We requested that tutors record actual time spent on completion of each component in the battery. Average time to complete the Laubach check-ups ranged from 26 to 58 minutes for Skill Book 1 and 4 exercises, respectively. Completion of the OTAFL averaged 30 minutes for level A (range 7–60) and 54 minutes for level C (range 15–90). The average time for students to complete the LES was 32 minutes (range 16–45), and tutors took an average of 17 minutes (range 10–30).

DISCUSSION

Characteristics of our sample closely matched those reported in a national survey of Laubach students and tutors (Sullivan et al.,

Figure 1
LES Confidence Scores by Task Frequency

Subscale

Subscale	Below median	Above median	Sig.
General Literacy	65	81	**
Commercially related	47	89	**
Transportation	44	78	**
Access to services	41	74	**
Work/school-related	26	61	**
Family/social-related	43	72	*
Total	46	78	**

100 80 60 40 20 0 20 40 60 80 100

Confidence Score

☐ average frequency rating below the median

▦ average frequency rating above the median

* 0.05 < p < 0.1. ** p < 0.05.

1990). The only striking difference was the proportion of students who had been in the program for more than one year: 30% in the national survey as compared with 68% in this study. This difference may simply be a result of our small sample; however, more than likely it is due to the extent of participation requested (the national survey involved only completion of a short satisfaction questionnaire). It is possible that tutors who had been working with their students longer were more willing to relinquish some of their tutoring sessions. It is also possible that the literacy council administrators had deliberately included more "experienced" tutor/student pairs in the lists of tutor names provided. A random sample of tutor/student pairs was requested (as in the survey by Sullivan), but two councils insisted on providing the names of selected student/tutor pairs that the administrator thought were "most likely to participate." It is probable that these tutors were viewed as more committed to the program or as better tutors, especially because program administrators needed constant reassurance from the research team that it was the *instruments* that were being evaluated, and *not* their program.

All of the instruments used in this project were rated as "acceptable" or "very acceptable" by over 80% of tutors and students surveyed. Not surprisingly, the Laubach program's own check-ups received the highest acceptability ratings from tutors and students who were familiar with them. Such skill-based assessments appear useful for quantifying student reading abilities and could be useful for initial placement purposes. It was clear, however, that Laubach tutors were very reluctant to administer these assessments prematurely (i.e., before their student had covered the corresponding skill book material). Tutors' concern that unsuccessful performance may undermine student self-confidence is supported by Bandura's (1982) efficacy theory.

The OTAFL (new to most) was also rated as acceptable by the large majority of respondents. Jones and Librande (1987) similarly reported that literacy students found the OTAFL enjoyable and many did better that they thought they would on this test. OTAFL scores were able to distinguish between Laubach students at different skill book levels. However, our data suggested a possible ceiling effect (i.e., higher-level students performed very well on the highest level of the test, even though they were still enrolled in a literacy program). Interestingly, a number of tutors in our focus groups predicted that their students might "ace" the OTAFL. The findings also substantiated a connection between frequency of routinely doing an

activity (e.g., using public transportation) and successful performance on enactive measures (e.g., reading a bus schedule). Unfortunately, it was not possible to determine whether literacy skills or other strategies (such as recognition or matching) were used to complete the tasks.

The newly developed LES was found to be internally consistent and contained a suitable mix of item difficulty for this population. Although space was provided at the end of each subscale (general literacy, work/school related, access to services, etc.) to add items, few tutors or students provided additions, indicating that the scale appears to have good content validity. The LES was also rated favorably by the majority of tutors and students, although less so than the other instruments. This may have been due to several factors: the LES is very different from typical adult literacy assessments that tutors and students may be more familiar with (such as phonics and other reading skill tests); respondents may have been exasperated by having so many items (one tutor commented that she felt unable to rate her student on many of the tasks); or, as one tutor commented, the three different rating scales created some confusion for both tutors and students.

Two of the expected indicators of performance—level in the literacy program and level of formal education—were related to scores on the enactive and affective measures in the expected direction. That is, students working on more advanced Laubach skill books and those with higher levels of education had higher levels of literacy-related self-confidence and scored better on the OTAFL. Confidence scores, OTAFL performance, and reported frequency of task performance were related to one another. These findings are consistent with Bandura's (1982) theory that there is a reciprocal relationship between self-efficacy, probability of attempting a given task, performance mastery, and enhanced self-efficacy.

Surprisingly, length of time in the program was not a good indicator of performance. No significant differences were found for either confidence ratings or OTAFL scores based on length of time in the Laubach or Core program. The tendency was the opposite of what might be expected: students who had been in the program for *less time* had *higher* confidence scores. Many of the tutors indicated that lesson times are often missed owing to holidays or competing demands (e.g., job or family) or are taken up discussing problems and current events. Thus, duration in the program does not necessarily translate into more hours of actual teaching time. It is also possible

that students with low confidence and/or functional abilities may be reluctant to leave the program and the help they receive from their tutor.

An interesting finding was that female students had significantly higher LES scores, were significantly more likely to complete all levels of the OTAFL, and scored significantly better on two of the levels. Gender differences could not be accounted for by level of formal education or level in the program (generally good predictors of performance on these two measures). Length of time in the program was also similar for the female and male students in our sample. Bandura's (1982) theory postulates that role modeling and performance feedback from role models with similar characteristics enhance efficacy expectations. Operation of literacy programs tends to be female dominated; all four of our program administrators and 71% of the tutors in our sample were women. All female students were matched with female tutors. While 11 of the male students were matched with male tutors, the apparent shortage of male tutors meant that 12 of the males in our sample had female tutors. It is possible that males may be less likely to admit to the stigma of illiteracy and approach literacy programs; and those who do may have exceptionally low literacy abilities. Until literacy programs begin routine tracking of their students, however, such speculations (and others, such as the characteristics of drop-outs) will be neither confirmed nor disconfirmed.

The present study is limited by a small sample size that precluded several analyses of interest. Generalizability is restricted by our having examined only two literacy programs (although the selected programs represent opposite ends of the spectrum in terms of curriculum structure). The fact that the various components in the battery were sometimes completed weeks, and even months, apart may have underestimated the extent of correspondence between the scores. Despite the aforementioned limitations, over the year and a half devoted to this project we learned a great deal about the challenges inherent in evaluating literacy programs, and we can offer concrete recommendations to the field.

CONCLUSION

An ideal evaluation protocol would entail comprehensive assessment of cognitive, enactive, and affective components of literacy (Finlay & Harrison, 1992)—at program entry and periodically to examine

student progress. Unfortunately, such a protocol is problematic on several counts. Because new students may be particularly apprehensive and therefore not perform to their full potential, a "settling-in" period prior to assessment has been advocated (Sticht, 1990). A counterargument is that delayed assessment may fail to capture early gains. Self-efficacy, in particular, has been found to increase dramatically following initial program participation (Bandura, 1982).

A comprehensive assessment strategy also requires several sessions for completion—only 13 of our 38 pairs completed the entire battery within a two-to-six-week time frame. Not surprisingly, the extensive time committment was the principal complaint of the tutors. Because volunteer tutors and students arrange their own meetings at mutually convenient times (although the program recommends at least weekly sessions), we found weeks often elapsed between sessions. This has important implications for both program delivery and evaluation. Sacrificing lesson time for assessment time is one issue. Another is the utility of data not controlled from the perspective of program juncture relative to assessment.

Of the three recommended components of literacy assessments (Finlay & Harrison, 1992), cognitive tests were the most problematic. Our tutors rejected the notion of using any of the currently available standardized reading tests for reasons similar to those cited in the literature (Sticht, 1990). Judging by our experience with the Laubach check-ups, tests of actual reading abilities are acceptable when based on the program curriculum, when administered in the familiarity of the established tutor-student relationship, and only when student mastery of content appears assured and confidence will not be undermined. The label itself is important—our focus groups strongly favored using "check-ups" or "progress assessments" in contrast to the word "test." Given the variability of current literacy training approaches, the onus is on a central body such as the Ministry of Education, to define a set of essential, basic skills (for reading, as well as numeracy and other domains of literacy) that would constitute a minimum curriculum for all literacy programs. Only then can a standardized cognitive instrument be developed to measure student progress within literacy programs, as well as to compare relative outcomes between programs.

In contrast, the enactive and affective assessment tools—the OTAFL and the newly developed LES—were acceptable to tutors of both literacy programs, suggesting the possibility of widespread adoption in the field for other basic literacy training programs (whether

curriculum based or not). Indicators of functional literacy and literacy-related self-confidence appear to be relevant and interesting to literacy students and tutors alike. Moreover, both tools were feasible for tutor-based administration. However, time required for administration—about 30 minutes for each level of the OTAFL and the complete LES—must be addressed. Jones and Librande (1987) reported that the concentration of literacy students begins to wane after about half an hour.

The high internal consistency ratings do not suggest that the 45-item LES be shortened. However, the correspondence between confidence and frequency ratings indicates that the latter can be dropped. We recommend that both confidence and perceived importance ratings (each requiring about 10 minutes) on the LES be completed by all literacy students at baseline, as part of their intake interview. Literacy task importance ratings are a more structured mechanism for obtaining student outcome expectancies (or personal goals) and a means of providing direction for tutors.

We further recommend that that the OTAFL not be administered until after a settling-in period. Mastery of each OTAFL level should dictate the timing of admistration of subsequent levels (until a ceiling effect occurs for the final level, C). Continued, periodic administration of LES confidence ratings (requiring only about 10 minutes) is warranted beyond the OTAFL given preliminary indications of relatively high difficulty on at least 7 of the items, and the fact that self-confidence should continue to increase with literacy progress.

We recommend that student assessment procedures be incorporated into tutor training workshops to facilitate ongoing evaluation. We further recommend that literacy programs gather background information on registrants (especially gender and highest level of formal education) and establish tracking mechanisms. Speculations are that many students remain in literacy programs for less than 18 months (Finlay & Harrison, 1992).

This study has provided valuable baseline data for adult literacy students who have been in their program for at least three months, a time frame in which students would likely have become more comfortable in their tutoring sessions. The baseline information has provided some evidence of the acceptability, internal consistency, and convergent and discriminative properties of the chosen instruments. A current follow-up study of these students will provide informa-

tion about the responsiveness to change of the LES. A national survey is being planned to address issues of recruitment and retention.

REFERENCES

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist, 37*(2), 122–147.

Breen, M.J. (1992). Writing for your audience. *Health Promotion, 31*(1), 2–6.

Creative Research Group. (1987). *Literacy in Canada: A research report*. Ottawa: Southam News.

Fingeret, H.A., Ton, A., Dyer, P., Morley, A., Dawson, J., Harper, L., Lee, D., McCue, M., & Niks, M. (1994). *Lives of change: An ethnographic evaluation of two learner-centred literacy programs*. Vancouver.

Finlay, A., & Harrison, C. (1992). Measuring "success" in reading in adult basic education: A United Kingdom perspective. *Journal of Reading, 36*(3), 164–172.

Gorman, T.P. (1981). A survey of attainment and progress of learners in adult literacy schemes. *Educational Research, 23*(3), 190–198.

Jackson, G.B. (1990). *Measures for adult literacy programs*. American Institutes for Research, Washington, DC: Association for Community Based Education.

Jones, S. (1989). Tests for adult basic education and adult literacy. In M.C. Taylor & J.A. Draper (Eds.), *Adult literacy perspectives* (pp. 217–226). Toronto: Culture Concepts. (ERIC Document No. 333117)

Jones, S., & Librande, L. (1987). *Ontario Test of Adult Functional Literacy*. Ottawa: Carleton University Centre for the Study of Adult Literacy.

Karlsen, B., & Gardner, E. (1986). *Adult basic learning examination*. San Antonio, TX: Psychological Corporation.

Kirsch, I. (1985). *NAEP profiles of literacy: An assessment of young adults*. Princeton, NJ: National Assessment of Educational Progress.

Librande, L., & Jones, S. (1988). *Ontario Test of Adult Functional Literacy: Administration guide*. Ottawa: Carleton University Centre for the Study of Adult Literacy.

Lichtman, M. (1974). The development and validation of R/EAL, an instrument to assess functional literacy. *Journal of Reading Behavior, 1*(2), 167–182.

Metz, E. (1989). *The issue: Adult literacy assessment.*. ERIC Clearinghouse on Reading and Communication Skills. (ERIC Document No. 332165)

Meyer, V., & Keefe, D. (1988). The Laubach way to reading: A review. *Lifelong Learning, 12*(1), 8–10.

Montigny, G., Kelly, K., & Jones, S. (1991). *Adult literacy in Canada: Results of a national study*. Ottawa: Statistics Canada.

Myers, A.M., & Gonda, G. (1987). Evaluation principles in practice. *Canadian Journal of Program Evaluation, 2*, 23–35.

Ontario Public Health Association. (1989). *Literacy and health project: Phase one*. Ottawa: Author.

Open Learning Network. (1994). *Literacy Link Eastern Ontario: Recognition for Learning pilot project*. Kingston, ON: Author.

Padak, N.D., & Padak, G.M. (1991). What works: Adult literacy program evaluation. *Journal of Reading, 34*(5), 374–379.

Perrin, B. (1989). Literacy and health: Making the connection. *Health Promotion, 28*(1), 2–5.

Psychological Corporation. (1986). *Canadian adult achievement test*. Toronto: Harcourt Brace Jovanovich.

Selman, M. (1991). *Adult ESL program evaluation guide*. Vancouver: Ministry of Advanced Education, Training, & Technology.

Shell, D.F., & Murphy, C.C. (1989). Self-efficacy and outcome expectancy mechanisms in reading and writing achievement. *Journal of Educational Psychology, 81*(1), 91–100.

Sticht, T.G. (1990). *Testing and assessment in adult basic education and English as a second language*. San Diego, CA: Applied Behavioral and Cognitive Sciences. (ERIC Document No. 317867)

Strecher, V.J. (1986). The role of self-efficacy in achieving health behavior change. *Health Education Quarterly, 13*, 73–91.

Sullivan, A.M., White, S., & Batstone, R. (1990). *A survey of learners and tutors*. St. John's, NF: Laubach Literacy Canada.

Thomas, A.M. (1991). *Adult literacy volunteer tutor program evaluation kit*. Vancouver: Ministry of Advanced Education, Training, & Technology.