

Self-Assessment, French Immersion, and Locus of Control

BONNY N. PEIRCE, MERRILL SWAIN, and DOUG HART

Ontario Institute for Studies in Education

This article compares the self-assessments of French proficiency made by approximately 500 Grade 8 students in two different French immersion programs ('early' and 'middle') in Toronto, Canada. Two self-assessment benchmarks are used: the perceived language proficiency of francophone peers and the difficulty represented by specific everyday tasks in French. The study investigates: (1) the extent to which self-assessment is a valid and reliable indicator of tested proficiency in French immersion programs; (2) how benchmarks influence correlations of self-assessment with tested proficiency; (3) whether self-assessment research can inform or support current theories of second language learning and assessment. The results indicate that: (1) self-assessments of language proficiency correlate only weakly with objective measures of language proficiency; (2) self-assessment measures on specific tasks are more highly correlated with tested proficiency than are global self-assessment measures; (3) irrespective of program, students agree on the relative difficulty of oral and literacy tasks in French under specific conditions of reception and production. These findings are explained with reference to current research on self-assessment, Spolsky's Conditions for Second Language Learning (1989), and the authors' construct of 'locus of control' in a communicative event. It is argued that the locus of control operates at the 'interface' (Bachman 1989) between language assessment and second language acquisition research.

1. INTRODUCTION

In recent years, there has been increasing interest in the role of self-assessment in language learning and teaching (Oscarson 1978, 1984, 1989; Davidson and Henning 1985; LeBlanc and Painchaud 1985; Henner-Stanchina and Holec 1985; Blue 1988; Bachman and Palmer 1989; Blanche and Merino 1989; Janssen-van Dielen 1989). Such interest is a logical outcome of increased interest in learner-centred language teaching and self-directed language learning (Knowles 1975; Holec 1980; Riley 1985; Dickinson 1987; Nunan 1988; Brown 1991). There is general consensus in the literature that language learning is enhanced if the learner takes initiative in the language learning and assessment process and if responsibility for the management of language learning and assessment is shared by both the language teacher and the language learner. Rubin (1975) and Naiman, Fröhlich, Stern, and Todesco (1978) have argued, for example, that the most successful language learners are those who regularly engage in self-assessment as part of their learning strategies. The types of

evaluation tools described in the literature range from the formal and analytical (computer generated assessment, standardized questionnaires) to the informal (self-observation and journal writing). The uses of self-assessment are equally diverse, ranging from its use in university placement (LeBlanc and Painchaud 1985) to its use in the certification of language teachers (Haughton and Dickinson 1988).

Because self-assessment has traditionally been viewed as informal and subjective, much of the current research on self-assessment has concentrated primarily on investigating whether self-assessment instruments and self-ratings of language proficiency are valid and reliable. Oscarson (1978: 4–5), for example, describes research carried out in Sweden in which learners' self-assessments correlated with the assessment of instructors at around 0.60, while self-assessments and formal tests correlated at 0.50. The results of LeBlanc and Painchaud's (1985: 679) study indicate a correlation of 0.53 between a self-assessment placement questionnaire at the University of Ottawa and a standardized English proficiency test. These promising results, however, are not always duplicated in other studies. Janssen-van Dielen (1989: 30, 41) found no consistent significant relationship between performance on a test of Dutch as a second language and a parallel version of that test in self-assessment format. Blanche and Merino (1989: 324) report on two studies in which there were no significant relationships between students' self-assessments of language proficiency and actual classroom/test performance. Ready-Morfit's (1991) study indicates that self-assessments can be unreliable if there is a clearly perceived advantage in mis-assessment. She argues that because of changes in the placement and registration procedures at the University of Ottawa in 1987, there has been a tendency for some students to inflate their self-assessments. As a result, self-assessment no longer enjoys such a high rate of success at the university.

The research on self-assessment is not conclusive because, as Bachman and Palmer (1989: 25) argue, self-assessment studies differ in design and purpose. Davidson and Henning's (1985) study on the applicability of Rasch modelling for the development of self-rating scales indicates that little confidence can be placed in the particular self-ratings they examined. In contrast, Bachman and Palmer's (1989) investigation of the trait structure of an experimental self-rating test indicates that self-ratings can be reliable and valid measures of communicative language ability. Bachman and Palmer (1989) investigated the trait structure through the use of the multitrait multimethod design and the confirmatory factor analysis procedure. They conclude that despite the apparent differences between their study and that of Davidson and Henning (1985), the studies demonstrate the complementary roles of IRT modelling and factor analysis: while IRT is a useful tool for test development, factor analysis is one approach to test validation.

Debates in the self-assessment literature extend to the framing of questions and the appropriate use of benchmarks in self-assessment instruments. More than a decade ago, Oscarson (1978: 14) made the claim that most learners can

determine their language ability provided they have a standard by which to compare themselves. More recently, researchers have been paying greater attention to what kind of standard is used for the purpose of comparison. The research of LeBlanc and Painchaud (1985), for example, indicates that researchers can expect higher correlations between self-assessment measures of specific tasks related to the students' situations (posters on campus, articles in a student newspaper) and tested proficiency than self-assessment measures of global indicators of proficiency (the general understanding of texts) and tested proficiency. Bachman and Palmer (1989) used three question types in their study: 'ability' questions (How many different names of well-known American people and places do you know?); 'difficulty with production' questions (How many different kinds of grammar mistakes do you make in English?); and 'recognition' questions (Can you tell who polite English speaking people are by the kind of English they use?). Of these three question types, they found that the most effective was the 'difficulty with production' question.

Our study has three purposes: First, we wish to contribute to the debate on the validity and reliability of self-assessment as a measure of language proficiency. Much of the literature described above has focused on the self-assessments of adult language learners; our study focuses on the self-assessments of school children. Second, we wish to investigate benchmark effects on self-assessment. Our study incorporates the use of two different benchmarks against which students assess their language proficiency. Third, we wish to investigate whether there are any generalizable findings on self-assessment that can contribute to our understanding of second language learning and assessment theory. This objective is motivated by Bachman's (1989) call for further investigation into the 'interface' between language testing and second language acquisition (SLA) research. Because our self-assessment data are drawn from two different French immersion populations, which have been administered the same self-assessment questions and proficiency tests, we are in a position to determine which self-assessments are *independent* of the respective programs of instruction. In other words, if both groups respond in similar ways to the same questions, notwithstanding their different language programs, we can conclude that we have a potentially generalizable finding that may contribute to a broader understanding of second language learning and assessment.

2. THE STUDY

The data discussed in this paper are taken from an evaluation that was conducted comparing the linguistic outcomes and social character of early and middle French immersion programs in a large-scale French immersion research project in Toronto, Canada, in which approximately 500 students participated (Hart, Lapkin, and Swain 1988; Lapkin, Hart, and Swain 1991). The early immersion students began their program in kindergarten where all their instruction was in French. As they proceeded through the grades, instruction in English was gradually introduced until an approximate balance between the two languages of instruction was reached at grade 5. Middle immersion students in

this study began their program in grade 5 where they received about half their instruction in French and half in English. Prior to entering the immersion program, the students had studied French as a subject since kindergarten. Twenty-six classes of students were tested at the end of grade 8, by which time the early immersion students had accumulated about 5,300–6,040 hours of instructional time in French, and the middle immersion students had accumulated approximately 2,040 hours. Two sorts of data were gathered: questionnaire data (which included the self-assessment questions) and data from tests of French reading, writing, speaking, and listening.

3. METHODOLOGY

A detailed questionnaire and French proficiency tests were administered to the 26 grade 8 classes. In the questionnaire, the immersion students were given two benchmarks against which to assess their own French language skills. The first benchmark was the perceived language proficiency of francophone peers (see Table 1 below). The second benchmark was the difficulty represented by a set of

Table 1: Francophone peer benchmark

My French is:	about the same	somewhat worse	much worse
Speaking	3	2	1
Listening	3	2	1
Writing	3	2	1
Reading	3	2	1

Question to students: Now, thinking about a person your age whose first language is French, (for example, someone from northern Ontario, Quebec or a French-speaking country), how would you compare your French to theirs in the following areas?

specific everyday tasks conducted in French. For the purpose of this study, four of these specific tasks have been selected to serve as points of comparison with objective measures of proficiency (see Table 2 below). These measures have been chosen because the linguistic demands of the measures are closest in equivalence to the objective measures of proficiency described below for each respective skill.

French tests were developed for receptive skills (listening, reading) and productive skills (speaking, writing). The tests were developed with reference to the Canale and Swain framework of communicative language testing (1980), using formats that would allow specification of psychometric characteristics. This involved using quasi-realistic material and providing thematic links between tasks where possible. The tests had to allow for an initially unknown range of student skill levels, both within and between programs. The final tests used for the purpose of comparing self-assessment data with tested proficiency data were the following:

Table 2: *Difficulty with specific tasks benchmark*

	not at all	with much difficulty	with some difficulty	with little difficulty	without any difficulty
(a) Explain the plot of a mystery book or movie to someone face to face	1	2	3	4	5
(b) Listen to and understand a French radio show	1	2	3	4	5
(c) Explain the plot of a book or movie to someone in a letter	1	2	3	4	5
(d) Read and understand a French newspaper or magazine	1	2	3	4	5

Instruction to students: Please circle how well you can do the following in French.

1. A test of listening comprehension, the *test de compréhension auditive, niveau C* (TCAC): The TCAC requires students to answer multiple choice questions based on passages to which they have just listened. There are fifteen questions based on seven passages. The passages are drawn from actual French radio broadcasts in different formats (for example, a news item or interview segment) and there is a mix of male and female voices.

2. An open speaking task: In this speaking task, which is thematically linked to the final passage of the listening test, students were required to comment on the strictness of their own and/or friends' parents and to provide examples. It is scored using a four-point scale of fluency. '0' reflects uneven rhythm, inappropriate stress patterns, and use of frequent and prolonged pauses often in inappropriate places. The top rating of '3' reflects native-like rate of speech, rhythm, stress and intonation patterns, use of liaison, and avoidance of overly long pauses. The speaking task was administered to a random sample of students on a withdrawal basis, always in a session following that for the TCAC.

3. A test of reading comprehension, the *test de mots à trouver, niveau C*: This is a cloze test based on a text about the 'Abominable Snowman' or 'Yeti' purportedly resident in the Himalayas. The 'acceptable' method of scoring was used, yielding a maximum score of 25.

4. An open writing task: In this writing task, which is thematically linked to the reading passage, students were asked to state what they thought about reports on 'strange creatures' and specifically, their own opinion as to whether the 'Yeti' exists. The task received a global judgement measure out of 2 (0, 1, 2) with '0' indicating the use of simple sentence structures and a high number of grammatical errors and '2' indicating the use of complex sentence structures and

relatively few grammatical errors. This task was administered immediately after the cloze test, in the same testing session.

With the exception of the speaking test, all test and questionnaire results are based on populations of students, not samples. The speaking test was administered to a random sample of 8 students per class.

Early and middle immersion students differ very substantially in their exposure to instruction in French. As noted above, total accumulated hours of instruction in French to the end of grade 8 in middle immersion is less than half that in early immersion. This gap is reflected in tested French proficiency on the measures described above (see Appendix). On average, early immersion students outperform those in middle immersion in all skill areas: speaking, listening, writing, and reading (see also Hart, Lapkin, and Swain 1988: 33). Our study investigates whether actual differences in the proficiency between the two programs are paralleled in students' self-assessments with reference to the two benchmarks given in Tables 1 and 2 above. (For results focusing on group differences which include all third-language learners as well as the random sample of French immersion students, see Swain in press.) In the Results section we address the following four questions:

1. How do French immersion students in each program compare their proficiency in French to that of francophones their own age?
2. How do French immersion students in each program rate their French proficiency with reference to the difficulty of doing everyday activities (specific tasks) in French?
3. How do the students' self-assessments in (1) and (2) correlate with the tests of language proficiency given in each program?
4. Which results are independent of the program of instruction?

Thereafter, in the Discussion section, we examine the relationship between self-assessment and objective measures of proficiency, benchmark effects, and language learning and assessment theory.

4. RESULTS

1. *How do French immersion students in each respective program compare their proficiency in French to that of francophones their own age?*

On a scale of 1 to 3 where 3 represents 'about the same' as francophones, 2 'somewhat worse' than francophones, 1 'much worse' than francophones, the results are given in Table 3.

Table 3 indicates that self-assessments referenced to perceived francophone peer standards bear little relationship to program.¹ There are virtually no program differences in the case of self-assessments of writing and reading skills. Early immersion students appear marginally more likely to offer higher self-assessments of their speaking and listening skills than students in middle immersion.

*Table 3: Self-assessments of proficiency against francophone standards by program**

	Early immersion			Middle immersion			Kendall's tau-C
	Mean	SD	N	Mean	SD	N	
Speaking	2.16	0.57	189	2.00	0.63	279	.09
Listening	2.48	0.64	190	3.00	0.66	282	.12
Writing	2.12	0.73	188	2.00	0.67	281	.04
Reading	2.33	0.65	188	2.00	0.69	288	-.01

* Based on a 3-point scale: 3 = 'about the same'; 2 = 'somewhat worse'; 1 = 'much worse'. Program codes: Early immersion = 2; middle immersion = 1. Correlations are between program and self-assessments.

Table 4: Self-assessments of difficulty in doing specific tasks in French by program*

	Early immersion			Middle immersion			Kendall's tau-C
	Mean	SD	N	Mean	SD	N	
(a) Explain the plot of a mystery book or movie to someone face to face (<i>speaking</i>)	3.66	.95	201	3.27	1.01	316	.19
(b) Listen to and understand a French radio show (<i>listening</i>)	3.35	.99	203	3.10	.94	315	.13
(c) Explain the plot of a mystery book or movie to someone in a letter (<i>writing</i>)	3.86	.89	203	3.51	1.05	317	.18
(d) Read and understand a French newspaper or magazine (<i>reading</i>)	3.86	.79	202	3.56	.83	316	.20

* Based on a 5-point scale: 5 = 'without any difficulty'; 4 = 'with little difficulty'; 3 = 'with some difficulty'; 2 = 'with much difficulty'; 1 = 'not at all'. Program codes: early immersion = 1; middle immersion = 2. Correlations are between program and self-assessments.

2. *How do French immersion students in each program rate their French proficiency with reference to the difficulty of doing everyday activities in French?*

On a scale in which students state how much difficulty they experience in performing specific activities in French, a 5 represents 'without any difficulty', a 4 'with little difficulty', a 3 'with some difficulty', a 2 'with much difficulty', and 1 'not at all'. The results are given in Table 4.

Program differences are somewhat larger in the case of self-assessments of difficulty in doing everyday activities than was found for self-assessments in terms of perceived francophone peer abilities, and are, moreover, consistent across skill areas. Early immersion students have higher self-assessments than middle immersion students.

3. *How are the students' self-assessments in (1) and (2) related to specific measures of tested proficiency in each respective program?*

In order to answer this question, we examined the Kendall's tau-C correlations between the two respective measures of self-assessment and the objective proficiency measures. The results are given in Tables 5 and 6.

In general, there are weak correlations between the two respective measures of self-assessment and the objective measures of proficiency. The highest correlations (.25) are those between measures of self-assessment on 'specific tasks' and the literacy skills of reading and writing for the middle immersion students. In general, there are higher correlations for the 'specific tasks' benchmark than the more global 'francophone peer' benchmark.

4. *What results are independent of the program of instruction?*

It is noteworthy (see Figures 1 and 2 overleaf where the group means of each respective group have been plotted) that irrespective of the program of instruction, when 'francophone peer' is used as a benchmark, both groups demonstrate greater confidence in their receptive skills (listening and reading) than their productive skills (speaking and writing). However, when 'specific tasks' are used as

Table 5: Correlations between measures of self-assessment against francophone peers and measures of tested proficiency

	Early immersion	Middle immersion
Speaking: Fluency	.14	.12**
Listening: TCAC	.14*	.07
Writing: Global	-.01	.14***
Reading: Cloze	.16**	.19***

* <.05; ** <.01; *** <.001.

Table 6: Correlations between measures of self-assessment against specific tasks in French and measures of tested proficiency

	Early immersion	Middle immersion
Speaking: Fluency	.19**	.17**
Listening: TCAC	.23***	.22***
Writing: Global	0.12*	.25***
Reading: Cloze	.20***	.25***

* <.05; ** <.01; *** <.001.

NB: The speaking test on which the fluency measure is based was administered to a simple random sample. While the other correlations are based on population data, and significance test results are not strictly appropriate, they have been provided as a source of additional information.

benchmarks, both groups demonstrate greater confidence in literacy tasks (reading and writing) than oral tasks (speaking and listening).

5. DISCUSSION

We introduce the discussion of our results by reiterating the research questions outlined at the beginning of this paper: First, for the French immersion population, to what extent is self-assessment a valid and reliable indicator of tested proficiency? Second, how are the correlations influenced by the type of benchmark used? Third, to what extent can self-assessment research inform our understanding of second language learning and assessment?

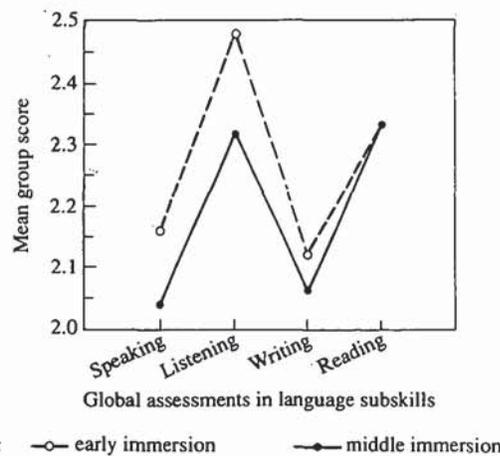


Figure 1: Self-assessments of proficiency against francophone peers

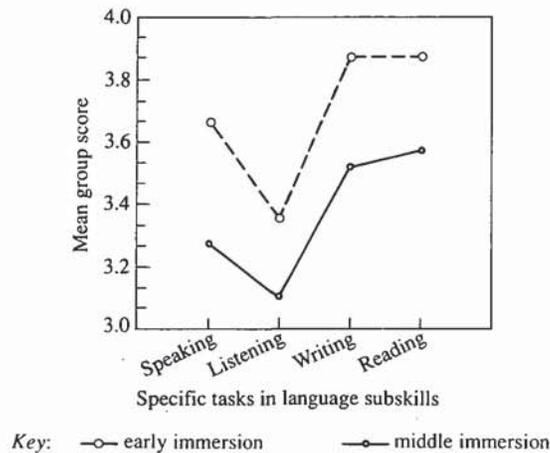


Figure 2: Self-assessments of proficiency against specific tasks in French

5.1 The validity and reliability of self-assessment

Notwithstanding the benchmark effects discussed in section 5.2 below, our results indicate that for both French immersion populations, self-assessment is only a weak indicator of tested proficiency. While no definitive explanations can be formulated, a number of reasons for this finding can be considered. First, the vast majority of French immersion students have little access to French or francophones outside the classroom. Over half of the students in the study reported that they had not had a conversation outside of class with a francophone in the month prior to the self-report (Hart, Lapkin, and Swain 1988). For this reason, many French immersion students do not have an authentic native speaker peer standard by which to compare their proficiency in French. In addition, in the month prior to the self-report, only approximately a third of the students had watched French TV or listened to French language radio (*ibid.*). Second, the kind of language training that French immersion children receive is initially relatively informal and experiential rather than formal and analytical (Allen, Swain, Harley, and Cummins 1990; McVey, Bonyun, Dicks, and Dionne 1990). The kind of training may differ from that of the adults in other self-assessment studies in which higher correlations with tested proficiency have been found (Oscarson 1978; Davidson and Henning 1985; LeBlanc and Painchaud 1985; Von Elek 1985).

5.2 Benchmark effects on self-assessment

The finding that a situational benchmark produces higher correlations with tested proficiency than a more global benchmark is consistent with the findings of LeBlanc and Painchaud (1985). In addition, the relative effectiveness of this benchmark, which focuses on the 'difficulty' represented by specific tasks in the

second language, lends support to the findings of Bachman and Palmer (1989) which lead them to argue that 'foreign/second language users may be more aware of the areas in which they have difficulty than they are of the areas they find easiest' (ibid.: 23). It must be borne in mind, however, that although the second benchmark used in our study is more effective than the first one, the correlations with tested proficiency remain weak.

5.3 *The locus of control in self-assessment*

When 'francophone peer' is used as a benchmark, both early and middle immersion students rate themselves as more proficient in receptive skills (listening and reading) than productive skills (speaking and writing). However, when 'specific tasks' are used as benchmarks, students in both programs rate themselves as more proficient in literacy tasks (reading and writing) than oral tasks (listening and speaking). This result could not have been anticipated on the basis of the available literature on self-assessment. For this reason, we have focused the greater part of our discussion on this finding.

Conventional wisdom in the field of second language learning suggests that students are more confident about their receptive skills than their productive skills. Spolsky's ninth condition for second language learning is illustrative of this observation.

Condition 9. Receptive Skills stronger than Productive condition (typical, graded): Receptive language skills (understanding speech or written text) usually develop before productive skills (speaking, writing) and usually develop to a higher level. (Spolsky 1989: 17)

Given Condition 9, it is not surprising that when students are asked to compare their French proficiency to that of francophone peers, students in both early and middle French immersion perceive their receptive skills to be more developed than their productive skills (Figure 1). What is surprising, however, is that when the students are given specific tasks as self-assessment benchmarks, they are no longer more confident about their receptive skills than their productive skills (Figure 2)—they are more confident about their literacy skills than their oral skills. Furthermore, while students in both groups are *most* confident about their global listening skills with respect to other 'global skills' (Figure 1), both groups are *least* confident about the given listening task with respect to other 'specific tasks' (Figure 2). Although these self-assessments bear only a weak relationship to tested proficiency, what is interesting is the perception of relative difficulty shared by *both* groups of students.

These results would seem to provide counter-examples to Spolsky's ninth condition for second language learning, as described above. We would like to argue, however, that while Condition 9 applies for global self-assessments of language proficiency (as indicated in Figure 1), Condition 9 can be qualified with reference to what we call the 'locus of control' in a communicative event. In our formulation, the locus of control is said to reside with the participant (or participants) who exercise dominant control over the rate of flow of information

in a communicative event. If a communicative event takes place in 'real time', i.e. when the language learner has little time to process information and cannot reflect on what is being communicated, the locus of control does not reside with the language learner. When the language learner can control the rate of flow of information in a communicative event and reflect on what is being communicated, the locus of control does reside with the language learner.

To illustrate: when a learner is engaged in encoding or decoding written material, the learner has the opportunity to reflect on the encoding or decoding process and attempt to develop the necessary 'schemata' (see Coady 1979; Carrell and Eisterhold 1983; Carrell 1983, 1989) to read or write a given document. However, when a learner is engaged in oral activities, the activities take place in real time, and the learner has little opportunity to reflect on the encoding and decoding process. Under these conditions, we suggest that the working memory of the learner can be overloaded, compromising the learner's control over cognitive processes that have not yet achieved full automaticity (see McLaughlin 1990, 1992). Thus while the locus of control in literate activities is in the language learner's favor, the locus of control in oral activities is not. Furthermore, in oral production (speaking), the learner has greater control over the rate of flow of information than in oral reception (listening) because the learner can speak slowly, pause, and make use of a variety of 'strategies' (see Canale and Swain 1980; Wenden and Rubin 1987) in order to enhance communication. Thus the locus of control is more favorable to the learner in oral production than oral reception.

The locus of control construct helps to explain the unexpected results in our study: When students were asked to assess the relative difficulty of the four second language tasks (Table 2), they indicated (Figure 2) that writing a letter in French or reading a French newspaper is easier than explaining a plot to someone face to face, or listening to a radio show. We argue that the locus of control is more favorable to them in the former two literacy tasks (in which there is maximal control over the rate of flow of information) than in the latter two oral tasks (in which there is minimal control over the rate of flow of information). Furthermore, we suggest that the learners found listening to a radio show the most difficult of the four tasks because in such a task they have less control over the rate of flow of information than they have for the other three tasks, including the oral production task. This lack of control is exacerbated by the fact that a radio broadcast does not provide any paralinguistic information that might enhance comprehension (see Pennycook 1985). In sum, the locus of control construct offers an adequate explanation for the results in Figure 2.

The locus of control construct is theoretically important to research on second language learning and assessment. First, it helps to explain why learners' perceptions of the relative difficulty of receptive and productive communicative tasks may not be consistent with learners' global assessments of receptive and productive language skills. When the locus of control resides with the language learner, there is a concomitant effect on the language learner's perception of the relative difficulty of a particular communicative task and consequently the

learner's self-assessments of language proficiency with respect to that activity. We suggest that a favorable locus of control leads to the perception that a communicative activity is relatively 'easy'; this in turn leads to relatively high self-assessments of ability. An unfavorable locus of control leads to the perception that a communicative activity is relatively 'difficult'; this in turn leads to relatively low self-assessments of ability. In particular, the construct serves to qualify Spolsky's ninth condition of second language learning, which can be refined as follows:

Condition 9: Receptive Skills stronger than Productive condition, when the locus of control resides with the language learner (typical, graded).

Second, the locus of control is a useful construct to explain other anomalous results in the self-assessment literature. In their research on the development and validation of the semi-direct Portuguese Speaking Test, Stansfield, Kenyon, Paiva, Doyle, Ulsh, and Cowles (1990) found that although students had approximately the same results whether they were given a live oral interview or a simulated taped interview, 90 per cent of the students perceived the taped test as more difficult than the live interview. During the taped test, the students had no control over the rate of flow of information in the test as there were timed pauses in the test; in the live interview, the interviewer could adapt the conversational pauses in accordance with the learner's responses. We would argue that the live interview was perceived to be relatively easier than the taped test because the locus of control was more favorable to the students in the live interview than in the semi-direct test.

Third, the locus of control is a construct that can be understood to operate at the interface between language assessment and SLA research. It is instrumental in explaining under what conditions a learner might assess some communicative tasks in the target language to be more difficult than others, and it is theoretically motivated by psycholinguistic research in SLA. Further research is needed to investigate the adequacy of this claim.

6. CONCLUSION

In our study of the self-assessments of French immersion students, the correlations between self-assessments of language proficiency as compared to measures of tested proficiency were weak, far weaker than many equivalent correlations in the literature for adults. This suggests that caution should be exercised in the interpretation of the self-assessments of language proficiency made by French immersion students. Possible reasons for this finding might include the fact that French immersion students do not have an authentic francophone peer standard by which to compare their own linguistic skills and that the relatively informal, experiential learning of French immersion students in their initial years does not promote the self-reflection characteristic of more formal, analytical adult language classes.

For both groups of French immersion students, the benchmarks that were used for the purposes of comparison made a difference to the correlations

between self-assessments and tested proficiency. In both programs, there were higher correlations between the 'specific tasks' benchmark and measures of tested proficiency than there were for the more global 'francophone peer' benchmark and measures of tested proficiency. This suggests that the more specific and focused a self-assessment instrument is, the greater the likelihood that there will be higher correlations with objective measures of proficiency.

When the benchmark was a 'francophone peer', students in both early and middle immersion ranked themselves closer to francophones on receptive skills than productive skills. When the benchmark was 'specific tasks', both groups of learners agreed that the oral tasks of speaking and listening were more difficult than the literacy tasks of reading and writing, with the oral reception task representing the greatest difficulty of all four tasks. We explained this anomaly with reference to our formulation of the 'locus of control' in a communicative event.

In sum, our study suggests that research on self-assessment in language learning can include not only a consideration of self-assessment as a valid and reliable measure of language proficiency, but a consideration of the specific conditions that might influence a language learner's self-assessments. Such conditions include the kind of benchmark used for the purposes of comparison, and the locus of control in a particular communicative event. Of particular significance to language educators, researchers, and testers is that the locus of control helps to provide a theoretical explanation for the anxiety that many language learners experience when they have to take formal language tests (see Peirce 1992). In such a communicative event, the locus of control does not reside with the language learner as the learner has little control over the rate at which information in the test must be processed. Indeed, the appeal of self-assessment as a measurement instrument is that it places the locus of control squarely in the learner's favor and should therefore be encouraged in regular classroom practice. Our study suggests, however, that the benchmarks used for the purposes of comparison should be authentic and situational, while focusing on the relative difficulty of communicative tasks.

(Revised version received March 1992)

ACKNOWLEDGEMENTS

This paper was completed while the first author held a Graduate Assistantship at the Ontario Institute for Studies in Education and a SSHRC Doctoral Fellowship. This support is gratefully acknowledged. Appreciation is also extended to Sharon Lapkin and Birgit Harley for useful comments on an earlier draft of this paper.

NOTE

¹ Strength of association between program type and self-assessment ratings is indicated by the magnitude of the rank-order correlation measure, Kendall's tau-C. Program is conceptualized as an ordinal variable in terms of magnitude of accumulated

instructional time in French. The tau-C, appropriate for rectangular tables, takes on a value from +1 to -1. Strictly speaking, the tau-C cannot be squared (as can the Pearson correlation coefficient) to yield the proportion of variation 'explained' by program. However, the Kendall's tau-C values are almost identical to Pearson correlation coefficient values for our data. The Pearson correlation would be strictly appropriate if we were prepared to make the assumption that the 'psychological distance' between any two adjacent self-assessment categories is the same; that is, that the difference between 'about the same' and 'somewhat worse' is the same as the difference between 'somewhat worse' and 'much worse'. It must additionally be assumed that the distances are the same for all respondents. It is highly unlikely, however, that these conditions are actually met. However, departures from them are also not likely to result in large, systematic distortions if the squared coefficient is used to approximate the proportion of variation explained.

REFERENCES

- Allen, P., M. Swain, B. Harley and J. Cummins.** 1990. 'Aspects of classroom treatment: toward a more comprehensive view of second language education' in B. Harley, P. Allen, J. Cummins, and M. Swain (eds.): *The Development of Second Language Proficiency*. Cambridge: Cambridge University Press.
- Bachman, L. F.** 1989. 'Language testing-SLA interfaces' in R. B. Kaplan (ed.): *Annual Review of Applied Linguistics, Volume 9, 1988*. Cambridge: Cambridge University Press.
- Bachman, L. and A. Palmer.** 1989. 'The construct validation of self-ratings of communicative language ability.' *Language Testing* 6/1: 14-29.
- Blanche, P. and B. J. Merino.** 1989. 'Self-assessment of foreign language skills: implications for teachers and researchers.' *Language Learning* 39/3: 313-40.
- Blue, G.** 1988. 'Self-assessment: the limits of learner independence.' *English Language Teaching Documents* 131: 100-18.
- Brown, H. D.** 1991. 'TESOL at twenty-five: What are the issues?' *TESOL Quarterly* 25/2: 245-77.
- Canale, M. and M. Swain.** 1980. 'Theoretical bases of communicative approaches to second language teaching and testing.' *Applied Linguistics* 1/1: 1-47.
- Carrell, P.** 1983. 'Background knowledge in second language comprehension.' *Language Learning and Communication* 2/1: 25-34.
- Carrell, P.** 1989. 'SLA and classroom instruction: Reading' in R. B. Kaplan (ed.): *Annual Review of Applied Linguistics, Volume 9, 1988*. Cambridge: Cambridge University Press.
- Carrell, P. and J. Eisterhold.** 1983. 'Schema theory and ESL reading pedagogy.' *TESOL Quarterly* 17/4: 553-73.
- Coady, J.** 1979. 'A psycholinguistic model of the ESL reader' in R. Mackay, B. Barkman, and R. Jordan (eds.): *Reading in a Second Language*. Rowley, Mass.: Newbury House.
- Davidson, F. and G. Henning.** 1985. 'A self-rating scale of English difficulty: Rasch scalar analysis.' *Language Testing* 2: 164-79.
- Dickinson, L.** 1987. *Self-instruction in Language Learning*. Cambridge: Cambridge University Press.
- Hart, D., S. Lapkin and M. Swain.** 1988. *Early and Middle Immersion Programs: Linguistic Outcomes and Social Character*. Toronto: Ontario Institute for Studies in Education.
- Haughton, G. and L. Dickinson.** 1988. 'Collaborative assessment by master's candidates

- in a tutor based system.' Paper presented at the 10th Annual Language Testing Research Colloquium, Chicago, Ill.
- Henner-Stanchina, C. and H. Holec.** 1985. 'Evaluation in an autonomous learning scheme' in P. Riley (ed.): *Discourse and Learning*. New York: Longman Press.
- Holec, H.** 1980. *Autonomy and Foreign Language Learning*. Strasbourg: Council of Europe.
- Janssen-van Dieten, A.** 1989. 'The development of a test of Dutch as a second language: the validity of self-assessments by inexperienced subjects.' *Language Testing* 6/1: 30-46.
- Lapkin, S., D. Hart and M. Swain.** 1991. "Early" and "middle" French immersion programs: French language outcomes.' *The Canadian Modern Language Review* 48/1: 11-40.
- LeBlanc, R. and G. Painchaud.** 1985. 'Self-assessment as a second language placement instrument.' *TESOL Quarterly* 19/4: 673-87.
- McLaughlin, B.** 1990. 'Restructuring.' *Applied Linguistics* 11/2: 113-28.
- McLaughlin, B.** 1992. 'What can psychology tell us about the good language learner?' Paper presented at the American Association for Applied Linguistics conference in Seattle, Washington.
- McVey, M., R. Bonyun, J. Dicks and L. Dionne.** 1990. *Middle Immersion Review: A Comparison of Ottawa Board of Education Students in Three French Immersion Programmes in Grade 6 and Grade 8*. Research Paper 90-01A. Ottawa: Ottawa Board of Education (mimeo.)
- Naiman, N., H. Fröhlich, H. Stern, and A. Todesco.** 1978. *The Good Language Learner*. Toronto: OISE Press.
- Nunan, D.** 1988. *The Learner-centred Curriculum*. Cambridge: Cambridge University Press.
- Oscarson, M.** 1978. *Approaches to Self-assessment in Foreign Language Learning*. London: Pergamon.
- Oscarson, M.** 1984. *Self-assessment of Foreign Language Skills: A Survey of Research and Development Work*. Strasbourg: Council of Europe.
- Oscarson, M.** 1989. 'Self-assessment of language proficiency: rationale and applications.' *Language Testing* 6/1: 1-13.
- Peirce, B. N.** 1992. 'Demystifying the TOEFL reading test.' *TESOL Quarterly* 26/4.
- Pennycook, A.** 1985. 'Actions speak louder than words: Paralanguage, communication, and education.' *TESOL Quarterly* 19/2: 259-82.
- Ready-Morfitt, D.** 1991. 'The role and limitations of self-assessment in testing and research.' Paper presented at the 13th Annual Language Testing Research Colloquium, Princeton, NJ.
- Riley, P.** (ed.) 1985. *Discourse and Learning*. New York: Longman.
- Rubin, J.** 1975. 'What the "good language learner" can teach us.' *TESOL Quarterly* 9/1: 41-50.
- Spolsky, B.** 1989. *Conditions for Second Language Learning*. Oxford: Oxford University Press.
- Stansfield, C., D. Kenyon, R. Paiva, F. Doyle, I. Ulsh, and M. Cowles.** 1990. 'The development and validation of the Portuguese Speaking Test'. *Hispania* 73: 641-51.
- Swain, M.** (in press). 'Using assessment information in French immersion programs' in E. Shohamy and R. Walton (eds.): *Language Assessment for Feedback: Testing and Other Strategies*. Washington, DC: NFLC.
- Von Elek, T.** 1985. 'A test of Swedish as a second language' in Y. P. Lee, C. Y. Y. Fok, R.

Lord, and G. Low (eds.): *New Directions in Language Testing*. Oxford: Pergamon Press.

Wenden, A. and J. Rubin. (eds.) 1987. *Learner Strategies in Language Learning*. Englewood Cliffs, NJ: Prentice Hall.

APPENDIX

Tested proficiency on different instruments measuring speaking, listening, reading, and writing

	Mean	SD	N	Median
<i>Speaking: Oral Fluency Rating by Program (Max. =3)</i>				
Early immersion	1.72	0.61	75	2.00
Middle immersion	1.17	0.78	123	1.00
<i>Listening: Listening Comprehension (TCAC) by Program (Max. =15)</i>				
Early immersion	11.09	2.43	200	11.00
Middle immersion	8.42	3.07	319	8.00
<i>Writing: Global 'Good writing' Scores by Program (Max. =2)</i>				
Early immersion	1.17	0.54	198	1
Middle immersion	.90	0.56	316	1
<i>Reading: Cloze Results (TMTC) by Program (Max. =25)</i>				
Early immersion	15.69	3.56	200	16.00
Middle immersion	11.31	5.37	319	12.00