

# THE EVOLUTION OF LITERACY THROUGH LARGE-SCALE ASSESSMENT

TREVOR J. GAMBELL *University of Saskatchewan*  
DARRYL M. HUNTER *Saskatchewan Department of Education*

**ABSTRACT.** In this article, we explore recent conceptions of literacy through the differing lenses of three large-scale literacy assessments conducted in 1994: a provincial learning assessment in Canada, a national reading and writing assessment, and an international literacy study. The varying purposes, designs, and assumptions of these three literacy assessments are revealed through juxtaposition. Their unique features and their findings are described to illustrate new definitions of literacy, new directions in the measurement of literacy, and potential for literacy curriculum and instructional change. We find that measurement techniques are consistent with best instructional practice, that literacy now encompasses critical, ethical, and aesthetic dimensions, and that assessment results do not support the notion of a literacy crisis in Canada.

**RÉSUMÉ.** Dans cet article, nous analysons les récentes conceptions de l'alphabétisation au travers des lentilles divergentes de trois évaluations à grande échelle menées en 1994: une évaluation provinciale de l'apprentissage au Canada, une évaluation nationale de la lecture et de l'écriture et une étude internationale sur l'alphabétisation. La diversité des buts, des objectifs et des hypothèses de ces trois évaluations est révélée par juxtaposition. Les auteurs décrivent leurs caractéristiques propres et leurs résultats pour illustrer les nouvelles définitions de l'alphabétisation, les nouvelles orientations de la mesure de ce concept et les possibilités de modifier les programmes d'études à cette fin. Nous constatons que les techniques d'évaluation sont conformes aux meilleures pratiques pédagogiques, que l'alphabétisation englobe désormais des paramètres critiques, éthiques et esthétiques et que les résultats de ces évaluations démentent la notion d'une crise à ce sujet au Canada.

What we measure, it is often remarked, is what we deem important. Thus, it is not surprising that the demands for public accountability and program improvement have stimulated a number of provinces, individually and in concert, to establish large-scale assessment programs to garner information about the literacy skills of public school students as an indicator of educational health. The most prominent, and the first,

national literacy assessment in schools was in 1994, part of the Council of Ministers of Education, Canada's (CMEC's) School Achievement Indicators Program (SAIP). Simultaneously, Saskatchewan Education introduced a Learning Assessment Program (LAP) to monitor provincial educational outcomes. Both programs demonstrate the priority public officials have accorded to literacy as a fundamental skill "for a global economy and lifelong learning" (CMEC, 1995a, p.5). Recognizing that assessments conducted by ministries of education do not provide a comprehensive picture of literacy for an entire population, Statistics Canada (the federal crown corporation for statistical monitoring) cooperated in an International Adult Literacy Survey (IALS) in 1994 as part of a multi-country and multi-language assessment in seven comparable countries.

Taken together, the 1994 assessments paint a revealing, three-dimensional portrait not only of achievement in reading and writing, but also changing definitions and directions in literacy education in Canada. In this paper, we compare and contrast the test designs, procedures and findings of these three large-scale assessments. We plumb underlying conceptions of and assumptions about literacy in Canada. Through examining the definitions adopted by and tenets held for each program, we posit an evolution of the notion of literacy in an educational context.

#### THE SASKATCHEWAN LEARNING ASSESSMENT PROGRAM

Concerns emanating from provincial curriculum reform and instructional change prompted Saskatchewan educational authorities to develop their own provincial language arts assessment program (Saskatchewan Education, 1996) in 1994, rather than joining with their counterparts in the national SAIP. Alone among the provinces and territories at the time, Saskatchewan abstained because public officials were initially apprehensive that national testing might tarnish provincial curricular definitions of literacy. Literacy specialists conceived of literacy as socially constructed, as a process domain of the language arts, and as maturationally cultivated through emerging, developing, extending and specialised phases across Grades 1-12 (Saskatchewan Education, 1989). The province was in the midst of implementing an English language arts curriculum using that definition in its schools, as one means for achieving a set of Goals for Education in Saskatchewan derived from a public review process in the 1980s (Saskatchewan Education, Training and Employment, 1984). To ensure that assess-

ment procedures were compatible with the new literacy orientations, the Saskatchewan LAP architects explicitly designed an assessment to match blueprints somewhat different than those that underlay other large-scale assessments at the time.

Ostensibly, the purposes of the 1994 Saskatchewan LAP in Reading and Writing were not radically different from others: to collect baseline data for longitudinally tracking mainstream student progress; to describe how well students have acquired basic and higher order skills in reading, writing, and thinking; to provide a picture of strengths and weaknesses in reading and writing; to collect information about the home, classroom, and out-of-school environment in which language learning takes place; and to set performance standards or expectations for language arts achievement.

For the LAP, statisticians drew a stratified random sample of approximately 3400 students from the classrooms of Grade 5, 8, and 11 students, so as to represent the diverse urban, rural and northern school situations in the province. Just over half of the province's children attend schools in rural, mixed-farming areas and in remote boreal areas. A total of 160 classrooms in 141 schools received a 'lottery letter'. Only students who were enrolled in regular language arts programs participated; however, students from French-designated schools took part in this English literacy assessment. When the provincial tests were administered, students did either a reading or writing assessment, not both, for reasons of time and economy. Each grade level reading assessment consisted of two parts: *Read On 1!* contained four literary passages such as poems, short stories, and excerpts from plays or novels. *Read On 2!* featured four expository passages such as newspaper articles and excerpts from factual reports and essays. The single writing assessment form, called *Write On!*, collected data on students' writing skills and strategies. Students' writing efforts were divided into three sessions. In the first or prewriting session, students pored over a variety of passages based on a theme – Grade 11 was environmental concerns, Grade 8 was interpersonal communication, and Grade 5 was (outer) space – to spark ideas for writing. Teachers directed students to write in either the expressive or expository mode, and students self-selected one of four genres within that mode. In the second session, students completed a first draft of their writing. They also revised and edited another piece of writing provided in the assessment form. In the third or postwriting session, students toiled over the final draft of their writing.

The Saskatchewan test designers printed reading and writing tests in colour, tabloid newspaper format with graphics to make them more appealing. After students completed the assessment items, they checked off a 35-item closed-response questionnaire to indicate their reading and writing practices, preferences, and attitudes. Teachers too completed a questionnaire that inquired about the classroom, school, and community context for language arts instruction, and instructional techniques associated with cultivating literacy skills.

In the summer of 1994, contracted teachers trained for and scored student responses using pre-established criteria in five performance levels. Level 1 represented low performance; Level 5 represented highest performance. To assist with analysis and interpretation of results, a standards committee of educators and noneducator stakeholders met to formally set expectations for students, using a modification of the Angoff method (Angoff, 1971), an American judgmental process for establishing minimal test competencies. As such, the project both described actual student outcomes but also set socially-constructed targets for what they should be in the future, by directly engaging the public in an accounting of provincial literacy skills and attitudes. Thus, public officials interpreted test numbers in terms of the expectations established by the standards committee, in the absence of a national and international comparator.

Overall, the writing results showed half the students at all three grade levels performing in the middle of the five-point scale (Saskatchewan Education, 1996). These students produced a competent, but not particularly imaginative piece of writing. The largest proportion of Saskatchewan students at each of Grades 5, 8, and 11 organized their writing in a clear but mechanical way, although fewer students than expected achieved at the upper levels. Students in all three grades wrote uncomplicated sentences well, but were less effective writing complex sentences. The majority of students (more than expected) used vocabulary at the mid-range performance level. Their ideas and expressions were clear and adequate if somewhat unimaginative. More youth at all three grade levels should have been able to express ideas in the complex and imaginative ways typical of top performance levels.

At all three grade levels, students' reading comprehension or ability to gather meaning from what they read, generally exceeded expectations. Although fewer Grade 11 students than expected achieved high performance levels in this skill, overall, students demonstrated very strong skills in reading comprehension. Students were also assessed in higher

order thinking skills appropriate to their ages: Grade 5 students proved weak in making predictions from reading; Grade 8 students were proficient in clarifying and extending their thinking; Grade 11 performance in interpreting or evaluating when reading was notably short of expectations. In particular, Grade 11 students had difficulty defending their interpretations and evaluations about a piece of writing.

Grade 5 and 8 students demonstrated proficiency at identifying the author's purpose for writing a particular passage, and in recognizing that the author or other readers might have points of view different from their own. Grade 11 students were far below expectations in supporting their judgments about a reading passage with specific evidence from the text. Grade 5 students' ability to read critically, and to judge ideas, events, and characterizations was near expectation. Grade 8 students, on the other hand, were proficient at reading critically for a variety of purposes. Grade 11 students were far below expectations at evaluating literature using literary criteria.

The Saskatchewan LAP pioneered in its collection and measurement of information about both the processes as well as products of writing. Studies of writing process examine the stages or phases a writer goes through while composing a text, from conception to final copy. Writing processes vary among students, but typically stages or phases have been identified that include prewriting, drafting, revising, redrafting, editing, proofreading, and final copy/publishing. Literacy specialists hypothesize that a more complex and complete writing process produces a better piece of writing than a 'first draft equals final copy' process. In the Saskatchewan LAP, evidence of students' writing processes was available through their portfolios: teachers invited students to submit what they considered to be their best writing along with all notes and previous drafts.

To categorize the myriad writing processes used by students, Saskatchewan assessors initially hypothesized a number of writing process models and used them to analyze and classify students' writing strategies. Other models were added as they appeared in the data. Teachers scored the final drafts of the classroom samples using the same five-point scale and procedures as with the pencil-and-paper assessment forms. In addition, teacher-coders assigned genre and mode codes to the portfolio pieces. Table 1 shows the writing process models data.

When looking at the classroom writing process models used, the most striking finding was the "Unknown" category: a large number of stu-

**TABLE 1.** Classroom Writing Processes:  
1994 Saskatchewan Language Arts Learning Assessment

| Writing Process Model   | Percent of Grade Level Total |         |          | Average Score Achieved Level 1-5 | Confidence Interval for Mean |
|---|------------------------------|---------|----------|----------------------------------|------------------------------|
|   | Grade 5                      | Grade 8 | Grade 11 | Total Students Tested            |                              |
| Model A Unknown   | 17.0%                        | 41.1%   | 37.8%    | 2.90                             | 2.8 – 3.0                    |
| Model B First Draft = Final Draft                             | 12.0%                        | 7.6%    | 12.5%    | 2.68                             | 2.5 – 2.9                    |
| Model C First Draft->Editing->Final Draft                     | 18.5%                        | 6.8%    | 10.4%    | 2.91                             | 2.8 – 3.0                    |
| Model D* First Draft-> Revision->Editing-> Final Draft        | 18.9%                        | 20.1%   | 10.4%    | 3.00                             | 2.9 – 3.1                    |
| Model E Prewriting-> First Draft-> Final Draft                | 9.9%                         | 2.3%    | 8.8%     | 2.85                             | 2.7 – 3.1                    |
| Model F* Prewriting-> Drafting->Editing-> Final Draft         | 12.7%                        | 4.9%    | 7.0%     | 3.12                             | 3.0 – 3.3                    |
| Model G Prewriting->Drafting->Revision->Editing-> Final Draft | 10.9%                        | 17.2%   | 13.0%    | 3.13                             | 3.0 – 3.2                    |
| Personal/Artistic   | 76.4%                        | 70.4%   | 65.1%    | 3.01                             |                              |
| Functional/Explanatory  | 23.6%                        | 29.6%   | 34.9%    | 2.76                             |                              |

\*Models which are significantly different from Model B using Student-Newman-Keuls test, F-ratio 4.721, *p*:.05

dents did not submit enough material such that a writing process could be determined. Those students submitting an entire writing process sample displayed a wide variety of writing strategies. Without doubt, the most interesting finding was that the more stages followed in a writing process, the higher the quality of the finished product. Those students who engaged in more complex writing processes involving prewriting, drafting, revising, and redrafting scored significantly higher than those who employed a simpler process. Students generally considered that their best writing was for personal or artistic purposes, probably because they had control over the topic, purpose, and voice of such pieces. Students did show increased confidence in the quality of their functional or explanatory writing across Grades 5, 8 and 11. Over all three grades, students' personal-artistic writing scored higher than their functional-explanatory pieces.

Other symmetries were evident in Saskatchewan LAP results. The largest proportion of Saskatchewan youth at all three grade levels characterized themselves as good readers, but as average writers. Students' self-concept as readers improved as they progressed through the

school system. The numbers of students who classified themselves as average or good readers increased significantly between Grades 5 and 11. By high school, students' self-concept may be greater than their critical and evaluative reading ability actually warrants. But in writing, the opposite pattern was evident. The numbers of students who saw themselves as good or very good writers declined between Grade 5 and Grade 11. However, achievement scores suggest that the quality of their writing actually improved. That a provincial assessment would search for relationships between self-esteem and performance indicates a broad definition of literacy that encompasses affective elements.

#### **COMPARISON OF THE SASKATCHEWAN LAP AND NATIONAL SAIP LITERACY ASSESSMENTS**

Whereas the Saskatchewan LAP tracked progress toward provincial literacy objectives and goals, the national SAIP (CMEC, 1995a) set out to ascertain growth or change between two different age groups of students, Canada-wide. Both 13- and 16-year-old students completed identical instruments to determine if there were differences in performance, assuming that changes would indicate whether literacy skills were being fostered by schools in the intervening years. On the other hand, the Saskatchewan LAP designers sampled according to grade rather than age, with different instruments at each of Grades 5, 8, and 11; these grades were chosen because they were related to the mid- or end-points of the three bands in Saskatchewan's new elementary, middle years and secondary curricula. Cross-grade comparisons could not therefore be made because test items and tasks varied according to those developmental- or grade-specific aims of language arts programs. As such, Saskatchewan measurement specialists assumed that literacy skills are specific to the developmental and maturational level of students, whereas the national SAIP designers assumed that 13- and 16-year-old literacy skills differ only in degree. The same test instrument was given to both groups.

The Saskatchewan LAP and the national SAIP displayed different orientations in accommodating linguistic diversity, not surprising given the relative demographic proportions of minority populations within Saskatchewan and within Canada as a whole. Designing an instrument and criteria sensitive to the large and growing Aboriginal population was a priority in Saskatchewan, whereas creating an equitable assessment process for French- and English-language students was a priority for the SAIP creators. Nevertheless, both Saskatchewan assessors and

their SAIP colleagues presupposed that literacy is a broad societal goal for all populations, regardless of linguistic or ethnic background.

While the Saskatchewan LAP sampled by classroom, the SAIP sampled by both school and individual students to randomly select approximately 58,000 participants Canada-wide. Twenty-nine thousand 13-year-olds and an approximately equal number of 16-year-olds participated in eleven different geopolitical jurisdictions. Of this total, about 43,000 completed the reading and writing tasks in English, and about 15,000 completed them in French, not only in Quebec but also in Manitoba, Ontario, and New Brunswick.

A central difference between the provincial LAP and national SAIP lies in their reference points. The Saskatchewan LAP designers sought answers to questions about how well Saskatchewan students are achieving the objectives and aims of provincial curriculum, measured in relation to provincially-defined targets of performance. SAIP designers, on the other hand, set out to measure how well school systems are preparing students in reading and writing knowledge, abilities and skills between the ages of 13 and 16 using interprovincial comparisons. The SAIP was not directly curriculum-referenced: policy makers couldn't use a specific curriculum blueprint for constructing the test instruments because of the differing provincial curricular approaches to language arts. "While the existing provincial curricula would be respected, the assessment would not be narrowly linked to the curriculum of any particular province" (CMEC Technical Report, 1995b, p.5). A more generic approach to the content and design of test instruments had to be adopted to accommodate the differing curricular orientations of the eleven different provinces and territories involved. The SAIP program thus staked its test validity on an interprovincial consensus exogenously derived from two years of "debate, discussion, consultation, revision and experimentation" (CMEC, 1995a, p. 92), including provincial reviews of reading and writing approaches. On the other hand, the Saskatchewan LAP assumed that test validity must be endogenously grounded in those provincial curriculum and instructional objectives which form the basis of curriculum guides now being implemented in Saskatchewan schools. Implicit in both testing programs was the notion that validity does not reside in the test instrument itself, but rather in the uses made of test scores (Messick, 1989).

Both the SAIP and the Saskatchewan LAP were criterion-referenced, and adopted five-point scales with descriptors of student performance



in reading and writing, rather than traditional percentages or letter grades. However, Saskatchewan assessors did not embed the levels directly in the reading assessment tasks. National SAIP evaluators, on the other hand, used sophistication or complexity of reading passages and difficulty of reading tasks as central gradients for gauging student performance. In contrast, the Saskatchewan LAP relied on quality of student response to reading passages and questions as the measure of student performance. The Saskatchewan LAP assessors designed reading tests such that the vast majority of students had an equal opportunity to respond to most reading passages and questions. In other words, the criteria were applied not to the design of the assessment but during scoring to students' constructed responses. What the student submitted as a response, not the complexity of the task, served as arbiter of student reading achievement.

As devices for categorization, the performance descriptors in a scoring rubric are key to understanding the notion of literacy being operationalized because the descriptors illustrate sorting principles. Both SAIP and Saskatchewan LAP evaluation specialists assumed that a student's reading fluency depends on his or her personal experience brought to the reading task, the student's language base (vocabulary and language strategies), and the complexity of the textual information. Similarly, both took as a premise that writing fluency depends on personal experience with written language, the degree to which the students' language base allows expression of ideas, and the complexity of the writing task. For the provincial testing program, five key criteria were used to sort student papers in reading: literal and inferential comprehension; predicting and interpreting; critical reading principles; making evaluations and judgments; and demonstrating cross-cultural understanding and appreciation. With the SAIP on the other hand, there was less emphasis on Bloom's taxonomy, North American educators' conventional rubric for categorizing sophistication of cognitive skills. SAIP measurement specialists focused rather on those textual features to which students responded, on quality of judgment, and on the level of personal insight and understanding demonstrated.

Classroom teacher-scorers used holistic or total impression marking techniques in mass scoring sessions for both the Saskatchewan and national assessments. In both exercises, teachers were recruited from across the expanse of the province or country to ensure that marking had "curricular validity". For both assessments, student forms were scored according to one of five levels from 1 (lowest) to 5 (highest),

proceeding from holistic to analytic scoring of the writing assessment, and from detailed scoring to overall impression scoring of the reading component. Carefully worded criteria for each performance level were used by scorers to place each paper in a level. Scorers relied on anchor or exemplar papers which characterized student work, and undertook prescoring training to develop consistency in scoring, creating an interpretive community of teacher-scorers. Papers were double-scored to maintain scoring consistency and triple-scored when scores diverged by more than one level. Scorers on the provincial assessment achieved a .82 level of reliability on a 10% subsample of the 1583 writing papers which were double-scored.

Both the Saskatchewan LAP and the national SAIP designers used similar approaches to determine student writing skills. Both provided students with opportunities to prewrite, to draft and to revise leading up to a finished text. Thus, both viewed the writing process as a series of integrated and recursive stages. However, for the SAIP, voice was deemed a significant part of the writing process, whereas the Saskatchewan LAP placed greater importance on genre, mode, and the anterior phases and episodes of the writing process. As for the Saskatchewan LAP, student participants in the SAIP selected and submitted a classroom sample of writing. In both cases, the portfolio piece served as a measure of concurrent validity for the pencil-and-paper test. But whereas the national SAIP collected only the final draft of a classroom sample, the Saskatchewan LAP asked students to submit all preliminary notes, drafts, outlines, and doodlings as a package. Coding and categorizing of this work would yield a measure of students' writing strategies. Thus, Saskatchewan assessors gleaned data on all parts of the students' writing to assess both writing process as well as quality of the final product. An accompanying questionnaire probed students about their writing strategies while producing the classroom writing.

Neither assessment was intended to be comparative or summative of student achievement within specific classrooms, across schools or school divisions, because they were "low-stakes assessments" in that marks did not affect students' transcript marks or grade promotion. Neither SAIP nor Saskatchewan LAP assessors forayed into special education classrooms, assuming that the validity of the assessment would be jeopardized since such students were following programs far different than the mainstream. The Saskatchewan LAP had wider categories for exemption than did the SAIP, and probably yielded more homogenous scores as a result. Both the SAIP and the Saskatchewan Learning Assessment

Program relied on random, representative sampling techniques which meant the students, teachers, and schools participated anonymously to create provincial and national profiles for monitoring systemic outcomes rather than individual student or school-level outcomes.

Like the Saskatchewan assessment, the national SAIP study found that Canadian students were competent readers and writers at the age levels tested, but not particularly creative writers or critical readers. Nearly four-fifths of Canadian students achieved at least a functional level of reading and writing skills. In terms of gender performance, the national SAIP findings for 13-year-olds and 16-year-olds were very similar to those of the Saskatchewan study: females significantly outshone their male counterparts in both reading and writing. Nationally, there tended to be closer achievement at the mid-performance range between 16-year-old males and females. Nevertheless, neither on the provincial nor the wider national scene was there a closing of the gender gap as one might expect. In both testing exercises, questionnaire data revealed that boys watched more television, did less language homework, and engaged in markedly less leisure reading than girls (Nagy, 1997; Gambell & Hunter, in press).

SAIP evaluators took as a cardinal assumption that all Canadian students could be compared fairly using an identical test instrument and tasks. However, recent test analyses reveal that males and females responded differently to some test questions on the SAIP study. At issue is whether the test was biased or whether students' overall response patterns vary according to gender (CMEC Technical Report, 1995b; Froese, 1998). Another SAIP premise is that literacy could be measured equitably and translinguistically – in either English or French – with a translated instrument. However, Nagy (1997) and indeed even the SAIP's sponsors (CMEC, 1995a) suggest that comparability in complex testing across languages is very hard to attain. Neither the provincial nor national assessments took as a premise that students would perform differently according to geography.

#### THE SASKATCHEWAN, CANADIAN AND INTERNATIONAL ASSESSMENTS COMPARED

The Saskatchewan LAP compared elementary and secondary student performance against curricular expectations, and the SAIP revolved around interprovincial comparisons of adolescents' literacy skills. However, the primary objective of the International Adult Literacy Survey (IALS) (Statistics Canada, 1996) was to compare Canada's national

literacy profile with those of several key trading partners in the North Atlantic triangle. Whereas the SAIP and Saskatchewan LAP focused on school-based youth up to age 16, the IALS concentrated on adults aged 16 and older. Like the Saskatchewan LAP and SAIP, IALS was criterion-referenced and used a five-point scale of achievement. However, its validity derived less from curricular match or interprovincial consensus but rather on the psychometric properties of Item Response Theory, a sophisticated statistical model that uses actual response patterns to develop categories of achievement.

The SAIP designers did not develop a formal criterion-referenced standard, but rather portrayed the outcomes in terms of interprovincial comparisons and in relationship to the Canadian norm. Nevertheless, a standard or expected performance is inherent in any scale used to score student work or in a norm that might be used as a comparator. Therefore, those provincial ministry officials involved in designing the SAIP instruments embedded a standard in the SAIP scale, by predefining Level 2 as acceptable performance for 13-year-olds and Level 3 for 16-year-olds. On the other hand, IALS designers calibrated their assessment materials by doing statistical analyses of response patterns before psychometrically defining the performance levels. In that sense, the Saskatchewan LAP explicitly relied on the consensual judgments of stakeholders to define *a posteriori* provincial standards as a socially-constructed point of decision-making, whereas the national SAIP allowed ministry officials to bureaucratically define, *a priori*, an implicit standard. In contrast, IALS organizers assigned the task to measurement specialists to empirically set the standard.

Like the provincial and national assessments, the IALS adopted a 'continuum of skills' definition of literacy, disposing of the simplistic dichotomy of literate and nonliterate. In the IALS, literacy was defined as "the ability to understand and employ printed information in daily activities at home, at work and in the community, to achieve one's goals and to develop one's knowledge and potential" (Statistics Canada, Backgrounder, 1996, p.2). Since the goal was not to rank countries from most literate to the least, but to compare across cultures and languages, the survey sought to capture the social and economic characteristics that underlay observed literacy skills. Whereas Saskatchewan evaluators sought linkages between curriculum and instructional practices, and the SAIP evaluators sought linkages between literacy achievement and educational attainment, IALS designers sought to examine the social and economic impacts of different levels of literacy, the under-

lying factors which cause them, and how they might be amenable to policy intervention.

It is important to note that the IALS was restricted by its methods. Not a single piece of participant writing was collected or scored, undoubtedly because of the difficulties inherent in generating equitable comparisons across several languages. Rather, subjects were assessed directly in their home by trained interviewers (in itself an attempt to capture information in a realistic context); they were given tasks which involved manipulating prose, and using documentary and quantitative information. In other words, the IALS might best be called a sophisticated study of various reading skills, rather than an investigation of literacy, traditionally viewed as involving a measure of written production. As such, the IALS may misrepresent literacy. In contrast to the IALS which concentrated on functional reading tasks such as reading medicine labels, memoranda from the workplace, and newspaper weather forecasts, both the Saskatchewan LAP and the national SAIP employed texts which matched the types of reading materials found in school curricula and used for classroom instruction.

The IALS creators perceived a literate person not as someone who responds to text at various levels, but as an information processor. Its sorting principles gauged how well an individual could disembed information from a variety of consumer documents to perform lower-order and higher-order inferences. Like the SAIP, text difficulty or density and integration of skills were the primary gradients for measuring skill. But unlike the provincial and SAIP assessments, its notion of literacy encompassed quantitative information tasks; simple arithmetic operations and serial mathematical calculations with weather charts and actuarial tables were an explicit component that was measured on a scale separate from prose and document literacy. That may extend the concept of literacy beyond its traditional limits, even though the IALS authors argue that the underlying skills for processing numeric information are graduated like those for print information (Statistics Canada, 1996).

Although the primary purpose of the international adult literacy survey was to compare national Canadian literacy profiles with those in the United States and several European countries, it did yield literacy estimates for specific regional subpopulations across Canada, except the two territories. The IALS did include in-school youth aged 16 to 24, youth not in school full time, Ontario and New Brunswick residents whose mother tongue was French, seniors over age 65, and social assistant recipients and beneficiaries. None of the three assessments

systematically sampled persons living on Indian reserves, but surveyed those Aboriginal Canadians living off reserves; in all three cases, the numbers of Aboriginal participants were too small to permit separate analyses. As such, the architects of all three assessments assumed that Canadian literacy is a mainstream construction, coming out of a "standard" Canadian English (or French), that doesn't recognize dialects.

For both the national and international studies, designers presumed that literacy skills and definitions transcended mother tongue, and could be compared regardless of cultural differences. The IALS results demonstrated that both Canadian and the United States' literacy levels were approximately equal, but lower than those in Sweden. For the SAIP, French-speaking students' scores in reading at both age 13 and 16 were significantly higher than the Canadian norm. On the other hand, French-language participants in the SAIP writing assessment achieved below the overall Canadian results for both age groups. Geographically, francophone students everywhere outside Quebec did not perform as well as their franco-quebécois counterparts. Conversely, anglophone students in Quebec achieved results similar to their language peers across Canada.

Indeed, the international study showed that literacy attainments within Canada have patterns that are related to geography and to employment status. As one moved from west to east across the country on a regional basis from the Western provinces to Ontario to Quebec to the Atlantic provinces both adult and youth literacy levels declined. In Quebec and the three prairie provinces, the distribution of literacy skills along social class lines was considerably more equitable than in Ontario, British Columbia and the four Atlantic provinces (Willms, 1997); small sample sizes do not permit more detailed, provincial level analyses for the youth population. Not surprisingly, since testing repeatedly shows the connection between socio-economic status and achievement, the proportion of the unemployed population at each of the five literacy levels decreased as one moved up the five-point scale of literacy achievement. And parents and children in low-income households participated less frequently in literacy-enhancing activities, such as allocating a dedicated reading place or time in the home, than did those with a higher standard of living (Schalla & Schellenberg, 1998). In terms of Canadian youth, the literacy levels of 16-to-24 year olds were consistently higher than those of adults about to leave the labour market. As the authors succinctly concluded, "there appears to be little to support earlier predictions and present concern of a rapid erosion of either

educational quality or of the adult skills base." (Statistics Canada, Backgrounder, 1996, p.3)

#### THE EVOLUTION OF LITERACY THROUGH LARGE-SCALE ASSESSMENT

Literacy can be defined in a number of ways: as a conduit for the transmission of culture; as a medium of expression and thought; as a vehicle for communication; as a workplace skill for getting things done; or even as an instrument of national survival (Graff, 1987). Regardless of how they are viewed, reading and writing are undoubtedly important public goals. Ministries of education have for a long time used curricula as vehicles for propounding conceptions of literacy. However, with the proliferation of assessment programs across Canada, powerful instruments are available to public officials through which literacy may be defined as well as gauged.

#### *Paradigm shift*

The multiplying number of assessment programs illustrate a paradigm shift now underway among policy makers. This transition means the public education systems' effectiveness will be framed less in terms of inputs in the form of total operating grants distributed or capital equipment purchased, learning materials available, or pupil-teacher ratios, but increasingly in terms of outcomes. Whereas the curriculum guide as a form of input into the classroom for teaching and learning was once a ministry's primary tool for shaping reading and writing instruction, testing programs are becoming increasingly influential in exemplifying contemporary and emerging notions of literacy. Current assessment practices demonstrate a linkage with classroom practice, arm's length oversight concerned with promoting system learning, intensified concern with accountability, and respect for professional autonomy through process evaluation and anonymous sampling (Mawhinney, 1998). Traditional patterns of provincial control over educational reform agendas are changing.

A primary assumption in many assessment programs is that literacy is measurable using a linear scale. Instructional experts often suggest that capturing the complexity of recursive and dynamic literacy learning processes with simple, linear assessment tools is impossible. Literacy assessment is now attempting to recognize this complexity through the use of portfolio techniques to assess writing over time and across place and in a variety of dimensions, and through interactive and response-based reading tasks in a variety of genres and for different purposes. At

the same time, we might also ask whether the IALS definition of literacy confuses competence, which is what language permits an individual to potentially accomplish, with performance, which is what an individual actually does accomplish with language (Green, 1972). The best a scale can do is pinpoint performance at a particular point in time to create a still-life literacy portrait.

Second, literacy is not the mastery of a body of literature. As revealed in the three assessments, literacy has moved away from understanding and explicating a literary canon. Literacy notions in provincial and national assessments have also moved beyond simple, functional notions of dealing with expository texts useful to the world of citizenship and consumer economics (as exemplified in the international study), to one which permits a critical, aesthetic, and ethical engagement with texts. Literacy is not only doing things in day-to-day situations, but also critically analyzing events and ideas which are represented in a variety of texts.

A third assumption is that literacy assessment programs should model best practice for classroom practitioners. Whereas the resource package or the bibliography of materials may have once been central to instructional planning, a scoring rubric or description of performance levels on a scale, along with exemplary writings, are becoming important instructional tools for classroom teachers. Measurement techniques, like definitions of literacy, have evolved in sophistication; for example, portfolios are now used in several large-scale assessments. Assessments do shape educators' instructional behaviour (Wideen et al., 1997) but not always in deleterious ways. Alternative assessment practices can reinforce and validate new instructional techniques, such as process orientations to writing, higher order and critical appraisals of text, and metacognitive strategies for reading.

Of course, centrally-mandated curricula and assessment have long been tools of normative control: "how a society selects, classifies, distributes, transmits and evaluates the educational knowledge it considers to be public, reflects both the distribution of power and the principles of social control" (Bernstein, 1977, p.55). Tests have influenced instruction, usually in restrictive ways; secondary teachers in particular do admit that grade twelve provincial examinations in English language arts drive instruction and become a less-than-hidden curriculum (Anderson et al., 1990). "Low-stakes" random-sample assessments at other grade levels may have a similar effect on pedagogy and classroom practice, particularly as curricula edge towards an educational outcomes model (Reigeluth, 1997). With this pragmatic consideration, the challenge for assessors and educators is to make large-scale assessment



techniques model as closely as possible curricular and instructional approaches. Rather than being an appendix on or the antithesis to instruction (Barlow & Robertson, 1994) – a weapon of mass distraction from the core activities of schools – large-scale assessments can become an integral part of a cycle of improvement and even curriculum change.

### **Public school literacy crisis?**

What can we conclude about public education, literacy and its definition in Canada from these three assessment programs? First and foremost, the much-touted media and public concern about a possible deterioration in literacy outcomes from Canadian schools is unfounded. A Canadian media-manufactured crisis (Willinsky, 1990) has been effectively defused by recent large-scale assessments. In fact, the SAIP authors concluded: “The relatively high proportion of 16-year-olds at level 5 – a level that required students to demonstrate a very high calibre of skills – is also reason for education authorities to be pleased” (p.94). While this finding does not give cause for dismissing literacy concerns in education, it does lay to rest the myth of schools as failing to produce literate students. Indeed, the international study found that since a parallel 1989 study, “those leaving the labour force have been replaced by an incoming cohort of young people who are collectively much better educated and more literate” (Statistics Canada, Backgrounder, 1996, p.3).

It is only to be expected that with increasing numbers of students for whom English is a second or other language or dialect (13% of participants in the national assessment spoke a language in their home other than the one in which they were assessed), literacy achievement rates in the school population will be more widespread with increased numbers below the average range. In fact, the IALS found that the proportion of immigrants performing at the lowest performance levels was larger than those test-takers born in Canada, but that there were also proportionally more immigrants at top levels of performance than in the other six industrialized countries participating. Canada was unique among these countries in having such a large proportion of immigrants operating at the top performance level. This finding was attributed to the Canadian government’s policy of selecting skilled immigrants, and its acceptance of large numbers of immigrants on humanitarian grounds. Such conditions reflect changing demographics and immigration policies, and do not constitute a literacy problem, but rather an educational challenge of linguistic diversity. We must be careful not to confuse

social, economic, demographic and policy trends with educational or literacy standards.

### *Process methodologies*

The great variety in students' writing processes in the classroom writing samples can lead to ambiguous conclusions with respect to curriculum, instruction and data collection procedures. For example, we do not know whether writing processes are linked to writing genres, and we do not know the links between writing process knowledge, use, and classroom instruction. Likewise, if teachers and students are not metacognitively aware of writing process we wonder how reliable is data collection of writing process material from all phases of writing and thinking. But more important is the finding that the more steps students take or make in their process of writing a piece, the better the product as indicated by holistic scoring. This latter finding is important for it strongly suggests that writing process instruction can lead to superior writing. As such, it validates the inclusion of writing process methodology in new elementary, middle years and secondary language arts curricula. This important finding is a legacy of large-scale literacy assessment rather than elective educational research.

New large-scale assessment programs in Canada acknowledge that reading and writing are social constructions wherein students derive meaning from text or through the production of text as they interact within their various social, economic and cultural environments. Understandably, the IALS couldn't adopt a social-constructivist orientation to reading and writing given the intractable problems of comparing across several countries and languages. In the Saskatchewan LAP, reading was defined as a "cognitive process in which readers construct meaning from the text they read by connecting new knowledge to the knowledge they already possess" (Saskatchewan Education, 1996, p.18). The national SAIP also recognized that reading is far more than decoding written words but involves interpreting written material in light of personal experience. Personal experience and language base, including context and language strategies, are deemed to be important variables in literacy performance. Writing was seen in both made-in-Canada assessments as a contextualized communicative act that has a purpose, an audience, and a recursive process. Moreover, by employing holistic or general impression scoring techniques, public officials are recognizing that reading and writing encompass a broad and integrated range of abilities that can not be measured as discreet and isolated skills in

exclusively multiple-choice formats. Meaning does not reside in the text proffered for student reading, nor in the written text provided by the student for marker reading, but is found in the reader's response to both. Above all, the assessments assume that literacy is an integrated set of observable and measurable skills, that it constitutes more than knowledge and appreciation of literature, and that it is manifest in relationship to texts as diverse as newspaper articles, business letters and magazine excerpts.

In other words, the concept of literacy itself is evolving in Canada. Theoretical and technological advances have transformed literacy from reading-as-deciphering and writing-as-grammatical proficiency into a rich, coherent and sophisticated construct. Literacy now embraces both the ability to use and critique information in a daily consumer and home context, and the ability to interact with text to create personal meaning. Rather than viewing a person as being either literate or illiterate, all three assessments assume that there is a structured continuum of abilities and performances: criterion content and performance standards describe not only the students' current levels of literacy but also offer explicit targets for future achievement.

### *Pedagogical improvement*

Rarely do large-scale assessments offer clear direction for curriculum and instruction or validate curriculum initiatives, but such is the case in the Saskatchewan LAP, the national SAIP and the international IALS. They have demonstrated that assessment can be done on a large-scale basis yet still be sensitive to individual student language processes. For example, the introduction of classroom sample evaluations validates ways of collecting information for individual student assessment in classrooms. In Saskatchewan, compared to the national SAIP assessment, although the broad sweep of findings is similar, there is more sophisticated information about reading and writing processes which, in turn, allow for more direct pedagogical improvements. The SAIP has also yielded an, as of yet untapped, wealth of information to participating ministries of education about students' preferences, practices and interests in the various provinces, which can be used to modify programs. In a consumer-oriented society, can a curriculum guide that has been shaped according to the empirically-documented literacy tastes, attitudes and habits of Canadian adolescents be far off?

To the extent that assessment drives instruction in classrooms such developments are constructive. The extent to which large-scale assess-

ments can validate alternative means of classroom assessment is also a positive outcome of literacy measurement. An understanding of how curriculum reform and assessment might serve each other's purposes rather than being antagonists will need to be fostered in Canada if testing is to inform and advance the educational agenda in positive ways.

#### REFERENCES

- Anderson, J., Muir, W., Bateson, D., Blackmore, D., & Rogers, W. (1990). *The impact of provincial examinations on education in British Columbia: General report*. (Report to the British Columbia Ministry of Education.) Victoria, BC: British Columbia Ministry of Education.
- Angoff, W. (1971). Scales, norms and equivalent scores. In R.L. Thorndike (Ed.), *Educational measurement* (2nd. ed.), (pp. 508-600). Washington, DC: American Council on Education.
- Barlow, M., & Robertson, H.J. (1994). *Class warfare: The assault on Canada's schools*. Toronto, ON: Key Porter Books.
- Bernstein, B. (1977). *Class codes and control* (Vol. 3). London: Routledge & Kegan Paul.
- Council of Ministers of Education, Canada. (1995a). *School achievement indicators program: 1994 reading and writing assessment*. Toronto, ON: Author
- Council of Ministers of Education, Canada (1995b). *School achievement indicators program: 1994 reading and writing assessment. Technical report*. Toronto, ON: Author.
- Froese, V. (1998, June). *Girls and boys reading: What's the DIF*. Paper presented at the annual meeting of the Canadian Society for Studies in Education, Ottawa, ON.
- Gambell, T. & Hunter, D. (in press). Rethinking gender differences in literacy. *Canadian Journal of Education*.
- Graff, H. (1987). *The labyrinths of literacy: Reflections on literacy past and present*. London: The Falmer Press.
- Green, J. (1972). *Psycholinguistics: Chomsky and psychology*. Middlesex, UK: Penguin.
- Mawhinney, H. (1998). Patterns of social control in assessment practices in Canadian frameworks for accountability in education. *Educational Policy*, 12(142), 98-109.
- Messick, S. (1989). Validity. In R.I. Linn (Ed.), *Educational measurement* (3<sup>rd</sup> ed.), (pp 12-103). Washington, DC: The American Council on Education and the National Council on Measurement in Education.
- Nagy, P. (1997, March). *Secondary analysis of data from Canada's indicators program*. Paper presented to the annual meeting of the American Educational Research Association, Chicago, IL.
- Reigeluth, C. (1997, November). Educational standards: To standardize or to customize learning? *Phi Delta Kappan*, 79(3), 202-206.
- Saskatchewan Education. (1989). *Policy for English language arts kindergarten to grade twelve for Saskatchewan schools: Summary paper*. Regina, SK: Author.
- Saskatchewan Education. (1996). *1994 provincial learning assessment in language arts (reading and writing)*. Regina, SK: Author.

## *The Evolution of Literacy*

Saskatchewan Education, Training and Employment. (1984). *Directions: The final report*. Regina, SK: Author.

Shalla, V., & Schellenberg, G. (1998). *The value of words: Literacy and economic security in Canada*. Ottawa, ON: Statistics Canada.

Statistics Canada. (1996). *Reading the future: A portrait of literacy in Canada*. Ottawa, ON: Human Resources Development Canada, and the National Literacy Secretariat.

Statistics Canada. (1996). Backgrounder on the international adult literacy survey (IALS). Pamphlet, 8 pp. Ottawa, ON: Human Resources Development Canada, and the National Literacy Secretariat.

Wideen, M., O'Shea, T., Pye, I., & Ivany, G. (1997). High-stakes testing and the teaching of science. *Canadian Journal of Education*, 22(4), 428-444.

Willms, J.D. (1997). *Literacy skills of Canadian youth*. Ottawa, ON: Statistics Canada.

Willinksy, J. (1990). The construction of a crisis: Literacy in Canada. *Canadian Journal of Education*, 15 (1), 1-15.

**TREVOR GAMBELL** is Professor of Curriculum Studies, College of Education, University of Saskatchewan, Saskatoon.

**DARRYL HUNTER** is Director of Assessment and Evaluation, Saskatchewan Department of Education, Regina.

**TREVOR GAMBELL** est professeur d'études sur les programmes au College of Education de l'Université de la Saskatchewan, à Saskatoon.

**DARRYL HUNTER** est directeur de l'évaluation et de l'appréciation au ministère de l'Éducation de la Saskatchewan, à Régina.