PART I:
THE NATURE OF LANGUAGE PROFICIENCY

1 *The nature of language proficiency*

Birgit Harley,
Jim Cummins,
Merrill Swain,
and Patrick Allen

Theoretical background

Since ancient times, philosophers and psychologists have debated appropriate ways of conceptualizing the nature of language proficiency and its relationship to other constructs (e.g., "intelligence"). The issue is not just an abstract theoretical question, but one that is central to the resolution of a variety of applied educational issues. Educational policies based on hypotheses about how language proficiency develops as a function of different classroom treatments, different experiences in the environment, and different social contexts clearly imply certain conceptions about what it means to be proficient in a language.

The context for our investigation of the nature of language proficiency will be described briefly in relation to the applied, theoretical, and methodological considerations that influenced the conceptualization and implementation of the study.

Applied issues

Several examples will serve to illustrate the point that the conceptualization of the nature of language proficiency has a major impact on a variety of practical and policy issues in education. In the area of language teaching methodology, for example, the predominant emphasis until recently has been on the teaching of grammar. The implicit conception of language proficiency, as it has been operationalized in second language classrooms, has entailed viewing proficiency as little more than grammar and lexis. The recent movement toward communicative language teaching has been associated with a broader view of language that includes not just its grammatical aspects, but also the ability to use language appropriately in different contexts and the ability to organize one's thoughts through language. That is to say,

7

the recent emphasis on communication in language teaching is expressed in attempts to develop students' sociolinguistic and discourse competencies in addition to their grammatical competence. In short, the conception of what it means to be proficient in a language has expanded significantly.

In a very different context, it is clear that in bilingual programs for minority students in the United States, alternative conceptualizations of the nature of language proficiency have played a major role in policy and practice. A characteristic of most of these programs has been the establishment of entry and exit criteria whereby students must be declared of limited proficiency in English to enter the program and of sufficient proficiency in English to follow regular all-English instruction in order to exit from the program. One of the most controversial aspects of the bilingual education policy in the United States has been determining exactly what constitutes sufficient proficiency for the student to survive academically in a regular English-language classroom. In perhaps the majority of bilingual programs, what has been considered as "full English proficiency" amounts essentially to fluency in English; that is, the ability to function adequately in face-to-face situations and use English appropriately in a conversational context. However, the evidence suggests that many children who have exited from bilingual into all-English programs continue to experience academic difficulties in English (see Cummins 1984 for a review). This raises the question of how different aspects of language proficiency (e.g., grammatical, discourse, and sociolinguistic competencies) are related to the ability to manipulate language in academic contexts.

A similar issue arises in the area of psychological assessment. When minority students experience academic difficulties, educators frequently wonder whether the student has some learning disability or simply an inadequate grasp of English. Such students may appear to have overcome difficulties in English since they frequently understand and speak English relatively well. However, when they take IQ and other psychological tests in English, they often show more poorly developed verbal than performance abilities. This has led to many minority students being labeled as "learning disabled" and getting a one-way ticket into special education classes. Ortiz and Yates (1983), for example, report that Hispanic minority students in Texas were overrepresented by 300 percent in the "learning disability" category. Once again, the problem is directly related to the way in which language proficiency has been conceptualized – specifically, the extrapolation from conversational fluency in English to overall proficiency in the language and the judgment that students have sufficient proficiency in English so that verbal IQ tests will not discriminate against them on the basis of language.

Theoretical issues

When we began our investigation of the nature of language proficiency, two stages were evident in the way this issue had been conceptualized and investigated.

The first stage was essentially a continuation of factor-analytic work carried out between 1920 and 1970 by researchers such as Thurstone, Carroll, Guilford, Spearman, Burt, and Jensen, whose aim was to discover the basic structure of mental abilities, including language abilities. John Oller was a principal figure associated with this first stage. What distinguished his work from that of the earlier factor-analytic investigators was that its theoretical basis derived from applied linguistics rather than from psychometrics. Thus, although Oller employed essentially the same analytic methods, namely, variations of exploratory factor analysis, and his results were similar to those of several other investigators (Burt, Spearman, Jensen) in showing a strong general factor that incorporated both language and nonlanguage abilities, his empirical results were interpreted within a coherent framework of relatively sophisticated theoretical constructs (e.g., the pragmatic expectancy grammar) derived from applied linguistics (see Cummins 1984; Oller 1983).

Oller's (1981) provocative claim on the basis of his research and theory that language proficiency was largely indistinguishable from intelligence spurred considerable theoretical controversy and focused further research on the nature of language proficiency. The second stage of research into this issue was characterized by the use of a much more sophisticated form of factor analysis than that used by Oller. Confirmatory factor analysis allowed the explicit testing of theoretical models and was first applied to the field of language proficiency by Purcell (1983) and Bachman and Palmer (1982). Bachman and Palmer's work drew on the theoretical model of communicative competence developed by Canale and Swain (1980). Their results showed clearly that more than just one general factor could be extracted from language proficiency data. In fact, both trait and method factors could be distinguished. Taking into account the work of Bachman and Palmer and also the cogent criticisms of other researchers (see Oller 1983), Oller admitted that his earlier position was overstated and that there probably were nontrivial dimensions of language proficiency in addition to a general factor.

Two frameworks for conceptualizing the nature of language proficiency were particularly influential in the design of our study. The first was the communicative competence framework developed by Canale and Swain (1980). This framework initially distinguished grammatical competence, sociolinguistic competence, and strategic competence. The framework was later refined by Canale (1983) into grammatical, discourse, sociolinguistic, and strategic competence. The framework was

designed primarily to facilitate the process of curriculum development and language assessment in second language teaching.

The second framework involved a distinction between the use of language in context-embedded and context-reduced situations (Cummins 1984). The former is typical of face-to-face interactions where the communication is supported by a range of contextual cues, while the latter is typical of many academic contexts and involves primarily linguistic cues to meaning.

The integration of these two frameworks resulted in a 3×3 matrix comprised of measures of grammatical, discourse, and sociolinguistic competence assessed in oral and written productive modes and by multiple-choice written tests. Although any testing situation is likely to be less context-embedded than naturalistic, face-to-face interaction, the oral measures were conceived as being relatively more context-embedded than the written measures, which were more typical of academic (context-reduced) assessment procedures.

Methodological issues

Bachman and Palmer's work had clearly demonstrated the utility of confirmatory factor analysis, and this procedure was therefore adopted as a primary means of testing hypotheses about interrelationships among components of language proficiency. We were also aware that hypotheses about the nature of proficiency could be tested through other means than confirmatory factor analysis. For example, comparison between native speakers of a language and second language learners could reveal which aspects of second language (L2) proficiency were acquired most rapidly and help elucidate the ways in which proficiency becomes developmentally differentiated.

Similarly, an investigation of issues such as the relationship between different aspects of first and second language (L1 and L2) proficiency or the relationship between age and L2 acquisition could point to significant implications for the conceptualization of the nature of proficiency. For example, if discourse skills were strongly related across languages but grammatical skills were not, this implies a distinction between grammar and discourse and generates hypotheses regarding the causes of the cross-lingual discourse relationships. For example, it might be that discourse is more strongly related to general cognitive abilities than is true for grammar. Similarly, differential acquisition rates for different aspects of proficiency by older and younger learners can provide evidence relevant to theoretical conceptualizations of the nature of proficiency. Thus, a variety of strategies was employed to investigate the interrelationships among components of the construct of language proficiency.

Large-scale proficiency study

The primary purpose of the large-scale proficiency study (Allen et al. 1983; Swain 1985) was to determine whether the three hypothesized traits, representing key components of language proficiency, could be empirically distinguished. It was hypothesized that grammatical, discourse, and sociolinguistic competence (Canale and Swain 1980; Canale 1983) would emerge as distinct components of second language proficiency which may be differentially manifested under different task conditions. A secondary purpose of the study was to develop a set of exemplary test items and scoring procedures that could be used, or modified for use, in further studies involving the measurement of the hypothesized traits. A final purpose was to examine the target language proficiency of the second language learners tested in relation to that of native speakers.

Sample and design

A total of 198 students was involved in the study, including 175 grade 6 early French immersion students from the Ottawa region, and 23 grade 6 native speakers from a regular francophone school in Montreal. The immersion students had received 100 percent of their schooling in French in kindergarten to grade 2 or 3, after which they had been taught in English for a gradually increasing portion of each day. At the time of testing, about 50 percent of their school subjects were being taught in French, and the other 50 percent in English. This sample of classroom second language learners was selected because of the theoretically interesting and educationally innovative nature of their intensive school-based language learning experience, and because they were at an age where they were sufficiently proficient in the second language to be able to cope with a wide range of language tasks.

A multitrait-multimethod analysis was used to examine the extent to which grammatical, discourse, and sociolinguistic dimensions of the immersion students' French proficiency were distinguishable. To measure proficiency on each trait, three methods were used: oral production, multiple-choice, and written production. A matrix with 9 test cells was created, consisting of 3 tests of grammar, 3 of discourse, and 3 of sociolinguistics (see Table 1.1). The oral production task for each was administered to a subsample of 69 immersion students and 10 native speakers. The next three sections describe how the three traits were operationalized in the testing instruments.

12   *Birgit Harley et al.*

TABLE 1.1 OPERATIONALIZATION OF TRAITS IN
LARGE-SCALE PROFICIENCY STUDY

Trait / Method	*Grammar* Focus on grammatical accuracy within sentences	*Discourse* Focus on textual cohesion and coherence	*Sociolinguistic* Focus on social appropriateness of language use
Oral	Structured interview Scored for accuracy of verb morphology, prepositions, syntax	Story retelling and argumentation/ suasion Detailed ratings, e.g., identification, logical sequence, time orientation and global ratings for coherence	Role-play of speech acts: requests, offers, complaints Scored for ability to distinguish formal and informal register
Multiple choice	Sentence-level "select the correct form" exercise Involved verb morphology, prepositions, and other items (45 items)	Paragraph level "select the coherent sentence" exercise (29 items)	Speech-act-level "select the appropriate utterance" exercise (28 items)
Written composition	Narrative and letter of suasion Scored for accuracy of verb morphology, prepositions, syntax	Narrative and letter of suasion Detailed ratings much as for oral discourse and global rating for coherence	Formal request letter and informal note Scored for ability to distinguish formal and informal register

Operationalization of grammar trait

The grammar trait was measured in terms of accuracy in French mor-
phology and syntax, with an emphasis on verb and preposition usage.
The focus on verbs and prepositions was motivated by their importance
as subsystems of French grammar, and also by previous research, which
had indicated that these were areas of French grammar where one was
most likely to find variability among the immersion students. Note that
variability in performance on the measures was a statistical necessity for
factor analysis.

The grammar oral production task consisted of a guided individual
interview in which the interviewers' questions were designed to elicit a

variety of verb forms and prepositions in French, as well as some re-
sponses that were sufficiently elaborated to score for syntactic accuracy.
The content of the interview questions (e.g., favorite pastimes, trips
taken) was designed to focus the subject's attention on communication
rather than the code. Scoring was based on the student's ability to
produce grammatical forms and structures accurately in the context of
particular questions.

   The second grammar test was a written, group-administered multiple-
choice test. This test also assessed knowledge of the verb system, prep-
ositions, and other forms in sentence-level test items. The student's task
was to fill in the gap by selecting the correct response from three alter-
natives provided.

   The third grammar test, written production, consisted of two short
compositions – one a narrative and the other a letter of request. In
the narrative, the students were given a theme and an opening line
which prompted the use of past tense and plural verb forms. The re-
quest letter involved persuading an addressee to give the student per-
mission to carry out an action. This composition test was parallel to
a discourse written production test, which also involved a narrative
and a request letter. All four compositions were scored for grammati-
cal accuracy in verbs, prepositions, and other rules of syntax and
morphology.

Operationalization of discourse trait

The discourse trait was defined as the ability to produce and understand
coherent and cohesive text. Accuracy of cohesive elements (i.e., spe-
cific linguistic realizations of coherence) was included in the assess-
ment because errors in such elements tended to disturb the coherence
of a text.

   For the individual discourse oral production test, the student was
required to retell the story of a silent movie about a mole and a bulldozer.
In addition, there was a role-playing task in which the student had to
take the part of the mole and convince some addressees not to carry out
an action that would have been harmful to the mole. This argumentation-
cum-suasion task was designed to parallel a letter of suasion in the
discourse written production test. Discourse oral production was rated
on a series of five-point scales both globally in detail, focusing, for
example, on the student's ability to make clear and accurate reference
to characters, objects, and locations; to produce logically coherent text;
and to perform the basic tasks required.

   The second discourse test was a multiple-choice test consisting of short
written passages from each of which a sentence had been omitted. The

student was required to select from three alternatives the sentence that best fit the context of the passage.

The discourse written production test, like the grammar written production test, consisted of a narrative and a request letter. All four written production tasks were rated for discourse coherence and cohesion, mostly on the same kinds of features assessed in the discourse oral production test.

Operationalization of sociolinguistic trait

The sociolinguistic trait was defined as the ability to produce and recognize socially appropriate language in context. The individual oral production test involved a set of slides with taped descriptions. The slides represented situations of different levels of formality to which the student had to respond appropriately with a request, offer, or complaint. The objective of this test was to determine the extent to which students were able to shift register, using formal markers of politeness in formal situations and using them less often in informal situations. The selected formal markers were based on what native speakers were found to be using in the formal situations. The final score was a "difference" score; this was arrived at by taking the number of formal markers produced in informal situations and subtracting them from those produced in formal situations.

The items in the sociolinguistic multiple-choice test each involved three grammatically accurate choices for expressing a given sociolinguistic function. The student's task was to select the most appropriate choice in each situation. Scores were weighted according to the responses selected by native speakers.

The final test in the matrix, a sociolinguistic written production task, required the student to write a formal request letter and two informal notes, all of which can be categorized as directives. The request letter written as part of the discourse written production test was also scored for sociolinguistic proficiency. Scores were again difference scores, obtained by subtracting formal markers produced in the notes from those produced in the letters.

Reliability and generalizability of scores

The component within-test scores were combined to produce a single overall score for each of the 9 trait-method cells in the matrix. The composition of each of these overall scores was calculated to maximize validity and reliability. On the multiple-choice tests, the reliability of the immersion students' total scores ranged from .58 on the sociolinguistic test to .75 on the discourse test. Generalizability studies (Cronbach et

TABLE 1.2 CONFIRMATORY FACTOR ANALYSIS – LISREL

	Factor 1 General	Factor 2 Written	Uniqueness
Grammar oral production	.53	—	.72
Grammar multiple-choice	.49	.55	.47
Grammar written production	.68	.39	.38
Discourse oral production	.30	—	.91
Discourse multiple-choice	.41	.42	.65
Discourse written production	.20	.66	.52
Sociolinguistic oral production	.23	—	.95
Sociolinguistic multiple-choice	.47	.24	.72
Sociolinguistic written production	−.03	.49	.76

$X^2 = 14.13$, df $= 21$, p $= .864$

al. 1972) were conducted on those cells for which sufficient data were available: the sociolinguistic oral production test and the three written production tests. G-coefficients for these tests, based on the subsample of orally tested students, were comparable to the multiple-choice test reliabilities.

Testing the model of proficiency

In order to determine whether the three traits – grammatical, discourse, and sociolinguistic competence – could be distinguished empirically, two kinds of analyses were performed: factor analysis, and a comparison of the group means of the learners and native speakers.

The factor analysis based on the 69 orally tested immersion students failed to confirm the hypothesized three-trait structure of proficiency. Instead, confirmatory factor analysis by means of the LISREL program (Jöreskog and Sörbom 1978) produced a two-factor solution. One of these factors was interpretable as a general language proficiency factor (see Table 1.2); it had positive loadings from all cells in the 9-test matrix except for the sociolinguistic written production test. The highest loadings on this general factor were from the three grammatical tests. The second factor was interpretable as a written method factor; it had loadings from the three multiple-choice tests and from all three written production tests. The tests loading on this method factor appeared to be tapping the kind of literacy-oriented linguistic proficiency that is typically learned in classrooms.

A different kind of result emerged from comparisons of immersion and native speaker scores on the various tests. On all three grammar tests, the immersion students' mean score was considerably lower than that of the native speakers, and they also scored generally lower on the sociolinguistic tests than did the native speakers. On the discourse tasks,