
The Development of the Canadian Language Benchmarks Assessment

Bonny Norton Peirce and Gail Stewart

In this article the authors describe the development of a new language assessment instrument that will be used across Canada to place adult newcomers in instructional programs appropriate for their level of proficiency in English. The development of the instrument represents one step in a lengthy process of federal and grassroots initiatives to establish a common framework for the description and evaluation of the language proficiency of adult newcomers who speak English as a second language. The authors, who were the test developers on the project, provide an introduction to the development of the instrument, referred to as the Canadian Language Benchmarks Assessment (CLBA). They describe the history of the project and challenges they faced in the test development process. In addition, they give an account of how the instruments were field tested, piloted, and scored. They conclude with a brief discussion of work in progress on the ongoing validation of the instrument.

Introduction

The development of the Canadian Language Benchmarks Assessment (CLBA) represents one step in a lengthy process of federal and local initiatives to establish a common framework for the description and evaluation of the language proficiency of adult newcomers to Canada. Two reasons for the development of the CLBA are provided in the document, *Language Benchmarks: English as a Second Language for Adults* (Citizenship and Immigration Canada [CIC], n.d., p. 1).

- Different programs use different names to describe the same level. A level in one program may be called "Intermediate." In another program that same level may be "Level 7" or perhaps "Advanced." There is no common way to describe the levels.
- One language program does not usually accept ESL certificates from another program because ESL programs do not have a common language to describe what students have learned.

In this article we first describe the history of the project and the test development mandate. We then discuss the challenges we faced in attempting to address the sometimes conflicting demands of the mandate. This is followed by a more detailed description of the instruments and the field testing and pilot testing procedures. We then turn to a description of how the

instruments are administered and scored. The conclusion addresses current work in progress on the ongoing validation of the CLBA.

Because of the scope of the CLBA project, three texts are essential complements to this article. The first text, *Language Benchmarks: English as a Second Language for Adults* (CIC, n.d.) is the original draft document that served as the test specifications for the CLBA. Our mandate was to develop the CLBA in accordance with this document. Because for security reasons we are not able to provide examples of the tasks we developed, the reader is referred to this draft Canadian Language Benchmarks (CLB) document for examples of sample tasks. The second text is *Canadian Language Benchmarks: English as a Second Language for Adults/English as a Second Language for Literacy Learners Working Document* (CIC, 1996). This document is a revised version of the draft CLB document and provides an introduction to the CLB, the theoretical approach adopted, and CLB descriptors. The third text, *A Report on the Technical Aspects of Test Development of the Canadian Language Benchmarks Assessment* (Nagy, 1996)¹ provides a detailed account of the design and rationale for the pilot study of the reading and writing assessments and the results obtained. It also includes a brief assessment of the listening/speaking instrument.

History of the Project

In its annual report to Parliament in 1991, Employment and Immigration Canada (now Citizenship and Immigration Canada) indicated its intention to improve the language training offered to adult newcomers by improving language assessment practices and referral procedures (Immigration Canada, 1991). As Rogers (1993) indicates:

In announcing its new immigrant language training policy, Employment and Immigration Canada stressed that a key to developing the most effective training possible is to clearly relate the training to the individual needs of clients. To do this, reliable tools are needed to measure the language skills possessed by clients against standard language proficiency criteria. For federally funded training this will mean that real client language needs can be met and that clients will have access to equivalent types and results of training regardless of where they settle in Canada. (p. 1)

An important innovation of this new policy was the emphasis placed on partnerships between the federal government and local organizations involved in immigrant language training (Rogers, 1994). In this spirit, in 1992 CIC organized a number of consultation workshops to consider what potential benefits there might be in having a set of national language benchmarks to offer to English as a Second Language (ESL) learners, teachers, administrators, and agencies serving immigrants. In March 1993 the federal govern-

ment established the National Working Group on Language Benchmarks (Taborek, 1993) to oversee the development of a language benchmarks document that would describe a "learner's abilities to accomplish tasks using the English language" (CIC, n.d., p. 3). This group comprised stakeholders from across the country (see Appendix A) who met regularly throughout the development of the draft CLB document. Two resources that were influential in the development of this document were *The Certificate in Spoken and Written English* published in Australia (Hagan et al., 1993) and the College Standards and Accreditation Council pilot project (CSAC), which developed benchmarks for ESL programs in the Ontario college system (CSAC, 1993). In 1995 the draft CLB document was field tested extensively with stakeholders from various parts of the country (Crawford, 1995), and following that field testing the revised CLB document (CIC, 1996), as described above, was produced. This document defines 12 benchmarks that describe learner performance in each of three skill areas: listening/speaking, reading, and writing.

In March 1995 the Peel Board of Education in Mississauga, Ontario, was contracted to develop assessment instruments that would be compatible with the draft CLB document (Calleja, 1995). The project team comprised two test developers, Bonny Norton Peirce (University of British Columbia) and Gail Stewart (University of Toronto), and two Peel Board representatives, Tony da Silva (project manager) and Mary Bergin (project coordinator). The test developers were assisted by Philip Nagy (OISE/University of Toronto), the measurement consultant; Alister Cumming (OISE/University of Toronto) the principal consultant; and a team of assessment specialists.² The assessment instruments were developed from April 1995 to April 1996.

Because the contract for the development of the assessment instruments ran concurrently with the contract for field testing of the draft CLB document, it was this document rather than the revised CLB document that was used to determine the initial specifications for test development. As the tasks for the CLBA were developed, they were taken into the field and compared with the descriptors in the draft CLB document. In this way it was possible to allow the task-writing stage of test development to feed into the refinement of the test specifications (Lynch & Davidson, 1994). Test development and revision of the draft CLB document thus became an iterative process, culminating in the revised CLB document.

Our mandate was to work with the draft CLB document to develop a task-based assessment instrument that would address benchmarks 1-8 for the three separate ESL skill areas. The intended purpose of the assessment was to place learners into ESL programs most suitable to their needs. We were also contracted to develop an outcomes instrument to assess learner progress in ESL programs. It is important to note, however, that the outcomes instrument will be valid only to the extent that ESL curricula are

consistent with the objectives of the CLB—an issue that was beyond the scope of our project. Stakeholders had indicated that the instruments should be flexible enough to apply in a range of program placement circumstances, from integrated classrooms to separate skills applications. For this reason it was deemed important to develop instruments that would treat the three skill areas separately and provide diagnostic information for use by instructors.

Each CLBA kit contains the following documents: *Introduction to the CLBA* (Bergin, da Silva, Peirce, & Stewart, 1996); *Listening/Speaking Assessment Manual* (Stewart & Peirce, 1996); *Reading and Writing Assessment Manual* (Peirce & Stewart, 1996a); a CLBA Client Profile Form; a videotape for the listening/speaking assessment; a photo-story; five photo-spreads; a Listening/Speaking Assessment Guide; and a Listening/Speaking Assessment Form. Each kit also contains eight prototype assessment forms: four for reading and four for writing, divided into Stage I and Stage II, placement and outcomes respectively.

Test Development Challenges

The CLBA had to comprise tasks representative of the functions and activities outlined in the various stages of the draft CLB document. These tasks include the day-to-day tasks that adults need to accomplish in order to function successfully in Canadian society. These tasks had to represent increasing levels of difficulty for most learners and be culturally accessible to people from a wide variety of backgrounds. In addition, the CLBA had to be user-friendly; that is to say, the instruments had to be designed for efficient, reliable, and cost-effective administration and scoring. Furthermore, it needed to be accountable to ESL learners and teachers.

A central priority in the test development process was the development of culturally accessible tasks. In this regard, task-based assessment is a double-edged sword. It is appealing because the tasks that are assessed can be seen as relevant to learner needs and authentic in communicative intent (Canale & Swain, 1980). On the other hand, task-based assessment can be challenging because so many relevant tasks may assume knowledge of cultural practices that are unfamiliar to some candidates. Indeed, to a greater or lesser extent all language assessment instruments assume some kind of cultural knowledge on the part of the learner, whether this knowledge is about assessment practices, testing conditions, item formats, or background information. Knowledge about culture assumes both knowledge of content (Courchène, 1996) and knowledge of social relationships and structures (Sauvé, 1996). We wanted to ensure that most learners would be able to access the various tasks; however, it was neither possible nor desirable to strip the assessment content of its cultural context. To do so would have been contrary to the spirit of the draft CLB document and would have resulted in bland, in-

authentic content that would have little meaning or relevance to learners of ESL in Canada. The focus of our test development was therefore on the development of culturally accessible tasks rather than culturally "free" tasks. We were concerned, furthermore, that the desire to separate language skills into three distinct skill areas (listening/speaking, reading, and writing) as specified by the draft CLB document would not be compatible with the more holistic approach to language competence implicit in task-based assessment (Brindley, 1995; McNamara, 1995; Wesche 1987). This is an issue we had to struggle with throughout the test development process. Because the field had indicated that separate-skills evaluation was important, we sought by diverse means to prompt language production that would not cause weakness in one skill area adversely to affect results in another.

Another important consideration in the test development centered around conditions of administration. Because most learners would be required to take three assessments on the same day, timing was a key concern. It was important that the three components of the assessment each be lengthy enough to ensure reliable placement, but not so long as to tax the stamina of the learner and the resources of the assessment center. In addition, materials had to be designed for reliable administration in variety of assessment situations, from large centers to itinerant settings.

Finally, we sought to be as accountable as possible throughout the test development process. Issues of authenticity and cultural diversity remained major challenges in this regard (Peirce & Stewart, 1996b). We recognized the importance of seeking input from all the major stakeholders in the CLBA, in particular learners of different cultural backgrounds, ESL teachers and teacher trainers, community stakeholders, the NWGLB, and Citizenship and Immigration Canada. Our approach to accountability was informed by the *Principles of Fair Student Assessment Practices for Education in Canada* (Wilson, 1996) and current literature on accountability in language assessment (Cumming, 1994; Elson, 1992; Lacelle-Peterson & Rivera, 1994; Moore, in press; Peirce & Stein, 1995; Shohamy, 1993).

In addressing these test development challenges, we gained valuable input from a wide variety of stakeholders and our team of assessment specialists. In the field testing stage, tasks were sent to these assessment specialists for their comment and critique. Regular meetings were held with members of the National Working Group on Language Benchmarks (NWGLB). In addition, a Cultural Advisory Group was struck in the region of Peel, consisting of members of service agencies, settlement workers, and English language learners, who reviewed the assessment instruments and materials in their development stages and gave valuable input. Furthermore, we included a wide variety of tasks and item types in each of the instruments in order to increase the opportunities available to learners to perform at their best.

Development of the Instrument

The CLBA has three separate instruments: a Listening/Speaking Assessment, a Reading Assessment comprising two parallel forms, and a Writing Assessment comprising two parallel forms. The parallel forms of reading and writing are for the purposes of program placement and outcomes respectively. All three instruments have a Stage I assessment and a Stage II assessment, with Stage II being more complex and demanding than Stage I (Nagy, 1996). The tasks in Stage I are relatively short and related to information of a personal nature, whereas the tasks in Stage II are longer, more cognitively demanding, and related to information at the community level. In all cases learners must achieve an advanced placement in Stage I before they are eligible to proceed to Stage II. In keeping with the spirit of the CLB documents, the tasks in Stage II are parallel in type to the tasks in Stage I.

In developing the listening/speaking assessment, we considered first and foremost the comfort of the client and the flow of the interaction. To this end we developed a one-to-one conversation in which the learner is guided from content that is simple and familiar toward material that is more challenging. Wherever possible we introduced the element of choice, so that a learner can direct the conversation toward topics that she or he considers most relevant. In an effort to create materials that would be interesting and accessible to a wide range of learners from different backgrounds, we consulted with learners, instructors, assessors, and representatives of various cultural agencies to determine which themes and topics would be most suitable.

The prompts for the listening/speaking assessment consist of verbal questions and instructions from a live interlocutor (the assessor-facilitator), photographs, a photo-story, video materials, and audio materials. In creating specifications for the photography, we examined the initial A-LINC assessment (Tegenfeldt & Monk, 1992), which makes effective use of visual prompts. More than 200 photographs were taken. These were examined by ESL learners, instructors, and members of the Cultural Advisory Group for clarity, accessibility, and relevance. Video and audio prompts were professionally recorded, tested with learners, and reviewed by ESL professionals, and then revised accordingly. The listening/speaking tasks, which are described more fully in the revised CLB document, are summarized as follows.

Stage I Listening/Speaking Tasks

Task Type A: Follows and responds to simple greetings and instructions;

Task Type B: Follows and responds to questions about basic personal information;

Task Type C: Takes part in short informal conversation about personal experience;

Task Type D: Describes the process of obtaining essential goods and services.

Stage II Listening/Speaking Tasks

Task Type A: Comprehends and relates video-mediated instructions;

Task Type B: Comprehends and relates audio-mediated information;

Task Type C: Discusses concrete information on a familiar topic;

Task Type D: Comprehends and synthesizes abstract ideas on a familiar topic.

In the initial development of the reading and writing assessments, a team of task-writers worked with us to create a bank of tasks according to the specifications of the draft CLB document. The task-writers were Enid Jorsling (Peel Board of Education), Donna Leeming (Peel Board of Education), Kathleen Troy (Mohawk College), and Howard Zuckernick (University of Toronto). The task-writers were instructed to study the draft CLB document and create materials that would be relevant to newcomers, appropriate in length and level, and equitably accessible to learners from diverse cultures settled in different parts of the country.

At the end of the task-writing phase, 160 original tasks had been created, 80 for reading and 80 for writing. Numerous stakeholders responded to the format and content of the original tasks, which were accordingly eliminated or revised before field testing. Following the field test procedures, tasks were assembled to create various forms for pilot testing so that psychometric data could be gathered and analyzed. The reading and writing tasks, which are described more fully in the revised CLB document, are summarized as follows.

Stage I Reading Tasks

Task Type A: Reads simple instructional texts;

Task Type B: Reads simple formatted texts;

Task Type C: Reads simple unformatted texts;

Task Type D: Reads simple informational texts.

Stage II Reading Tasks

Task Type A: Reads complex instructional texts;

Task Type B: Reads complex formatted texts;

Task Type C: Reads complex unformatted texts;

Task Type D: Reads complex informational texts.

Stage I Writing Tasks

Task Type A: Copies information;

Task Type B: Fills out simple forms;

Task Type C: Describes personal situations;

Task Type D: Expresses simple ideas.

Stage II Writing Tasks

Task Type A: Reproduces information;

Task Type B: Fills out complex forms;
Task Type C: Conveys formal messages;
Task Type D: Expresses complex ideas.

Field Testing and Pilot Testing

In the development of the CLBA, we distinguished between *field testing* and *pilot testing*. In the field testing, which served as a preparation for the pilot testing tasks went through a trial run in which we sought to reduce weaknesses in the tasks, hone the task instructions, and assess the time learners needed to complete the tasks. In the pilot testing we sought to collect data from a wide range of learners for the purposes of measurement and analysis.

We field tested the listening/speaking instrument in the Peel region, working closely with two experienced assessors, Carolyn Cohen and Audrey Bennett. Twenty-two learners of varying levels of proficiency in English were interviewed. The interviews were videotaped and carefully analyzed. Furthermore, we field tested the Stage II listening tasks in a group setting at the School of Continuing Studies, University of Toronto. Through this process a number of prompts were revised or eliminated and the scoring procedures refined. A more extensive pilot study of the listening/speaking instrument is recommended for future research (see Conclusion to this article).

Our field testing objectives for listening/speaking were to determine to what extent the assessment format and content facilitated the production of a learner's best possible language sample. We wanted to find out whether the structure of the assessment put learners at ease and allowed them to draw sufficiently on their own background and experiences. In addition, in the course of the field testing, we worked on the transitions between tasks so that the learner would perceive the conversation as a natural progression and not a series of unrelated tasks.

We had all learners in the field test begin with the first task in the Stage I assessment and progress through the conversation until threshold was reached. Threshold was identified by the assessor as the point at which a learner's language began to break down. At that point previously confident learners gradually lost confidence and sometimes began to apologize for their expression. During the field test we asked the assessors to take the clients progressively beyond threshold so that we could ascertain whether our assumptions about the progressive difficulty of the prompts were justified. In a regular assessment, however, the assessor takes the learner to this threshold, pushes only briefly beyond it to confirm that the learner is struggling at that level, and then quickly brings the conversation back to a level at which the learner is comfortable. The assessment is always terminated with pleasantries and reassurance.

Following the development of the listening/ speaking assessment, a study was conducted in which 17 assessors responded to statements about the validity and quality of the instrument. There were 30 statements included in the study, with space for additional comments. Responses were scored on a 5-point scale, with 5 representing the most favorable response to the instrument. The data were analyzed by our measurement consultant. For each statement in the study, an average score out of 5 was reported. Average scores ranged from a minimum of 3.18 to a maximum of 4.35. The following are reported averages on responses to some key statements: "The CLBA Listening/Speaking Assessment offers clients adequate opportunity to demonstrate their proficiency in listening/speaking" (4.06); "The CLBA interview becomes progressively more challenging for clients" (4.06); "The tasks are relevant to adult newcomers to Canada" (3.94). As a result of the feedback obtained, we were able to make further refinements to the listening/speaking assessment.

The reading and writing assessments each underwent a field test and a pilot test. During the reading and writing field testing phase, we gathered responses from a variety of sources including learners, instructors, assessors, and administrators with regard to the cultural accessibility of the tasks, the average length of time required for task completion, the clarity and simplicity of the instructions, and the relative ease of administration. To this end we sent the tasks out into the field and gathered qualitative feedback from teachers and learners as well as quantitative information on learner performance. One strategy we adopted was to provide teachers with a chart on which they recorded their observations of learners performing the field test tasks. If a learner indicated, for example, that he or she did not understand a word or an instruction, the teacher recorded this, often providing an explanation for the confusion.

The institutions involved in the piloting process were the Dixie Bloor Neighborhood Centre (Toronto), the Halifax Immigrant Learning Centre, the Ottawa Board of Education, the Peel Board of Education, and Vancouver Community College. There were 12 pilot forms in total: six for reading and six for writing. Because we wanted the final product to comprise two parallel forms (one placement and one outcomes) for each of reading and writing at Stage I and Stage II respectively, the following breakdown was necessary for the pilot process: Reading Stage I: three forms; Reading Stage II: three forms; Writing Stage I: three forms; Writing Stage II: three forms. By piloting three forms rather than two, we gave ourselves room for attrition. The total number of participants in the pilot process was 1,140 with a total number of 2,280 forms administered.

The reading and writing pilot study was designed, analyzed, and interpreted by our measurement consultant (Nagy, 1996). The primary purpose of the pilot study was to determine whether the three forms in each respective

stage were equivalent in difficulty. For this reason each participant in the pilot responded to two forms from either reading or writing. We then chose the two forms that had the greatest equivalence (for reading and writing respectively, and for Stage I and Stage II) as our placement and outcomes assessments. Furthermore, on the advice of our marking team, David Progosh (University of Toronto) and Howard Zuckernick (University of Toronto), we reworded some of the writing prompts, simplified vocabulary, and made the task objectives clearer.

At present, therefore, the CLBA comprises eight forms in total: four for reading and four for writing. Of the four forms for each respective skill, two constitute the placement assessments and two the outcomes assessments. Of the two placement assessments, one is a Stage I assessment and one a Stage II assessment. Likewise, of the two outcomes assessments, one is a Stage I assessment and one a Stage II assessment. In his report, Nagy (1996) notes the following:

The final tests are sufficiently reliable. On the 4-point [benchmark] scale, about 90% of students (slightly more for reading, slightly less for writing) would receive identical scores, or scores within one point of each other, if writing both [placement and outcomes] tests. (p. 21)

In a low-stakes placement test, these findings were deemed satisfactory. If this had been a high-stakes, gatekeeping test for college entrance, job entry, or immigration, we could not have been complacent.

Administration and Scoring Procedures

Implicit in the draft CLB document was the assumption that language tasks could be placed in hierarchical order, in which, for example, a task at benchmark 3 would be defined as easier than a task at benchmark 4. Although we attempted to create listening / speaking, reading and writing tasks of increasing levels of complexity and were generally successful for approximately 70% of learners (Nagy, 1996), we were concerned that a hierarchy of tasks did not “bias for best” for 100% of learners (Swain, 1984). For example, learners who were competent at tasks such as letterwriting (a supposedly challenging task) but had had little experience of filling out forms (a supposedly easier task) may have been placed at a benchmark level that did not do justice to the range of their writing proficiency. In choosing to bias for best, we did not want to penalize those learners whose language proficiency, for a variety of social, cultural, and historical reasons, did not fit neatly into a given hierarchy of tasks. For this reason we have given learners credit for their performance on a range of tasks at each respective stage, and have based their benchmark placement on a composite score that reflects their performance on all tasks attempted in a given stage.

The listening/speaking assessment, administered on a one-to-one basis, can take between 10 and 30 minutes to administer. The assessor is also the interviewer/facilitator, and scoring takes place at the time the instrument is administered. For this reason we had to devise a system that could be used reliably and unobtrusively by a trained assessor while engaged in a conversation with the learner. The assessor works throughout the interview with two documents—an Assessment Form and an Assessment Guide. Because all CLBA assessors have been thoroughly trained and tested, it is assumed that they are familiar with the interview protocol. However, the Assessment Guide is kept handy on the table to serve as a reminder of procedures, prompts, key decisions, and scoring procedures. On the Assessment Form, the assessor records information about the learner's performance and makes diagnostic notes for use in placement and instruction.

During the assessment the assessor engages the learner in a conversation and prompts her or him to give independent responses on a range of tasks. The assessor is trying to determine to what extent the learner is able to take a "long turn," or to direct the conversation. When it is clear that a learner is struggling with what we call "independent production," the assessor moves through a series of guided prompts to facilitate the production. The assessor continues to prompt the client until a proficiency threshold is reached, at which point the interview is terminated and a benchmark assigned in accordance with the eight benchmark descriptors included in the Assessment Guide.

The reading assessment and the writing assessment can be administered on a one-to-one basis or in a group setting. Learners may be given up to 45 minutes to complete a Stage I assessment and up to an hour to complete a Stage II assessment in both the reading and writing assessments. However, many learners complete the assessments in much less than the allotted time.

In the reading assessment, the client responds to a range of tasks, each of which comprises several items. The total number of tasks in each stage is four. The total number of items in Stage I is 30 and the total number of items in Stage II is 32. For each task the total number of correct item responses (the Task Score) is converted into a Performance Indicator of 1, 2, or 3. A score of 1 indicates that a learner has achieved limited success on the task; 2 indicates marginal success; 3 indicates successful performance. This conversion was devised by our measurement consultant in order to maintain the relative weight of the respective tasks and to ensure equivalence across the placement and outcomes instruments. The Performance Indicators are totalled to render a Composite Score with a minimum of 4 points and a maximum of 12, which is then converted to a benchmark.

In the writing assessment, we reviewed samples of writing reflecting the full range of proficiency of the participating learners. We drew a distinction between primary and secondary objectives in the successful execution of a

task. We defined *primary* objectives as those that address the task-based nature of the prompt. These include the extent to which the writer addresses the purpose and scope of the task and the intended audience. The *secondary* objectives include the extent to which the writer has adequate control of grammar, spelling, and mechanics. A learner's response to each task is given a Performance Indicator of 1, 2, 3, or 4, with 4 representing success on the task. Each task has a set of criteria to guide the decision-making process and a set of four exemplars, representing a Performance Indicator of 1, 2, 3, or 4 for each of the tasks assessed. As with reading assessment, the writing Performance Indicators are totalled to render a Composite Score (in this case with a minimum of 4 points and a maximum of 16), which is then converted to a benchmark. Because there are four tasks in each of the respective stages, in both the placement and outcomes instruments (i.e., a total of 16 tasks), we needed to select 64 exemplars from the pilot study and include these in the *Reading and Writing Assessment Manual*.

Conclusion

The CLBA, like the revised CLB document, remains a low-stakes work in progress, representing one contribution to nationwide attempts to improve the language learning opportunities and integration of new Canadians. It is the result of collaboration among many learners, teachers, administrators, federal and provincial officials, and assessment specialists across Canada. A motion passed by the TESL Canada Board on November 25, 1996 represented another chapter in the unfolding story of the CLBA (McMichael, personal communication, November 26, 1996). The motion read as follows.

That TESL Canada endorse the adoption of the Canada Language Benchmarks Assessment by language training providers and trainers in Canada and that the TESL Canada president inform the federal minister responsible for Citizenship and Immigration and all provincial ministers responsible for English language training of this endorsement.

The CLBA is as valid as the process that has generated it. Much work remains to be done to enhance its validity and reliability. In this regard, Nagy (1996) notes the following.

This project has made a good start on test development. We have dealt with the issues of equivalence of parallel forms of the tests, and with the hierarchical nature of the Benchmark skills. Priority issues for further development include examination of interscorer agreement, especially for the subjective decisions required in the Writing tests, investigation of the relationship between Reading, Writing, and Listening/Speaking skills, and collection and analysis of student data from the Listening/Speaking tests. (p. 22)

The Peel Board of Education is in the process of training assessors in different parts of the country to use the CLBA efficiently and effectively and has embarked on an interrater reliability study of the writing assessment (C. Cohen & T. da Silva, personal communication, October 24, 1996). Furthermore, work has begun on the development of a literacy assessment for learners whose needs are not met by the CLBA. In time, and with ongoing research, the CLBA may well meet the expectations expressed by da Silva (1996):

It is our hope that the CLBA ... will bring about integration and coherence in second language training in this country, and by extension, allow learners to move through the training and education system as efficiently as possible. (p. 1)

Notes

¹This report can be obtained from Tony da Silva, Director, Centre for Language Training and Assessment, 2 Robert Speck Parkway, 3rd Floor, Suite 300, Mississauga, ON L4Z 1H8.

²These assessment specialists were Margaret des Brisay (University of Ottawa), Helen Tegenfeldt (Vancouver Community College), Marian Tyacke (University of Toronto), and Mari Wesche (University of Ottawa).

Acknowledgments

We acknowledge the hundreds of learners and teachers across Canada who have generously contributed to the development of the CLBA. Discussions with members of the NWGLB have been invaluable. Geoff Brindley and Helen Moore gave us access to helpful Australian resources. We thank Caroline Clapham, Alister Cumming, and three anonymous *TESL Canada* reviewers for their insightful comments on an earlier draft of this article. The support of Citizenship and Immigration Canada is gratefully acknowledged.

The Authors

Bonny Norton Peirce is an assistant professor in the Department of Language Education, University of British Columbia. She has worked on the development of language assessment instruments internationally, including the TOEFL, the Test of Written English, and the University of Toronto's COPE test. Her research on language assessment has been published in *Applied Linguistics*, *Harvard Educational Review*, and *TESOL Quarterly*.

Gail Stewart is an ESL instructor and teacher trainer at the University of Toronto. Her experience in language test development includes the University of Toronto's COPE test, the Ontario College of Midwives' Language Proficiency Test, Citizenship and Immigration Canada's Citizenship Application Test, and, currently, the Canadian Language Benchmarks for Literacy Assessment.

References

- Bergin, M., da Silva, T., Peirce, B.N., & Stewart, G. (1996). *Introduction to the Canadian Language Benchmarks Assessment*. Mississauga, ON: Peel Board of Education.
- Brindley, G. (Ed.). (1995). *Language assessment in action*. Sydney: National Centre for English Language Teaching and Research.
- Calleja, F. (1995, April 7). Board to devise English skills test. *Toronto Star*, p. A7.

- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1-47.
- Citizenship and Immigration Canada. (1996). *Canadian Language Benchmarks: English as a second language for adults/English as a second language for literacy learners*. Working Document. Ottawa, ON: Minister of Supply and Services Canada.
- College Standards and Accreditation Council (CSAC), Ontario. (1993). *ESL Benchmarks, pilot project* (by Dianne Coons and Pat Parnell). Toronto: Author.
- Courchène, R. (1996). Teaching Canadian culture: Teacher preparation. *TESL Canada Journal*, 13(2), 1-16.
- Crawford, K. (1995). *Language Benchmarks report on field testing: Issues and recommendations*. Unpublished manuscript.
- Cumming, A. (1994). Does language assessment facilitate recent immigrants' participation in Canadian society? *TESL Canada Journal*, 11(2), 117-133.
- da Silva, T. (1996). Preface. In M. Bergin, T. da Silva, B.N. Peirce, & G. Stewart (Eds.), *Introduction to the Canadian Language Benchmarks Assessment* (p. 1). Mississauga, ON: Peel Board of Education.
- Elson, N. (1992). The failure of tests: Language tests and post-secondary admissions of ESL students. In B. Burnaby & A. Cumming (Eds.), *Socio-political aspects of ESL in Canada*. Toronto, ON: OISE Press.
- Hagan, P., Hood, S., Jackson, E., Jones, M., Joyce, H., & Manilis, M. (1993). *Certificate in spoken and written English* (2nd ed.). Sydney, Australia: NSW Adult Migrant English Service and National Centre for English Language Testing and Research (NCELTR).
- Immigration Canada. (1991). *Annual report to parliament, immigration plan for 1991-1995, year two*. Ottawa, ON: Employment and Immigration Canada.
- Lacelle-Peterson, M., & Rivera, C. (1994). Is it real for all kids? A framework for equitable assessment policies for English language learners. *Harvard Educational Review*, 64, 55-75.
- Lynch, B.K., & Davidson, F. (1994). Criterion-referenced language test development: Linking curricula, teachers, and tests. *TESOL Quarterly*, 28, 727-743.
- Moore, H. (in press). Telling what is real: Competing views in assessing ESL development. *Linguistics and Education*.
- McNamara, T.F. (1995). Modelling performance: Opening Pandora's box. *Applied Linguistics*, 16, 159-179.
- Nagy, P. (1996, April). *A report on technical aspects of test development for the Canadian Language Benchmarks Assessment*. Unpublished manuscript.
- Peirce, B.N., & Stewart, G. (1996a). *The Canadian Language Benchmarks Assessment reading and writing manual*. Mississauga, ON: Peel Board of Education.
- Peirce, B.N., & Stewart, G. (1996b, August). *Accountability in assessment: Challenges of authenticity and cultural diversity*. Paper presented at the 18th annual Language Testing Research Colloquium, Tampere, Finland.
- Peirce, B.N., & Stein, P. (1995). Why the "Monkeys Passage" bombed: Tests, genres, and teaching. *Harvard Educational Review*, 65(1), 50-65.
- Rogers, E. (1993, April). National working group on language benchmarks meets. *TESL Canada Bulletin*, p. 1-2.
- Rogers, E. (1994). Canadian federal language policy and the Benchmarks project. *TESOL Matters*, 3(6), pp. 1, 5.
- Sauvé, V. (1996). Working with cultures of Canada in the ESL classroom: A response to Robert Courchène. *TESL Canada Journal*, 13(2), 17-23.
- Shohamy, E. (1993). *The power of tests: The impact of language tests on teaching and learning*. Washington, DC: NFLC Occasional Papers.
- Stewart, G., & Peirce, B.N. (1996). *The Canadian Language Benchmarks Assessment listening/speaking manual*. Mississauga, ON: Peel Board of Education.
- Swain, M. (1984). Teaching and testing communicatively. *TESL Talk*, 15, 7-18.

- Taborek, E. (1993, fall-winter). The national working group on language benchmarks. *TESL Toronto Newsletter*, pp. 10-11.
- Tegenfeldt, H., & Monk, V. (1992). *Assessment interview: Language instruction for newcomers to Canada*. Vancouver, BC: Vancouver Community College.
- Wesche, M. (1987). Second language performance testing: The Ontario test of ESL as an example. *Language Testing*, 4(1) 28-47.
- Wilson, R.J. (1996). *Assessing students in classrooms and schools*. Scarborough, ON: Allyn and Bacon.

Appendix A: Members of the National Working Group on Language Benchmarks

- Jamie Baird, Victoria, BC
 Joan Baril, Thunder Bay, ON
 Bitu Bateni, North Vancouver, BC
 Elza Bruk, Calgary, AB (alternate, Sharon George)
 Raminder Dosanjh, Vancouver, BC
 Catarina Garcia, Charlottetown, PEI
 Maureen Gross, Edmonton, AB
 Artur Gudowski, (Co-Chair), Regina, SK
 Sutrisna Iswandi, Lethbridge, AB
 Mary Keane, Halifax, NS
 Grant Lovelock, Vancouver, BC
 Lynne McBeath, Fredericton, NB
 Pat Parnall, Peterborough, ON
 D'Arcy Phillips, Winnipeg, MB
 Eleanor Rogers, Kingston, ON
 Peggie Shek, Toronto, ON
 Elizabeth Taborek, Toronto, ON
 Martha Trahey, St. John's, NF
 Shailja Varma (Co-Chair), Ottawa, ON