DOCUMENT RESUME

ED 226 581 FL 013 475

AUTHOR McLean, Leslie D.; And Others
TITLE Cultural Ambassadors: Monitors in Core-French Classes. Informal Series/48.
INSTITUTION Ontario Inst. for Studies in Education, Toronto.
SPONS AGENCY Canadian Council of Ministers of Education, Toronto (Ontario).; Department of the Secretary of State, Ottawa (Ontario).
REPORT NO ISBN-0-7744-5060-6
PUB DATE 83
NOTE 120p.
AVAILABLE FROM OISE Press, 252 Bloor St. West, Toronto, Ontario M5S 1V6, Canada ($10.50).
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS Bilingual Students; *Cultural Awareness; Elementary Secondary Education; *Fles; *French; Language Attitudes; *Language Skills; Learning Experience; Official Languages; *Program Evaluation; Second Language Learning; Student Attitudes; Student Evaluation; *Teacher Aides
IDENTIFIERS *Canada

ABSTRACT

Beginning core-French students who had contact with part-time French monitors and students who did not were compared in five Canadian provinces in 1979-1980 as part of an evaluation of the Second Language Monitor Program. Twenty-eight schools and 56 classes were studied in Saskatchewan, Prince Edward Island, and Nova Scotia, along with 5 schools and 10 classes in Manitoba and Ontario. Half of the sample had monitors for the grades 5 through 9 French classes, which were primarily first or second-year studies. Each student was interviewed for about 30 minutes in both French and English, and a standardized test of French listening comprehension was administered to determine speaking and listening competence, cultural knowledge, quantity and quality of experiences in French class, and attitude toward learning French. Students who had monitor contact reported many more experiences in French class, and their experiences were judged of higher quality than those of control students. While monitors did not appear to provide experiences that increase competence in any conventional way, it is concluded that they add desirable and unique cultural elements. The English and French interview questionnaires and student background information questionnaire are appended, along with a checklist and instructions for interviewers and a classroom observation coding scheme. (SW)

CULTURAL AMBASSADORS

MONITORS IN CORE-FRENCH CLASSES

Leslie D. McLean
Ross E. Traub
Vincent A. Gaudino

This research was a joint project of the Council of Ministers of Education, Canada and the Department of the Secretary of State, Canada.

Informal Series/48

OISE Press
The Ontario Institute for Studies in Education

The Ontario Institute for Studies in Education has three prime functions: to conduct programs of graduate study in education, to undertake research in education, and to assist in the implementation of the findings of educational studies. The Institute is a college chartered by an Act of the Ontario Legislature in 1965. It is affiliated with the University of Toronto for graduate studies purposes.

The publications program of the Institute has been established to make available information and materials arising from studies in education, to foster the spirit of critical inquiry, and to provide a forum for the exchange of ideas about education. The opinions expressed should be viewed as those of the contributors.

252 Bloor Street West
Toronto, Ontario
M5S 1V6

CANADIAN CATALOGUING IN PUBLICATION DATA

McLean, Leslie D.
Cultural ambassadors : monitors in core-French classes

(Informal series / Ontario Institute for Studies in Education ; 48)
Co-published by Ontario Institute for Studies in Education.
Bibliography: p.
ISBN 0-7744-5060-6

1. Monitorial system of education - Evaluation. 2. French language - Study and teaching - Canada - Non-French-speaking students.* I. Traub, Ross E., 1938- II. Gaudino, Vincent. III. Ontario Institute for Studies in Education. IV. Title. V. Series: Informal series (Ontario Institute for Studies in Education) ; 48.

LB1029.M7M34 371.1'4124'0971 C83-094147-9

ISBN 0-7744-5060-6 Printed in Canada
1 2 3 4 5 XX 78 68 58 48 38

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The Second Language Monitor Program exists to promote the learning and use of Canada's official languages. But does it achieve this objective? A study was undertaken during the 1979-80 school year to answer this question, at least for a small part of the Program. The study was restricted to French Monitors who work part time; these individuals are in schools the equivalent of one day per week. Another restriction on the study was that the Monitors be employed in core-French classes; these classes study French for a short time each day, usually 20 minutes.

A substantial, carefully balanced sample of students (N=324) was obtained in three provinces--Saskatchewan, Prince Edward Island, and Nova Scotia. Twenty-eight schools and 56 classes were studied in these three provinces, and a further five schools and 10 classes in Manitoba and Ontario. Half of the sample was served by Monitors and the other half not (Monitor schools were assigned first by local decision, and Control schools were then matched to them). Classes were selected from grades 5-9, all but four of them enrolling students in their first or second year of French study.

Each student was interviewed individually for about 30 minutes, and a standardized test of French Listening Comprehension was given in each class in May 1980, after almost all Monitors had completed their work for the year. Interviewers were fluently bilingual and had attended a three-day training and practice session in Toronto to prepare to use the structured interview. Roughly half of the interview was in English and half in French, with a few instructions for the French tasks given in English. Interviewers made both detailed ratings of small tasks during the interview and global ratings of

French competence at the end. Attitude, knowledge, and experience scores were derived from the English interview. Each Interviewer worked in only one province.

Data from Saskatchewan, Prince Edward Island, and Nova Scotia were analysed in a multivariate analysis of variance that reflected the hierarchical design (students in classes in schools in provinces) and the multiple outcome measures.

Monitor and Control groups proved to be well matched on general characteristics. The Monitor groups, however, reported that they had more and better experiences in French. The Control groups were rated marginally higher on global competence, while the sums of the detailed ratings were marginally, but not statistically significantly, higher for Monitor groups. It seemed that interviewers in two provinces applied more traditional criteria than those in the third, favoring the Controls.

Monitors do not appear to provide experiences that increase competence in any conventional way, and if the ratings of two judges are accepted, Monitor students may exhibit slightly lower conventional competence. Monitors do add desirable and unique cultural elements, justifying their designation as cultural ambassadors.

1. BACKGROUND AND PURPOSE OF THE EVALUATION

The Second Language Monitor Program was introduced in 1973 to promote the learning and use of Canada's two official languages. University students working under the direction and supervision of professional second-language teachers take part in non-teaching activities at the elementary, secondary, and post-secondary levels. These "Monitors" work in schools for six to eight hours a week, and in return they receive a grant (approximately $3000 in 1977-78) plus the cost of return transportation from their home province to the host province. Monitors must ordinarily leave their province and must always enroll in a post-secondary institution.

The program is funded by the Department of the Secretary of State (Ottawa) and is administered jointly by a committee of provincial coordinators and a national coordinator in the Secretariat of the Council of Ministers of Education, Canada.

In April 1979, the Secretariat issued a request for proposals to evaluate the French-language segment of the Program. Proposals were screened by the Joint Evaluation Committee of the Council, and the Ontario Institute for Studies in Education (OISE) was selected to carry out the evaluation. This is OISE's report.

It was intended that the evaluation touch all aspects of the Program. Effects on students in schools served by Monitors was to receive priority, and benefits, if any, to Monitors were not part of the study. The high demand for Monitors suggested that school officials felt the Program was worthwhile and that it did not impose unreasonable demands. In responding to annual requests for feedback, teachers and officials involved in the

Program consistently reported to the Council of Ministers that Monitors played a significant role in promoting the learning and use of Canada's official languages. What was needed, therefore, was systematic and objective evidence of impact on students.

For reasons of economy and efficiency, the Joint Evaluation Committee decided that attention in this evaluation would be focussed on students who had worked with Monitors in core French classes during the students' first or second year of French study.[1] This decision shaped the evaluation to a very great extent, as detailed below.

Within the groups of students selected for study, attention was concentrated on competence in understanding and speaking French, on knowledge of French culture, and on attitudes toward French and French culture. In addition, observations were made in French classes in the attempt to document, if possible, the presence of Monitor-induced effects upon the learning environment of the French classroom. The same measurements and observations were made in a sample of classes in schools where no French Monitors had been assigned (but which were very much alike otherwise) in order to assess whether or not there were any differences that might be attributed to the work of the Monitors.

In summary, this is an evaluation of the part of the Second Language Monitor Program that involves grades 5-9 core French classes where students are in their first few years of language study. A survey at the beginning of the evaluation (details of which are reported below) suggested that classes such as these make up about 14% of the program. The primary goal was to discover whether or not there were measurable effects on students of the activities conducted by the Monitors.

The Monitor Program in Early Core French

The generally valid descriptions of the Program have already been given--each Monitor spends six to eight hours per week, working in one or more schools, supervised and directed by a qualified teacher of French as a second language. Regulations specify that Monitors are not to be used in place of regular teachers. French Monitors are university students whose

---

[1] Classes are designated "core" if they meet no more than 20 to 40 minutes per day, five days a week. Classes meeting 40 or more minutes per day, up to half-time are often called "extended," while 50 or more minutes is regarded as a form of "immersion."

first language is French (the great majority come from Quebec), but beyond that they differ widely, being enrolled in the full range of university undergraduate programs. A few are assigned to one school for the entire year, but most visit more than one--either changing after a period of time or rotating on a regular basis.

While total time spent learning is clearly important, recent research on second-language learning has shown that the nature and distribution of time are all important (see, e.g. Stern, 1978, and Lapkin, Swain, Kamin, and Hanna, 1980). When one assesses the amount and kinds of impact Monitors might have, it is important to know whether they increase the time students spend learning French or alter its character, or both.

Reports from provincial coordinators, teachers, Monitors, and students indicate that, in the vast majority of schools where Monitors work with core-French classes, students spend time on French in different ways but total time spent learning is not increased. Monitors most often work with small groups who are withdrawn from the regular French class. Some attempts were reported to launch French clubs or to have after-school activities, but both the Monitor and the school schedules seemed to rule out provision of extra time in this way.

The above description cannot be documented in a thorough way, since the evaluation was concentrated on student outcomes. Reports are so consistent, however, and the constraints under which Monitors operate in core-French programs are so similar, that we feel the description is accurate.

Very little can be said about the kinds of activities students experience with Monitors. In order to minimize possible rater bias, interviewers were not told which schools had Monitors and which did not, nor, of course, did they ask about Monitor activities. Monitors had completed their year's work before interviewers visited the schools, and only one Monitor agreed to our request to keep a log of activities. Any future study should give attention to Monitor activities.

Clearly, students ought to have more opportunity to speak French in a small group than a large one, but we don't know for sure that they did. We can only assume that the time spent with the Monitor was more intense for each student than it could be in the regular class. In cases where the Monitors worked with students withdrawn from a class, this intensity would also hold for those who stayed in class; in other words, there might

be a "monitor effect" indirectly caused by giving the remaining students a more intense time with the teacher.

Papers by Morrison, Walsh, Pawley, and Bonyun (1979) and Genesee (1979) have suggested that the intense time generated by early or late immersion could yield results superior to those achieved by core-French students who actually accumulated more time in class. No one really knows how long it takes to make what progress at what age and under what conditions, however, so it is difficult to say what one should reasonably expect the Monitors to achieve in terms of improved student competence. Two very recent studies (Hanna, Smith, McLean, and Stern, 1980; McLean, 1980), though supplying only weak evidence, suggest that our expectations should be very modest.

Hanna et al. (1980) found that about 50% of secondary school students showed modest gains on listening comprehension tests and oral proficiency interviews after an intensive four-week bilingual exchange experience, where a French English group spent two weeks in a French milieu and two weeks in an English one. Since that summer program added to the classroom experience, total time spent on language and the intensity of the experience exceeded that provided by Monitors.

From results of testing new French items in core-French classes having different amounts of instruction, McLean (1980) estimated the gains that might be attributed to an extra 180 or 360 hours of core French. Estimates ranged from 0 to 10% more items correct at the grade 6 level to 10% to 20% more correct at grade 10. In this evaluation of the Monitor program, a 10% gain (three or four items on the IEA Listening Comprehension Test,[2] for example) would be just detectable, but no students in core-French programs are able to work with Monitors the equivalent of an extra 180 hours during a single school year.

Thus, the main implication of recent research on second-language learning in a variety of core-French settings is that measurable effects of instruction can be expected to be small, relative to other sources of variation, and that such effects will be difficult to establish by statistical methods. When the teaching/learning experience is as small in time and variable in nature and intensity as Monitor schedules suggest

[2] IEA stands for International Association for the Evaluation of Educational Achievement.

it has to be, research on outcomes of education in general would predict that no differences attributable to Monitor activity would be found in such measures as test scores.

A modest, worthwhile language improvement goal was suggested by a provincial coordinator. "The effect of monitors," the coordinator said, "is to loosen the tongue and sensitize the ear." Operationally, one concludes that students should be more willing to speak French, should speak it more freely, and should exhibit better comprehension of spoken French.

Rationale for the Present Study

The design of this study was chosen to maximize chances of detecting Monitor effects, within the prevailing practical and financial constraints. On the one hand, the choice of beginning core-French classes was appropriate because of the number of Monitors so engaged, but, on the other hand, it is also where the effects are likely to be most dispersed. The study, therefore, was designed to case a wide "net" for effects, a net of the smallest mesh obtainable under the circumstances.

Students were interviewed individually in 33 schools (17 with Monitors and 16 without), in both English and French, in order to judge their competence in speaking and listening and to tap their knowledge of and feelings toward French language and culture. A standardized French listening test was given. In addition to the nearly 400 interviews with students, the interviewers observed classes and talked with teachers, submitting both quantitative reports (summary codings) and written descriptions. By these means, no important, general effect of the Program should have slipped through the net.

2. REQUIREMENTS OF THE CLIENT AND METHODOLOGICAL CONSIDERATIONS

The Joint Evaluation Committee imposed six constraints on the study. Three of these pertained to the grade levels of the students who were to be involved in the evaluation, the general nature of the program of French instruction they were experiencing and the length of time they have been studying French.

Specifically:

1. The focus of the evaluation was to be on students in grades 4 to 6.
2. The students were to be studying French in a "core" program, i.e. a program in which French instruction is provided every school day (usually) for a relatively brief period of time (e.g. 20 minutes).
3. The students were to be in their first or second year of the study of French.

The Sample and the Program

To discover whether or not it would be possible to work within these requirements, a cross-Canada survey was conducted during September and the first part of October 1979. The purpose of this survey was to determine the grade-level and program assignment of each part-time French-language Monitor. The survey results reported in Table 1 reveal that about two-thirds of the part-time Monitors had been assigned to core-French programs. Of these, however, fewer than one-fourth served programs for students in their first or second year of French-language instruction. Fewer still of the Monitors had been assigned to programs for students who were in grades 4 to 6 and also in their first or second year of French instruction.

Table 1

Actual or Estimated Number of Monitors Assigned to French Programs, by Province

Province	Core		Immersion	Total
	First or Second Year	Other		
Newfoundland	4*	8*	2*	14
Prince Edward Island	10*	4*	10	24
Nova Scotia	8*	22*	4	34
New Brunswick	2*	1*	2*	5
Quebec	0	40	—	40
Ontario	16*	128*	125	269
Manitoba	7	10	12	29
Saskatchewan	15*	15*	5*	35
Alberta	9*	46*	17*	72
British Columbia	13*	51*	23	87
Total	84	325	200	609
%	14	53	33	100

"First or Second Year" indicates first two years of French study.

*Estimated from comments of provincial and local coordinators, or from grade core begins, assuming equal distribution of Monitor over grades.

On the basis of the survey results, OISE (the contractor) presented the Joint Committee (the client) with a recommendation for modifying the first design requirement. This recommendation, which was accepted, was that the range of grades should be expanded to include grades 7-9. With this modification, the survey results suggested that it would be possible to satisfy requirements 1 to 3.

Of the remaining three constraints, two reflected a desire for results having a satisfactory degree of internal and external validity (Campbell and Stanley, 1963).

4. A matching design was to be employed in which students in contact with Monitors would be compared with students not in contact with Monitors.
5. The generalizability of any significant result was to be assessed.

A Comparative Study

In an attempt to meet these requirements, the contractor proposed, and the client accepted, a design that can be described as follows:

(a) A multistage sampling design would be employed in which the primary sampling unit would be the school board. The target number of boards was set at 10. They were to be widely distributed geographically; data would be collected in at least three different provinces. (The survey results suggested that it would be possible to meet this objective, and satisfy the client's desire for an assessment of the generalizability of Monitor effects.)

(b) The secondary sampling unit within each board would be the school. An attempt would be made to involve four schools from each board, for a total of 40 schools. Of the four schools per board, it was planned that two should be experimental in the sense that they employed part-time Monitors in the core-French programs of interest to the client. The other two schools were to be control schools; these would not enjoy the services of a Monitor.

(c) The schools within each board would be matched as well as possible in pairs. Each pair would include an experimental and a control school. The matching variables would be the socio-economic status of families in the neighborhoods served by the schools, the grade levels of the students in their first or second year of instruction in a core-French program, the

length of time devoted each week to core-French instruction, the training and experience of the French teachers, and the willingness of the teachers and the school administrations to accept the services of a Monitor.

(d) The tertiary sampling unit within schools would be the student. Because some of the data that were to be collected could not be obtained from all eligible students within each schooltime and resources did not permit this, a kind of matching of students would be attempted across pairs of schools, and these data would be obtained only from matched students. It was proposed that the eligible students in a school, whether the experimental or the control school of a pair, be stratified on the basis of sex; within each sex group, the students would be stratified by year of French instruction, first or second--this meant choosing students from different classes; within each sex-by-year-of-French-instruction group, the students would be ranked for achievement in French. Then one student from each third--top, middle, bottom--of rank order would be selected for study. This procedure would result in the selection of 12 students from each school.

In the foregoing description can be found the contractor's design objectives. Practical exigencies prevented the contractor from meeting all these objectives. Details of where they were not met, and why, are presented in two later sections of this report - The Actual Sample and its Selection, and Measuring the Match Between Schools.

The sixth and final requirement of the client dealt in a general way with the variables or measures on which the evaluation was to be based:

6. The measures were to be language proficiency or competence, attitude toward the learning of French, knowledge of French culture, and in-classroom behavior during French class.

The Measures

To operationalize this requirement, the contractor proceeded as follows:

French proficiency and competence were measures using two very different procedures--a standardized test of listening comprehension and a structured interview.

IEA Listening Comprehension Test

This test was designed for Population 1 of the IEA study of the teaching of French as a foreign language (Carroll, 1975). This instrument was chosen for use because it appeared to be an appropriate level of difficulty for

students in their first or second year of French. (Population 1 in the IEA studies included all 10-year-old students who were studying French [Carroll, 1975].)

The test and how it was administered and scored in this study can be described briefly as follows. The test requires approximately 20 minutes to administer and contains 35 multiple-choice items. Each item consists of a spoken question and four picture responses. The student chooses one picture in response to each question. All eligible students present on the day this test was administered were tested. A member of the school's staff, usually the French teacher, took responsibility for the administration. To standardize the procedure across schools, the questions were presented from a taped recording. Students' responses were scored by project staff. The score a student was assigned equalled the number of correct answers.

Individual Interview--French

The structured interview was prepared specifically for use in this study. The thirty items to this interview are given in Appendix A. The first 24 were in French, and students were expected to respond in French. For all but item 18, the student was requested to produce a response; item 18 asked the student only to read each of eight numbers in French. In items 25 to 27 the student was presented with a situation, described in English. The student was then asked to respond appropriately in French. The last three tasks of the interview, numbers 28 to 30, asked the student to repeat in French each of three French sentences read aloud by the interviewer. Each sentence was read twice before the student was asked to respond.

The standardized French interviews were conducted by trained project staff. During each different interview the same 30 items and opportunities for responding were presented and a standard record was prepared. For 26 of the first 27 items--item 18 excluded--the interviewer recorded two judgements:

1. The correctness of the response (in three categories--no response, incorrect response, correct response), and
2. The completeness of the correct response (in five categories--no response, single word, phrase, simple sentences or long sentence). (See Appendix A, The Interviews.)

In addition to these judgements, the interviewer made a note each time a student was prompted for a response. (For details, see Appendix A, Instructions to Interviewers.)

A different procedure was followed in recording a student's performance on item 18 and the final three tasks in the French interview. For item 18, a count was made of the number of numbers that the student read correctly. In addition, the interviewer judged the overall quality of response on a four-point scale.

The final three items in the questionnaire, it will be recalled, were "elicited imitation" tasks (see, e.g. Naiman, Frohlich, Stern, and Todesco, 1978). Each sentence the student was to repeat was divided into several parts--there were three parts to items 28 and 29 and four parts to item 30. The first recording made by the interviewer was of the number of parts of the sentence that the student *attempted* to repeat. The second recording was of the number of parts repeated correctly.

At the end of each French interview, the interviewer made five summary ratings of the student's performance (referred to below as "competency ratings"). The characteristics rated were pronunciation, fluency, grammar, vocabulary, and comprehension. Each was rated on a five-point scale, defined as follows:

Pronunciation
- 0 = often unintelligible
- 1 = obviously foreign but rarely unintelligible
- 2 = foreign but always intelligible
- 3 = sometimes foreign, always intelligible
- 4 = native-like

Fluency
- 0 = every utterance requires enormous obvious effort
- 1 = very hesitant, often long pauses before answering
- 2 = hesitant, must pause before answering
- 3 = rarely hesitates
- 4 = responds as well as in English

Grammar
- 0 = no control of syntax, often conveys wrong information
- 1 = fair control of syntax, can convey meaning of simple phrases, although there are frequent errors
- 2 = good control of syntax, conveys meaning well, can formulate simple sentences, there are still errors

3 = very good control, formulates more complex sentences, relatively few errors

4 = very good control, complex sentences, almost no errors

Vocabulary

0 = almost no vocabulary at all

1 = adequate for basic courtesy requirements

2 = simple social and school needs

3 = adequate for simple general conversations

4 = very good vocabulary, can handle routine social general conversations

Comprehension

0 = did not understand most of the questions asked, required frequent repetition, slow speech

1 = understood most of what was said, required slow speech, some repetition

2 = understood almost all of what was said, could follow slow speech with almost no repetition

3 = understood almost all of what was said at normal rate of speech, very few repetitions

4 = could follow all of what was said, no repetition required

In summary, 17 scores were derived from the information collected during a French interview. Four of these were based on the record of a student's responses to 23 of the first 24 interview items (item 18 excepted), those asking a response to a simple question. These scores were:

(1) the number of failures to respond (scale: 0-23)

(2) the number of correct responses (scale: 0-23)

(3) a quality score--the sum of the completeness-of-correct-response scores (scale: 0-92)

(4) the number of items for which the interviewer prompted the student in an effort to elicit a response (scale: 0-23)

A second group of four were:

(5)-(8) the same scores, but based on responses to the situations (items 25, 26, 27).

Each student's response to the remaining items--18, 28, 29, and 30--were used to derive four scores:

(9)-(10) number correct, and overall quality for responses to item 18

(11)-(12) the number of units attempted and the number correct for the elicited imitation tasks (items 28-30).

Finally, there were the global ratings:

(13)-(17) pronunciation, fluency, grammar, vocabulary, and comprehension.

Lower-bound estimates of reliability of the scales composed of more than one item are reported in Appendix B. These figures suggest that the scores derived from the French interview were highly reliable, at least as compared with the reliability estimates usually obtained for scores on standardized tests of achievement. Reliability estimates could not be obtained for the five summary ratings from the data that were collected

Individual Interview--English

Attitudes toward the learning of French, knowledge of French culture, and the quantity and quality of a student's experiences in French class were assessed from responses to 35 items presented in a 10-minute (approximately) structured interview conducted in English by the same interviewer. The same students who were interviewed to assess their competence in French also responded to the items of the English interview. (See Appendix A for a copy of the interview.) The English interview actually preceded the French interview. Because it was conducted in English it provided the interviewer with an opportunity to make the student feel comfortable before the assessment of competence in French was made.

For 23 of the items, the expected response was yes or no. Another nine items asked the student either to name as many examples as possible of a particular kind of activity or event or to indicate how frequently the French class engaged in a particular activity. The interviewer recorded the examples named to items of the former type and coded the frequency of an activity to items of the latter type. The scale for coding frequency had six points--never, once a month, twice a month, three times a month, once a week, and more than once a week. The remaining three items required the student to describe one thing that was liked about the French class, one thing that was disliked, and to give a reason for preferring one French class to another French class, if in fact the student had such a preference. (It was intended by this to find out whether or not students would state a preference for Monitor-assisted classes, if the student had experienced such classes.)

A student's responses to items 1 to 4 and 33 to 35 of the English

interview (see Appendix A) were used to obtain an attitude score. *Yes* responses to items 1, 4, and 33 to 35 each contributed one point to the score. Naming one thing that was liked about the French class also contributed one point. Thus the range of scores on this scale was -1 to 6.

A knowledge-of-French-culture score was derived from the responses to items 8, 9, and 16 by adding the codes assigned by the interviewer to the first two of these items, and adding one additional point if the student had been able to name a French holiday. The range of scores on this scale was from 0 to 7.

A student's responses to the remaining items in this interview were used as the basis for two scores, one reflecting the quality of the student's experiences in French class, the other reflecting the quantity of his or her experiences. The first score ranged from 0 to 17, and was obtained by counting the number of *Yes* responses coded by the interviewer for items 6, 10, 11, 13 to 15, 17, 19, 20, 22, 24, 25, 27 to 29, 31, and 32. The second score ranged from 0 to 35 and was obtained by summing the responses coded for items 7, 12, 18, 21, 23, 26, and 30.

The four scales derived from the items of the English interview do not achieve as well as might be desired the two associated objectives of the evaluation--to measure knowledge of French culture and attitudes toward French and French culture. The scale of knowledge of French culture was based on only three items, and thus sampled a very limited portion of the associated domain of knowledge. Results based on this scale are valid, nonetheless, although only limited generalizations can be based on them. In addition, it should be recognized that the scales assessing quality and quantity of experiences in French class also bear indirectly on knowledge of French culture. These scales assess dimensions of the opportunities students had in French class to learn about French culture.

The assessment of attitude toward French and French culture was least adequate of all. This was by design. Attitude must be measured by seeking information about a student's typical response. This type of response is easily faked by the student who can guess the intention of the interviewer. Also, typical attitude questionnaires, especially those that attempt to conceal the purpose of the questionnaire from the student, yield measurements of dubious quality, in part at least because the questions often deal

with situations, events, etc., outside the experience of the student. In the end, it was decided to concentrate on the student's attitude toward the learning of French. Questions about this can at least be related by the student to his or her experience in French class. Moreover, positive responses reflect an openness on the part of the student to continued learning of French and to continued acquisition of knowledge about French culture.

In summary, the English interview yielded the following scores:

(1) Attitude to Learning French (scale: -1 to 6)

(2) Knowledge of French Culture (scale: 0 to 7)

(3) Quality of Experiences in French Class (scale: 0 to 17)

(4) Quantity of Experiences in French Class (scale: 0 to 35)

Lower-bound estimates of reliability are reported in Appendix B for the scores derived from the English interview. These estimates are certainly smaller than the estimates usually obtained for attitude questionnaires. It should be noted, however, that these are lower-bound estimates of reliability, i.e., we can expect that the reliability of the scores is not less than these estimates. Also, three of the four scales are composed of few items, and one of these is composed of very few items indeed. It is well known that reliability is an increasing function of scale length. Enhanced reliability could be achieved by increasing the length of these scales. Finally, since results were aggregated to the classroom level, the reliability of individual scores was not as important as in studies of individual students. In other words, the group comparisons in this study should not be downrated on reliability grounds.

Classroom Observation

To obtain information on possible monitor-induced effects in the French class, it was decided, after observing several core-French classes, to concentrate on the interchange between teacher and students. The reason for this decision was that the principal method of instruction appeared to be question and answer, with the teacher asking a question either of the class as a whole or, more frequently, of an individual. These were observed to be rapid-fire sessions, with the next question following quickly upon receipt of a correct answer. At issue was whether or not observable characteristics of these exchanges were affected by the Monitors.

Because of the pace at which question-and-answer sessions are conducted, an observer can attend to only a few of the many characteristics of the exchanges between teacher and students. Consequently, a relatively simple classification scheme was devised for student responses. They were classified as group or individual responses. Each response from an individual if given in French was classified as correct or incorrect and rated for completeness (quality) on the same scale used in the French interview. If no response was made, the fact was noted, as was the occurrence of a response in English. In the event that a student's response was either in incorrect French or given in English, or if no response was forthcoming, note was taken of the teacher's next behavior--whether it was to prompt the student, to move to another student, or to provide an explanation in English. A prompt was further classified according to whether it was in English or French, and, if the latter, whether it was to elicit a response or to correct an incorrect response from the student.

3. THE ACTUAL SAMPLE AND ITS SELECTION

As of mid-October 1979, the pattern of Monitor assignments by grade level and program type was that described in Table 2. Approximately one-third of the Monitors had been assigned to immersion classes, and almost that number had been assigned at the high school level. The only provinces where a relatively large number of Monitors had been assigned to elementary and intermediate core-French classes were Prince Edward Island, Alberta, Nova Scotia, Saskatchewan, and British Columbia. It was not possible to estimate accurately how many of these Monitors had been assigned to work with students in their first or second year of the study of French. It was concluded, however, that approximately 85 Monitors (14%) had been so assigned.

This information was used to select the school boards that were invited to participate in the study. Several criteria were employed in making this selection. The boards that were chosen

(a) were clustered in several major centres

(b) apparently had assigned two or more Monitors to target core-French programs

(c) offered a choice of schools to match with Monitor-served schools, and

(d) were relatively near to boards employing full-time Monitors. (With this criterion satisfied, it was possible to reduce the costs involved in conducting the companion study of full-time Monitors.)

The initial selection of boards having been made, the provincial and local coordinators were contacted to arrange site visits to each board.

Table 2

Distribution of Monitors by Grade Level
Both Immersion and Core

Province	Grade Level K-6		7-9		10-13		Total
Newfoundland	2	4	—	4	—	4	14
Prince Edward Island	6	14	4	—	—	—	24
Nova Scotia	4	12	—	14	—	4	34
New Brunswick	2	—	—	3	—	—	5
Quebec	—	—	—	—	—	40	40
Ontario							
Ottawa	46	36	15	—	1	20	118
Toronto	41	4	—	5	—	37	87
Other	22	5	—	3	—	34	64
Manitoba	8	7	3	7	1	3	29
Saskatchewan	5	—	—	20	—	10	35
Alberta							
Calgary/Edmonton	17	15	—	7	—	5	44
Other*	—	—	—	16	—	12	28
British Columbia							
Coquitlam	9	—	—	5	—	3	17
Other*	14	38	—	11	—	7	70
Total	176	135	22	95	2	179	609
%	28.9	22.2	03.6	15.6	0.3	29.4	100
	51.1		19.2		29.7		

*Estimated

It was apparent after the first few site visits that the target population was smaller than had been estimated on the basis of the survey. There were two major reasons for this. The first was that all boards had a priority system for assigning Monitors to schools; in most cases the beginning grades of the core program had the lowest priority. If a board did not receive its expected quota of Monitors, the requests for Monitors to serve classes in the first two years of the core program were the first to be eliminated. The second reason was that overall policy on the starting grade for core French was not always applied by individual schools. For example, it was discovered during the course of a site visit that, while board policy specified that the core program start in grade 5, the Monitors had all been assigned to schools where the core program actually began in grade 3. Thus, it was discovered after the study had begun that there were no schools in this board that met the criteria for inclusion in the study.

Even by going into as many as five provinces, it was not possible to achieve the target of 40 schools, 20 with Monitors and 20 without Monitors, since the target population was relatively small and was dispersed over a very large geographic area. When a problem of any sort prevented one of the sample schools from participating in the study, it was usually not possible to find a replacement school in the same geographic area. To compound the problem, four schools that had initially agreed to participate decided in May of 1980--far too late to find replacements--that they could not. In one of these schools, the reason given was that the Monitor had been dismissed early in the year, hence the school could not be regarded as an experimental one. The reason given by the other schools for dropping out of the study was the "difficulty" of the test instruments; it was felt that they would provide a negative experience for the students.

In the end, the sample comprised 33 schools, 17 with Monitors and 16 without. Nine school boards were represented, and five provinces. The sample in three of the five provinces was of a substantial size (i.e. at least eight schools).

Sample Recruitment and School Characteristics

After a board had been chosen for inclusion in the study, initial contact was made with the provincial coordinator, informing him of this decision. This contact was followed by a letter to the five provincial coordinators

requesting their assistance in enlisting the cooperation of the designated boards in their provinces. A letter outlining the scope and purpose of the study was then sent to the local coordinators. All the provincial and local coordinators who were contacted agreed to help enlist schools. Site visits by a project officer were then arranged, usually through the office of the provincial coordinator.

During a site visit, the project officer usually met first with the provincial coordinator to obtain detailed information about operation of the Monitor Program in the province and then with the local coordinator to identify Monitor schools and to choose a Control school for each Monitor school.

The Control schools were matched as closely as possible to the Monitor schools. The basis for matching was the socio-economic status of the neighborhood served by the schools (as judged by the coordinators), the size of the core-French classes, the characteristics of French programs, and the training and experience of the French teachers. Table 3 contains information about each pair of schools in the sample on these characteristics. Inspection of this table reveals that matching was reasonably good on all variables except whether or not the teacher of core French was a Francophone. Very few teachers, only eight of 33, were Francophones.

Once the choice of schools in a board had been agreed upon, the project staff member and the local coordinator visited each school and met with the principal and French teacher(s). The purpose and scope of the study were explained. Then the target classes were identified, a description was given of how students within classes were to be chosen for the interviews, and a description was provided of the IEA Listening Test, the structured interview, the student questionnaire, and the classroom observation procedure. All but one school agreed at this point to participate in the study. A tentative timetable for the administration of the IEA Listening Test and the student questionnaire was then established with each school. It was stressed during this meeting that the results of individual students and the identity of teachers, schools, and boards would be kept strictly confidential.

Table 3

Characteristics of the Schools Participating in the Evaluation

					Class Characteristics				Program Characteristics			Teacher Characteristics		
Prov.	Board	School	Exp or Con.	SES	Grade	Class Size	Grade	Class Size	Days in Cycle	French Periods per Cycle	Min. of French per Cycle	Exper.	Special.	Franco.
1	1	1	E	1	7	10	8	9	6	3	135	x	x	
1	1	2	C	1	7	16	8	14	6	3	135	x		x
1	1	3	E	1	7	18	8	13	6	6	180		x	
1	1	4	C	1	7	23	8	19	6	5	200	x		
1	2	5	E	2	7	14	8	30	6	3	120	x	x	
1	2	6	C	2	7	17	8	16	6	4	120	x	x	
1	2	7	E	3	7	14	8	14	6	3	100	x	x	
1	2	8	C	3	7	20	8	29	6	3	105	x	x	
2	3	9	E	2	5	28	5	28	6	5	150	x	x	x
2	3	10	C	2	5	18	5	26	6	3	120	x	x	
2	3	11	E	2	5	30	5	28	6	3	105	x	x	x
3	4	12	E	4	5	30	5	26	5	4	100	x	x	
3	4	13	C	4	5	30	5	30	6	4	120	x	x	
3	4	14	E	5	5	26	5	16	6	4	100	x	x	
3	4	15	C	5	5	24	-	-	6	4	120	x	x	x
3	5	16	E	1	5	30	6	17	6	4	100		x	
3	5	17	C	1	5	9	6	15	6	6	120			
3	5	18	E	1	7	22	8	30	6	6	120	x		
3	5	19	C	1	7	13	8	12	6	4	135	x	x	
3	6	20	E	1	5	20	5	23	6	3	90	x		x
3	6	21	C	2	5	26	5	26	6	4	100	x	x	x
3	6	22	E	1	5	28	6	29	6	6	150	x	x	
3	6	23	C	1	5	28	6	29	6	6	150	x	x	

Table 3 (cont'd)

Characteristics of the Schools Participating in the Evaluation

					Class Characteristics				Program Characteristics			Teacher Characteristics		
Prov.	Board	School	Exp. or Con.	SES	Grade	Class Size	Grade	Class Size	Days in Cycle	French Periods per Cycle	Min. of French per Cycle	Exper.	Special.	Franco.
4	7	24	E	2	7	23	8	28	5	15	450	x	x	
4	7	25	C	2	7	23	8	28	5	15	450	x	x	
5	8	26	E	2	5	25	5	19	5	5	125	x	x	
5	8	27	C	2	5	27	5	30	5	4	100	x	x	
5	8	28	E	5	5	17	5	19	5	5	100	x	x	
5	8	29	C	5	5	31	5	13	5	4	85	x	x	
5	9	30	E	5	7	28	8	31	7	6	240	x	x	
5	9	31	C	5	7	31	8	31	7	6	240	x	x	
5	9	32	E	5	6	37	6	37	5	5	100	x	x	x
5	9	33	C	5	6	21	6	23	5	5	100	x	x	x

SES:

1 = Small Town/Rural

2 = Urban Middle Class

3 = Urban Upper Middle Class

4 = Urban, Inner City, Middle Class

5 = Suburban, Middle Class

4. METHOD

Development of Instruments and Training Interviewers

As noted previously, several data collection procedures were designed specifically for this study:

1. A student background questionnaire, designed to obtain information about age, sex, grade, etc. of each student in the sample
2. An English interview
3. A French interview
4. A classroom observation schedule

Copies of these instruments and instructions for their administration appear in Appendix A.

The interview and observation schedules were developed after the site visits to recruit schools had been made. During these visits, the project officer had occasion to observe several core-French classes. It became apparent during these observation sessions that the French interview would have to be undemanding, because of the level of French used in first- and second-year core-French classes.

Drafts of interview and observation schedules were pilot-tested in four classes in two schools in the Toronto area. Several revisions and additional trials of the instruments were required to obtain the final versions presented in Appendix A. The interview had to be made short enough to be completed in approximately 20 minutes (the time available for interviewing one student). The observation schedule had to include categories for coding classroom behavior that were both meaningful and usable.

The content of the English and French interviews was analysed for acceptability in relation to the curriculum of the core-French programs of the provinces and school districts within provinces that participated in the study. Also, most of the French coordinators in the participating boards and provinces reviewed the interviews and approved their use in the study.

As the instruments were being put in final form, interviewers were being recruited in each of the provinces involved in the study. All of the recruits--three men and two women--were qualified teachers of French.

In order to standardize the collection of data as fully as possible, the recruited interviewers--each of whom doubled as an observer--attended a three-day training session in Toronto. They first practised administering the interview schedules by interviewing one another and criticizing one another's efforts under the supervision of the project officer. Practice in completing the observation schedule was given by having the recruits view films and listen to related audiotapes of French classes. In addition, the interviewers were able to observe in two core-French classes and to interview at least one student. The results of these observation and interview sessions were analysed and discussed in order to develop a shared understanding of the procedures to be followed and the meaning of the ratings to be made.

Testing and Observation Procedures

During the month of April 1980, a contact person in each participating school was sent the proposed schedule of visits to the school by the interviewer and informed that the student background questionnaires and IEA tests were being sent under separate cover. At the earlier site visit by the project officer, it had been agreed that the French teachers in each school would administer the background questionnaire to their students as soon as convenient after it had been received and that they would administer the IEA test between the 15th and 30th of May, on a day when the interviewer would not be in the school.

The interviewers were scheduled to visit the schools between May 12 and June 10, 1980. The procedure was to telephone the contact person in the interviewer's assigned schools at least two days before the visit to confirm that he or she was expected, and to make last-minute adjustments

where necessary. Once on site, the interviewer performed the tasks as outlined in The Interviewers' Check List (see Appendix A).

Class observations and teacher interviews were fitted into the school schedule as best they could be. The work schedule was so tight that some target classes could not be observed during the school visits.

The students were withdrawn from classes and met the interviewer in a quiet location. First, the interviewer gave a brief explanation of the study and assured the student that the interview results would have no bearing on his or her school marks. No mention of Monitors was made; indeed, the interviewers were not even told which of the schools they were visiting had had Monitors. Next, the interviewer asked the questions on the English interview schedule. It was given first to accustom the student to the interview format and to give time to build rapport. After all the questions on the English interview had been answered, the interviewer told the student (in English) that he or she would be asked a series of questions in French. The student was instructed to answer each question in a complete sentence if possible, and to give the best possible response to each question.

Once the interviewers had completed all their assigned tasks in one school, they moved on the next.

5. MEASURING THE MATCH BETWEEN SCHOOLS--COMPARING STUDENT CHARACTERISTICS

When the study was designed each Monitor school was matched with a Control school, using characteristics of schools, core-French programs, and teachers. The closeness of the match that was achieved in these three respects has already been described.

The purpose of matching was to obtain two very similar samples of students, one that had interacted with a Monitor and another that had not. The objective was that the samples should also be comparable with respect to characteristics such as age, sex, grade in school, intelligence, vocabulary, reading ability, and the grade level at which the student had started to study French. Other characteristics on which comparability was sought were whether a student had studied French outside school, whether someone in the student's home spoke a language other than English (and if so, what that language was) and whether the student spoke a language other than English while at home (and if so, what that language was).

An attempt was made to collect information on all these characteristics. School records were searched for scores on tests of intelligence, vocabulary, and reading ability and students who participated in the study were administered a questionnaire (see Appendix A).

The association between school type (Monitor vs Control) and each item of the questionnaire was examined separately for the samples of students drawn from the pairs of matched classes in each pair of matched schools; there were 16 such pairs of schools--32 pairs of classes--plus one Monitor school for which the matched Control school had withdrawn from the study at the last minute. No investigation could be made of intelligence,

vocabulary, and reading test scores because this information was available for relatively few of the students, and then on a variety of tests. For all other variables, except for responses to the question whether French was being studied outside school, information is presented in a series of tables in Appendix C. The tables for most variables consist of simple cross-tabulations. The tables for age give means and standard deviations of age distributions (age was measured in months). There are no tables for responses to the question about the study of French outside school because all students reported that their only study of French was in school--a necessarily perfect match.

The closeness of the match between the students chosen from the paired Monitor and Control classes varied to some extent with the variable used to measure it. For grade level the match was perfect. The matches for ability rating and sex were very good; the match was perfect for 18 of 32 class pairs with respect to ability rating and for 14 of 32 class pairs with respect to sex. These results are to be expected, given that paired classes were in the same grade and that students within each class were stratified on sex and ability rating before the sample of students to be interviewed was chosen. That there was a difference in the distribution of students by sex and ability rating for a number of class pairs--the difference consisted of only one student being out of place for half of the pairs that were different in one or the other or both of these variables--is a consequence of having too few students in a particular stratum in a school or of absenteeism on the days that the interviews were held (replacement students did not always match the sex and ability-rating specifications).

The other variables used to measure the closeness of the match between samples had not been used to stratify students, hence the match on these variables need not have been close. And yet for at least two variables it was. When compared by mean age, there was no consistent difference between paired class samples across the 32 pairs. Moreover, when compared by the grade in which the study of French was begun, all but eight pairs of class samples were found to match perfectly, and these eight differed by at most one student.

The largest number of discrepancies were found for two variables beyond the contractor's control. Over 100 students reported that someone

at home spoke a language other than English; the pattern of differences between the paired class samples on this variable was complex. Also complex was the pattern of differences for these students in whether it was French or another language that was spoken instead of English. The paired class samples also differed, but to a less important extent, on the student's speaking a language other than English while at home. Only 32 said that they did, and differences among paired class samples were scattered on this variable. In Saskatchewan, Prince Edward Island, and Nova Scotia, the samples were satisfactorily balanced (see, e.g. Appendix Table C).

It is difficult to judge how much effect these differences between paired class samples might have had on the IEA test results and the scores derived from the French and English interviews. They can be expected to have added to the error variances used to judge the significance of the effects of interest in the analyses of variance reported later. The consequence of this should be to make the significance tests less sensitive than they would have been had the match been closer. Nevertheless, matching was so close on all the variables used to measure it that when the data were aggregated over types of schools within provinces and then over provinces, no association was found between type of school attended--Monitor vs Control--and any of these variables. This suggests that the overall effect of Monitors in the analyses of variance reported later was estimated without any biasing effects caused by student samples.

6. DATA ANALYSIS--STUDENT VARIABLES

As finally realized, the design for the data collections can be described as in Table 4. Within each province there were pairs of schools, one of which had been served by a Monitor and the other (a comparable school) not. In Prince Edward Island there were six pairs, in Saskatchewan four, and in Ontario only one. One comparison school in Manitoba had to withdraw at the last minute, leaving a gap there.

Within each school there were two classes, the grade level being noted within each cell of Table 4. In every instance, classes with Monitors were compared with classes without at the same grade level, and the classroom was the smallest unit served independently by a Monitor. The statistical analysis, therefore, was structured to reflect the hierarchical nature of the design--students within classes within paired schools and schools within provinces.

The analysis model and implications for significance tests are displayed in Table 5, and subsequent analyses of variance follow this pattern. The utility, if not necessity, of such a complex model was borne out by the analysis, which revealed that schools and classes were significant sources of variation. If not separately estimated and used properly in error terms, the resulting analysis would be suspect.

Only Saskatchewan, Prince Edward Island, and Nova Scotia were included in the analyses of variance for two main reasons. First, the others had very few schools and thus contributed very few relevant comparisons. The necessarily full analysis could not be carried out if they were included. Second, the Ontario classes scored considerably higher on the listening

Table 4

Schematic Description of Provinces, Schools within Provinces, and Grades within Schools Involved in the Evaluation

School Pairs	Class Pairs	Saskatchewan		Manitoba		Prince Edward Island		Ontario		Nova Scotia	
		Monitor	Control	Monitor	Control	Monitor	Control	Monitor	Control	Monitor	Control
1	1	7	7	5	5	5	5	7	7	5	5
	2	8	8	5	5	5	5	8	8	5	5
2	1	7	7	5		5	5			5	5
	2	8	8	5		6	6			5	5
3	1	7	7			7	7			7	7
	2	8	8			8	8			8	8
4	1	7	7			5	5			6	6
	2	8	8			5	5			6	6
5	1					5	5				
	2					6	6				
6	1					5	5				
	2					5	5				

Note: Numbers in cells are grade levels of classes.
Six levels per class, except two pairs have three students per class.

Table 5

Analysis of Variance Model and Derivation of Error Terms for Use in Analysis of Variables from the Structured Interviews in Three Provinces[1]

Symbol	Source and Type	df	Expected Mean Squares	Error Term
Province Effects				
P	Province (Fixed)	2	$\sigma^2_W + 16\sigma^2_P + 4\sigma^2_{S:P} + 2\sigma^2_{C:SP}$	S:P
S:P	School/Prov. (Random)	11	$\sigma^2_W + 4\sigma^2_{S:P} + 2\sigma^2_{C:SP}$	C:SP
C:SP	Class/Sch. + Prov. (Random)	14	$\sigma^2_W + 2\sigma^2_{C:SP}$	W
Monitor effects				
M	Monitor vs. Control (fixed)	1	$\sigma^2_W + 24\sigma^2_M + 2\sigma^2_{MS:P} + \sigma^2_{MC:SP}$	MS:P
MP	Mon. x Province	2	$\sigma^2_W + 8\sigma^2_{MP} + 2\sigma^2_{MS:P} + \sigma^2_{MC:SP}$	MS:P
MS:P	Mon. x Sch./Prov.	11	$\sigma^2_W + 2\sigma^2_{MS:P} + \sigma^2_{MC:SP}$	MC:SP
MC:SP	Mon. x Class/Sch. + Prov.	14	$\sigma^2_W + \sigma^2_{MC:SP}$	W
W	Within Class	268	σ^2_W	-

Note: See Table 4 for description of nested design.

[1]See, e.g. Winer (1962, pp. 191-195). The provinces were Saskatchewan, Prince Edward Island, and Nova Scotia.

comprehension test (see Figure 1), supporting reports that Ontario students were in an "extended" program. Data from all provinces are covered in the results and discussion sections; only the analysis of variance is based on a reduced sample.

Finally, multivariate techniques were required to provide valid statistical tests when there were numerous, correlated dependent measures. As noted in a previous section, five ratings of French competency were derived by the Interviewer from each interview. As would be expected, these are highly correlated, and under such circumstances a multivariate test is the most powerful way to use all the data. The "Multivariance" computer program (Finn, 1977) could accommodate the double nesting, unequal Ns, and different error terms in a multivariate analysis. The variance analyses and statistical tests for Monitor effects are supplemented by exploratory and descriptive analyses.

The four scores derived from the 23 items of the French interview (failures to respond, correct responses, number of prompts, and sum of quality ratings) were subjected to the same analysis as the variables in Tables 5 to 9, with a null result. Distributions of school means are displayed in Figure 4. Though each difference in median school average favors the Monitor group, these differences are not significant individually, or by the multivariate test. Neither were there differences between Monitor and Control groups on responses to item 18.

The three elicited imitation tasks yielded two scores, total unit responses and total number correct. The distributions of school means in Figure 5 reveal almost identical median scores and school mean spreads that might easily arise by chance from the same population.

Data Analysis--Observation Variables

Although 66 classes were involved in the study as a whole, only 52 classes were observed. Among these there were 21 pairs of classes, each pair consisting of a class from a Monitor school and one at the corresponding grade level from the matched Control school. The 21 pairs of classes were drawn from 12 pairs of schools. Nine pairs of schools were each represented by two classes, three pairs by only one class. The analysis was restricted to the observations made of these 21 pairs of classes.

Most classes were observed only once. Six classes among 21 pairs were observed twice. The two observation records for these classes were averaged to obtain a single record.

The observation record for a class yielded the following variables:

1. Total number of different questions asked by the teacher in the expectation of a response
2. Number of questions answered by the class as a whole (as distinct from questions answered by individuals)
3. Number of correct responses
4. Number of incorrect responses
5. Number of questions to which no response was given
6. Number of responses in English (when French was expected)
7. Number of prompts (from the teacher) to elicit responses
8. Number of prompts to correct responses
9. Number of prompts in English
10. Number of times the teacher moved to another student for a response
11. Number of times the teacher used English
12. Average quality rating assigned responses in French.

The frequency of number of occurrences of variables 2 through 11 in the preceding list seemed inappropriate for further analysis. They depend to some extent on the magnitude of variable 1, the total number of questions asked by the teacher. To correct for this problem, variables 2 to 11 were converted to proportions by dividing them by variable 1. It was the proportions for these variables that were subjected to further analysis.

The statistical analysis of the observation data consisted of treating each pair of classes as one unit of observation. The differences between paired classes for each observation variable were obtained. A correlated t test was applied to the distributions of these differences to test the null hypothesis that the mean difference for each variable was zero in the population.

The correlated t test that was used in this analysis is possibly inappropriate. One reason for suggesting that is that it ignores the fact that the 21 pairs of classes are not all independent; there are two pairs of classes from each of nine pairs of schools. One way to circumvent this problem is to obtain averages over pairs of classes within pairs of schools, thus treating the school pair as the unit of analysis and reducing the number of units to 12. This was done and the analysis was repeated.

7. RESULTS

Student Variables

The measures included language proficiency or competency (from the French interviews and the IEA Listening Test) and the group concerned with attitudes, knowledge, and experience in French (from the English interview).

Table 6 contains results of analysis of the IEA test scores. The distributions of Monitor and Control school means are displayed in Figure 1, showing clearly the extreme (but comparable) values obtained in the two Ontario schools and the slight difference between Monitor and Control schools that our analysis tells us is well within the bounds of chance variation.

The summary of MANOVA results for the five competency ratings from the French interview is presented in Table 7. As noted above, only the omnibus multivariate test is presented because of the high positive correlations among the ratings (see Appendix D, Table D1). Analysis of the other five competency or proficiency scores is summarized in Table 8. Table 9 is a parallel table for the attitude, knowledge, and experience variables derived from the English interview. The model for the analysis in Tables 7, 8, and 9 was presented in Table 5.

Figure 2 displays distributions of school means on the five competency ratings, showing that students in the Control schools were rated by the Interviewers as slightly more competent than students in the Monitor schools (per significant Monitor effect, Table 7). No other significant Monitor differences were found on competency or proficiency scores.

Table 6

Analysis of Variance of IEA Listening Comprehension Data from Saskatchewan, Prince Edward Island, and Nova Scotia

Sources	df	MS	F	P
Province Effects				
Province	2	735.82	10.74	.00
School/Prov.	11	68.53	1.21	.21
Class/Sch. + Prov.	14	56.71	4.25	.00
Monitor Effects				
Monitor vs. Control	1	1.05	.02	.89
Mon. x Prov.	2	9.26	.18	.83
Mon. x Sch./Prov.	11	49.92	2.62	.05*
Mon. x Cl./Sch. + Prov.	14	19.03	1.43	.14
Within Class	265	49.92	—	—

*Statistically significant

IEA Listening Comprehension Test

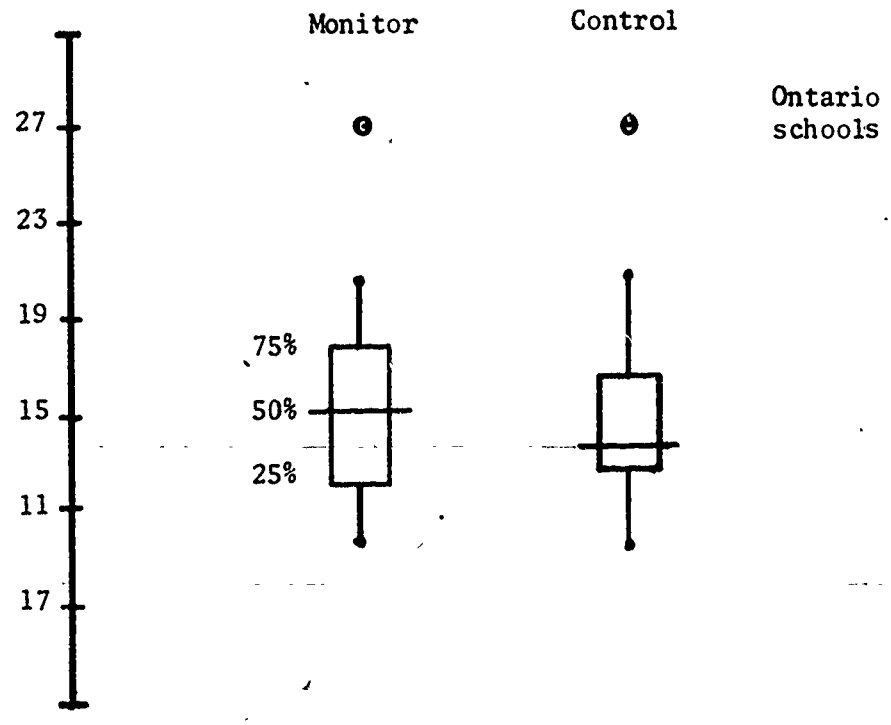

Fig. 1. Distributions of 33 school means (17 Monitor and 16 Control) from all five provinces: Horizontal line marks the median, box includes the middle 50% of school means, dots mark minimum and maximum observed means. Circled dots mark exceptionally extreme values (outliers) that have been set aside for calculation of the median and quartiles (box).

Table 7

Multivariate Analysis of Variance Results for Five Competency Ratings[1] from the French Interviews in Saskatchewan, Prince Edward Island, and Nova Scotia

Sources	df	F[2]	p
Province Effects			
Province	10;55	2.85	.04*
School/Prov.	55;70	2.71	.01**
Class/Sch. + Prov.	70;1261	1.05	.38
Monitor Effects			
Monitor vs. Control	5;7	5.37	.02*
Mon. x Prov.	10;14	1.53	.23
Mon. x Sch./Prov.	55;50	1.47	ns
Mon. x Cl./Sch. + Prov.	70;1261	0.95	.59
Within Classes	1261	-	-

[1]Pronounciation, Fluency, Grammar, Vocabulary, Comprehension

[2]Multivariate F, five variables, df vary with effect and error term.

*Statistically significant

**Highly statistically significant

Table 8

Multivariate Analysis of Variance Results for Five Variables Derived from the French Interview--Items 1-27[1]

Sources	df	F	P
Province Effects			
Province[3]			
School/Prov.[3]			
Class/Sch. + Prov.	70;1261	1.54	.00**
Monitor Effects			
Monitor vs. Control	5;7	0.98	.49
Mon. x Prov.	10;14	0.70	.71
Mon. x Sch./Prov.	55;50	1.85	.01**[2]
Mon. x Class/Sch. + Prov.	70;1261	1.71	.00**[2]
Within Classes	1261	—	—

[1]No. of failures to respond, no. correct, completeness of correct response, no. of prompts (except item 18), no. of units correct (Item 18)

[2]Traced primarily to one school in one province (see Appendix D, Table D4 for details)

[3]Province and school-within-province effects were not estimated in this analysis.

**Highly statistically significant

Table 9

Multivariate Analysis of Variance Results for Four Non-linguistic Variables[1] Derived from the English Interviews in Saskatchewan, Prince Edward Island, and Nova Scotia

Sources	df	F[2]	P
Province Effects			
Province[3]			
School/Prov.[3]			
Class/Sch. + Prov.	56;1032	1.77	.00**
Monitor Effects			
Monitor vs. Control	4;8	3.79	.05*
Monitor x Prov.	8;16	0.71	.68
Mon. x Sch./Prov.	44;44	4.07	.00**
Mon. x Class/Sch. + Prov.	56;1032	.98	.51
Within Classes	1032	—	—

[1]Attitude to French, Knowledge of French Culture, Quality of Experiences, Quantity of Experiences

[2]Multivariate F, four variables, df vary with effect and error term. Univariate tests can be made from results given in Appendix D.

[3]Province and school-within-province effects were not estimated in this analysis

*Statistically significant

**Highly statistically significant

Competency Ratings from the French Interview

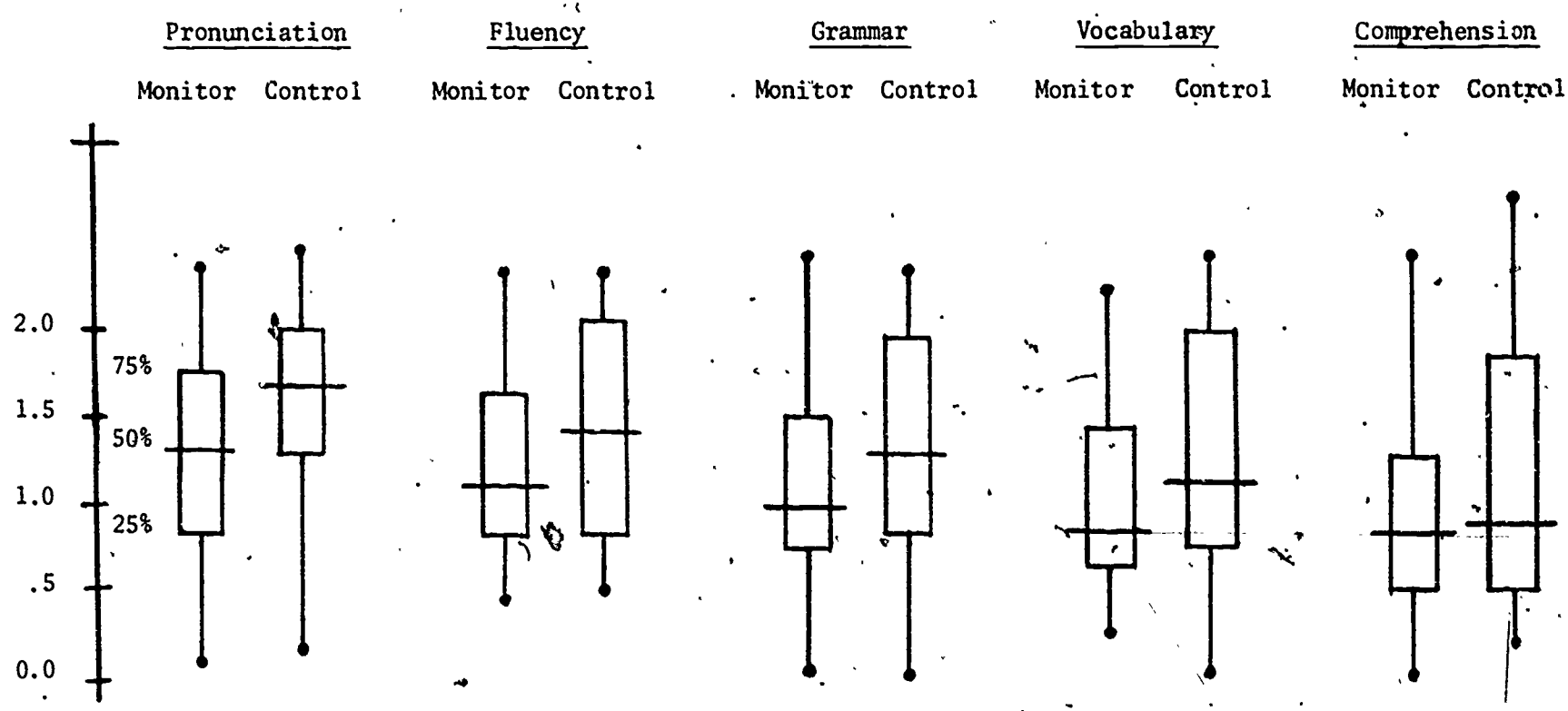

Fig. 2. Distributions of 33 school means (17 Monitor and 16 Control) from all five provinces: Horizontal line marks the median, box includes the middle 50% of school means, dots mark minimum and maximum observed means.

The significant Monitor effect on variables from the English interview (Table 9) can be seen in Figure 3. Students in Monitor schools showed better attitudes, more knowledge of French culture, and reported more and better experiences in French class.

Observation Study

The results of the Observation study revealed no important differences between Monitor and Control classes on any of the observation variables. (The distributions of differences for the 21 pairs of classes on the 12 observation variables are presented in tables in Appendix E.) The null hypothesis of zero mean difference was rejected for only one variable, the proportion of times during an observation session that students gave no response at all to a question ($t = 2.32$, $df = 20$, $p > .05$). The mean difference in proportions favored the Control classes by 0.02, which is to say that, when asked a hundred questions, Monitor classes like those in this study would be expected not to respond to two more questions than the corresponding, paired Control classes. This expected difference must be evaluated against the average proportions of times that students in both types of classes failed to respond to questions. These proportions were small--0.04 and 0.02 for Monitor and Control classes, respectively--and seem to reflect nothing of educational significance.

The analysis of the observation variables was repeated using paired schools as the unit of analysis. The results were substantially the same as for the analysis of data from paired classrooms.

It must be concluded that no evidence was found in the observation variables of important differences between Monitor and Control classes. This conclusion is based on the 21 class pairs (five provinces) included in this analysis, and generalizes to these school pairs.

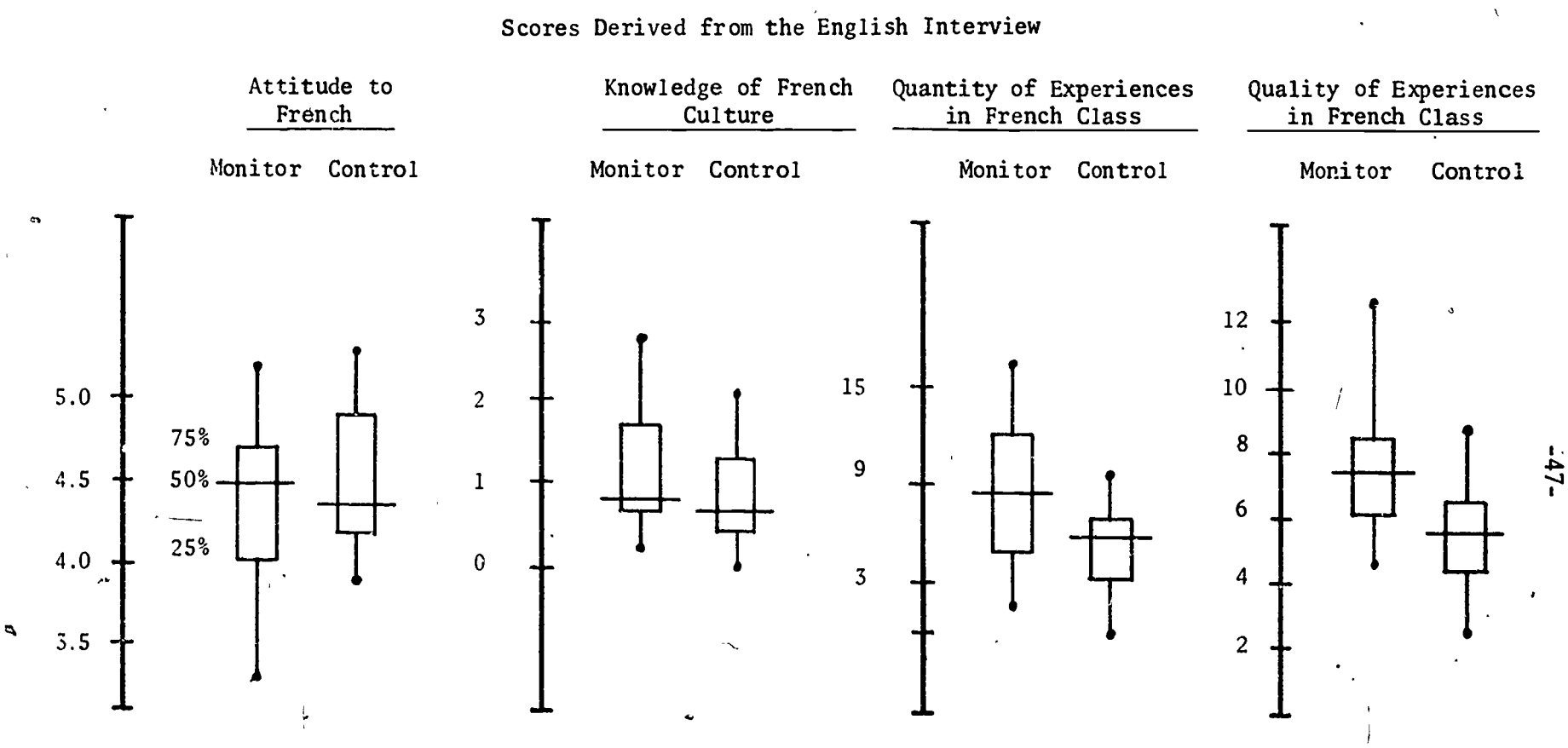

Fig. 3. Distributions of 33 school means (17 Monitor and 16 Control) from all five provinces: Horizontal line marks the median, box includes the middle 50% of school means, dots mark minimum and maximum observed means.

8. DISCUSSION

To the contractor, the most striking pattern in all the outcomes was the similarity between Monitor and Control groups. The matching of schools and balancing of samples had produced two groups that, in the contractor's judgement, were more alike overall than two random samples would have been.[1] Some differences were found, as noted in the previous section and discussed below, but the strong result is similarity.

Given that half of the schools had been served by a Monitor for much of the year (though we don't really know how much Monitor time they got), this similarity also implies that the aggregated impact of the Monitors, the effects averaged over 12 students per school, could not have been large. Arguments were advanced in the first section of this report why expectations for Monitor effects should be modest, and such expectations would seem to have been justified.

Competency and Proficiency

Consider first the scores on the IEA Listening Comprehension Test. The range and size of scores indicated that the test was suitable for all schools, with the possible exception of the two Ontario schools, where the test was too easy. Contrary to reports during our survey, the French program in these

[1] This is a subjective judgement based on many years of experience with data from comparative experiments. Researchers preoccupied with differences (including the present authors) rarely report on similarities, and time did not permit a thorough exploration of this conjecture.

schools could not be described as a core program (as that term is used in this study), in that French was studied for 40 minutes per day--about twice the time spent by students in other schools.

Monitor schools overall had a higher mean (and median) score; but variation among classes and schools (excluding Ontario and Manitoba) was large enough that the difference could easily be due to chance (see Table 6 and Figure 1). If we use this international standard, the data support a conclusion of no difference.

The IEA test was perhaps the sternest test of the students' ability to understand spoken French. The voices (in standard French) emanated from tape recorders of varying quality, with none of the visual or contextual clues that a listener has when speaking face to face. The listening task resembled very little the encounters between Monitors and students described to the contractor by teachers and coordinators. The IEA test was seen by the contractor as a benchmark, not very likely to be affected by Monitors but useful in view of the untried nature of the other measures. Had the survey and sample selection not taken so long, the IEA test would have been given early in the year and used (as a covariate) to reduce error variance and, if necessary, to equate the groups by statistical means.

The Interviewers, on the other hand, rated the Control groups higher than the Monitor groups on the five characteristics, pronunciation, fluency, grammar, vocabulary, and comprehension, and the difference cannot be readily attributed to chance (see Table 7, Figure 2, and Appendix D, Table D2). These ratings should be regarded as one global rating, as noted previously.

Before we discuss the global Interviewer ratings further, consider the last set of competency scores, those derived by summing rated responses to 23 of the items of the French interview (responses to short questions). The results for failure to respond, number of prompts, number correct, and quality/completeness (see Figure 4) were all in favor of the Monitor groups, though not to a statistically significant extent. These scores are also highly correlated and can be viewed as a different sort of measure of another dimension of competence, especially (recalling the provincial coordinator's words) whether Monitors succeeded in "loosening the tongue." The distributions of scores derived from the three elicited imitation tasks (Figure 5) were very similar in the two groups.

Scores Derived from 23 Questions of the French Interview

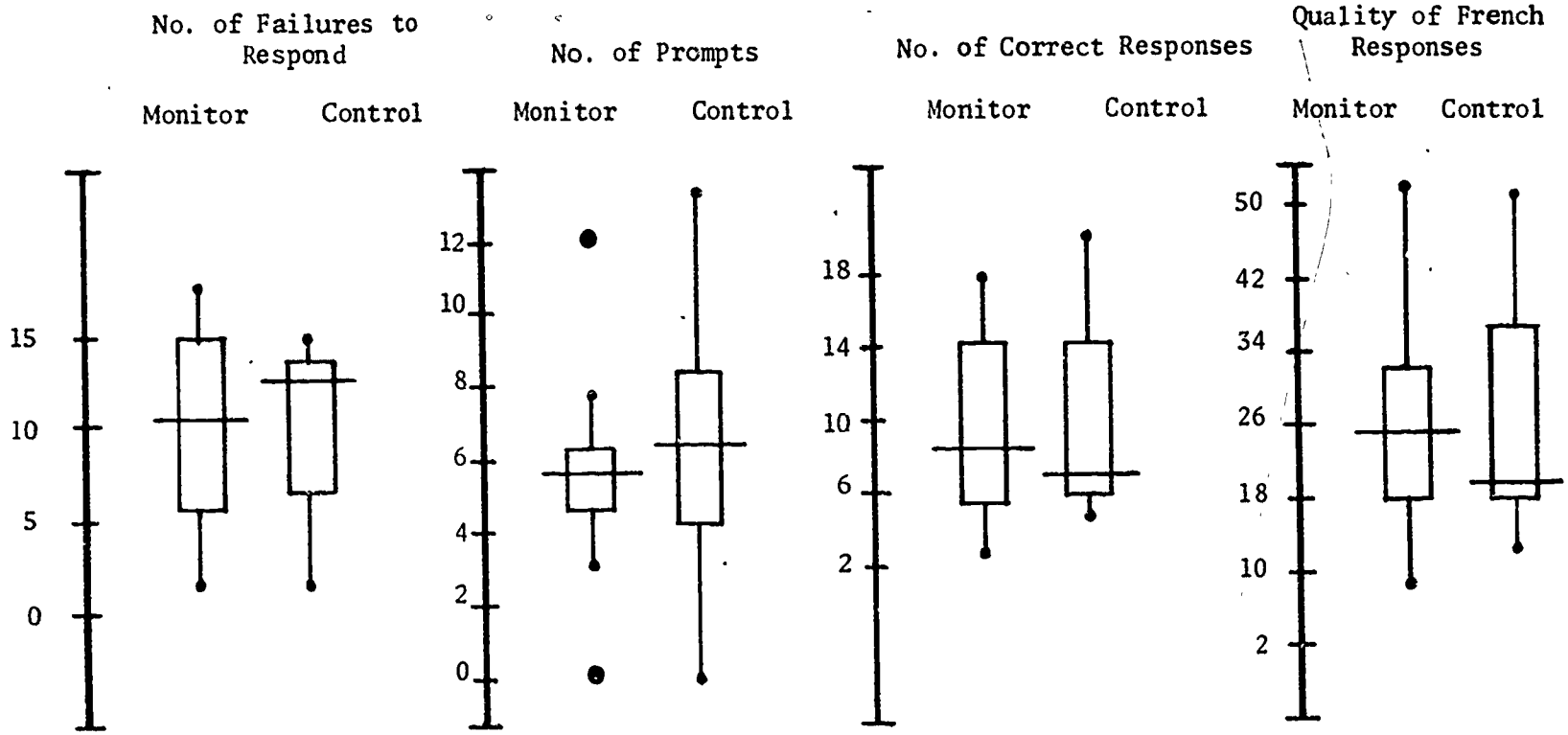

Fig. 4. Distributions of 33 school means (17 Monitor and 16 Control) from all five provinces: Horizontal line marks the median, box includes the middle 50% of school means, dots mark minimum and maximum observed means. Circled dots mark exceptionally extreme values (outliers) that have been set aside for calculation of the median and quartiles (box).

Scores Derived from the Elicited Imitation Tasks
Items 28-30 of the French Interview

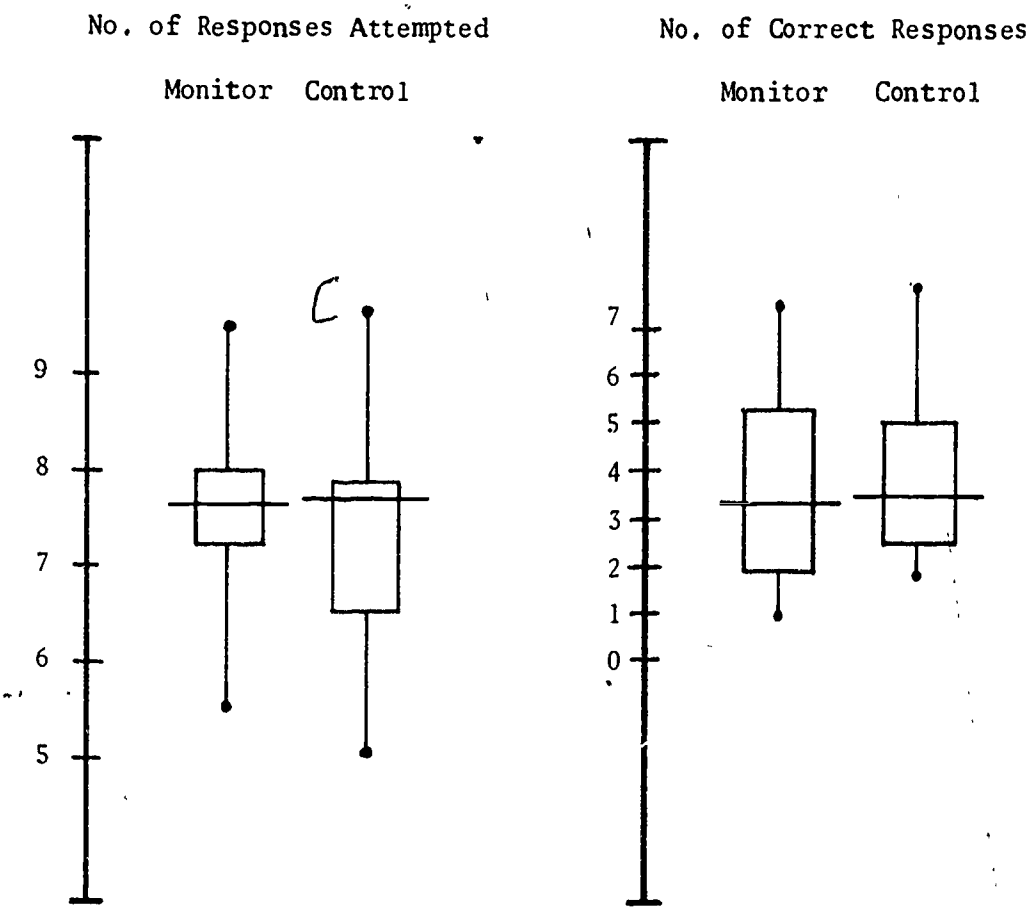

Fig. 5. Distributions of 33 school means (17 Monitor and 16 Control) from all five provinces: Horizontal line marks the median, box includes the middle 50% of school means, dots mark minimum and maximum observed means.

That these competency scores do not confirm (may even contradict) the global ratings, complicates the interpretation of the data. Recall that these competency scores showed high internal consistency reliability, indicating that together they measure some characteristics in a stable fashion. Before we probe the inconsistency of competence results, a few comments on the qualitative responses are in order.

Attitude, Culture, Experience

Monitor groups showed more positive attitudes and more knowledge of items of French culture, reported more experiences in French, and achieved a higher quality of experience score than the Control groups (see description of English interview for scores and Figure 3 for results). The differences are significant (Table 9), traceable mainly to the quantity and quality of experiences in French class.

The size and consistency of these results on variables that are aggregates of large numbers of responses are strong support for attributing the differences to the work of the Monitors. Students in these schools do not have opportunities outside school to meet French speakers and to learn systematically about French culture. In general, inside school and out, *apart from the Monitors*, the students in both groups had the same (minimal) opportunities.

Exploring the Inconsistent Results on Competence

The Interviewers rated the Control groups higher than the Monitor groups by their judgements made at the end of the French interview. When analysed apart from other measures, the differences were significant. These global ratings did not agree with the objective counts derived by summing ratings on individual items of the French interview, including the summation of the Interviewers' judgements of quality for each item of the interview (Figure 4, Quality of French Responses). In other words, the Interviewers' own judgements, item by item, added up in favor of the Monitor groups, whereas their summary judgements favored the Controls.

One question posed was,"Do Monitors 'sensitize the ear'?" The comprehension rating and the scores, number of Failures to Respond, and Number of Prompts were designed to help answer this question. Fewer prompts were required by students from Monitor classes and there were fewer failures to respond. The skewed and stretched out distributions, especially on prompts,

indicate that one cannot rely entirely on traditional statistical tests when faced with puzzling results.

Examining the comprehension ratings presented in detail in Table 10, one is struck with the low level of performance reflected by this and all ratings. Students in the first few years of core French develop very little competence, however defined. The observed rating difference in favor of the Control group apparently comes from both a few more low ratings in Monitor groups and one or two more middle ratings in Controls.

One hypothesis always present when people make subjective judgements (such as the ratings) is that there is unconscious bias in favor of one or the other groups under study. The training of Interviewers in this study has been described and it will be recalled that they did not know in advance which schools did and did not have Monitors. After reading Interviewer comments and hearing their verbal reports, we were quite confident that results were not affected by any tendency of Interviewers to favor one group over the other. The inconsistencies forced us to reexamine this possibility.

In two of the three provinces on which the main analyses are based, the Interviewers were experienced teachers. One had been a French coordinator. The Interviewer in the other province was a younger person, a very recent graduate of teacher training. The differences between Monitor and Control groups recorded by the younger Interviewer are all smaller than those recorded in the other two provinces. There is other evidence of difference among the Interviewers that may throw light on the failure of the two types of quality judgements to agree.

Figure 6 contains a graph of global rating results, comparing Monitor and Control groups means for the three provinces, combining all five ratings. Before they were averaged over the five ratings, data on each one were converted to standard scores, province by province, subtracting from each rating the provincial mean and dividing by the provincial standard deviation. The standard scores in each province thus had a grand mean (all scores) of zero and a standard deviation of one. Differences between Monitor and Control groups, if any, were preserved, but any tendency of an Interviewer in one province to rate consistently higher or lower ("mark" easier or harder) than Interviewers in other provinces was eliminated.

Table 10

Percentages of Students in Each Province Receiving Various Ratings on the Basis of the French Interview*

Scale Name	Categories	Provinces and Monitor vs. Control						Across Three Provinces	
		Saskatchewan		Prince Edward Island		Nova Scotia			
		Monitor	Control	Monitor	Control	Monitor	Control	Monitor	Control
Vocabulary	Little or none	21	13	12	7	8	16	14	11
	Simple basic	23	33	17	7	31	14	23	16
	Something more	6	4	21	36	10	21	13	23
Grammar	No command of rules--miscommunicates	14	20	8	2	12	14	11	10
	Fair	32	21	16	8	30	18	25	16
	Good to excellent	5	8	26	40	7	18	14	24
Comprehension	No understanding	35	34	16	7	12	21	20	19
	Some of slow speech	8	12	12	9	32	15	17	12
	Almost all if slow	6	4	10	20	4	8	8	12
	Understands normally	1	0	11	13	2	6	5	7
Fluency	Fragmented, primitive	20	21	9	2	12	15	13	11
	Very hesitant	17	18	13	10	24	18	17	15
	Some speech flow	13	10	20	26	13	12	16	17
	Approx. normal speech	1	0	8	12	2	6	4	7
Pronunciation	Hardly intelligible	39	34	20	8	29	26	28	21
	Usually intelligible	12	15	30	42	21	24	22	29
No. of Students (Total, 324)		48	48	66	66	48	48	162	162

*For illustration of results by category. Statistical analysis of these data is reported in Table 7 and in Appendix D, Table 2.

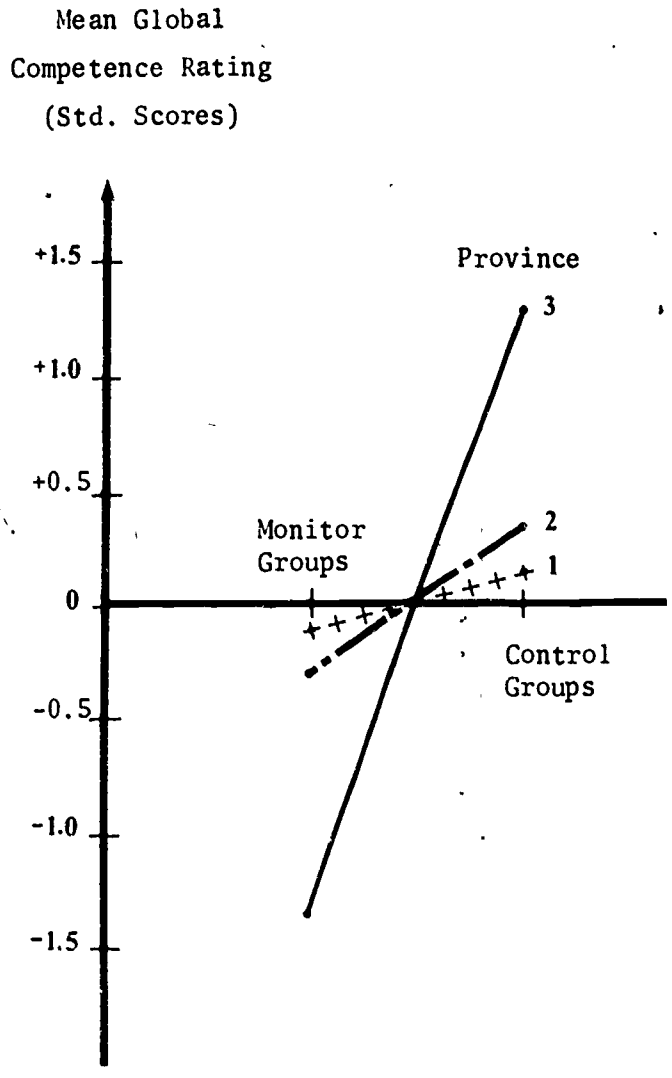

Figure 6. Average global competence rating (in scores standardized by province) for Monitor and Control groups in three provinces.

Figure 6 suggests a Monitor by Province interaction, and statistical analysis of these standard scores provides some support ($p = .08$).

One explanation for this interaction is that the different Interviewers actually applied different criteria in reaching their judgements, in spite of the efforts during training to standardize via definitions of fluency, comprehension, and so on. The higher ratings given to Control group students, especially by the experienced Interviewer in Province 3, may indeed document a monitor effect to which that Interviewer was highly sensitive--students who work with Monitors may become less like the usual core-French student. Their pronunciation may be non-standard, and therefore less attractive to Interviewers accustomed to "school French."

Research completed after the Monitor evaluation was begun suggested that age and language background can affect the judgements listeners make of French spoken by Anglophone students. Lepicq (1980) studied the acceptability of French spoken by 12-year-old immersion students to unilingual and fully bilingual "judges," half of whom were 12-year-olds and half of whom were adults. Both age and degree of bilingualism affected judgements in many ways too complex to go into here, but it was found that younger judges and bilingual judges were more tolerant. Since the context was different from the Monitor evaluation, such a study mainly sensitizes us to the likelihood of unwanted Interviewer effects. Had these results been available more effort would have been devoted to matching ages and backgrounds of Interviewers.

It is also possible that Monitors "sensitize the ear," but that they do so selectively with these beginning students and that there wasn't time for the sensitivity to mature and generalize to other speakers. On measures of willingness to respond and completeness of correct phrases, Monitor students were at least equal to the Controls.

If Interviewers in different provinces were using different criteria, there might be different patterns of correlation among the variables. Table 11 contains correlations among the five ratings, eight quality judgements derived from the items on the French interview and the IEA Listening Comprehension scores, calculated separately for each of the three provinces.

The French quality variables correlate moderately highly with ratings *and* IEA only in Province 1 (where the younger Interviewer worked). In the

Table 11

Within-Province Correlations Among Competence Measures
Values Less than 0.20 and Decimal Points Have Been Omitted

Province 1

	Item Level									Ratings			
	No. Cr. 23 Items	Qual. 23 Items	No. Cr. 3 Sits.	Qual. 3 Sits	No. Rep. E.I.	No. Cr. E.I.	No. Cr. Item 18	Qual. Item 18	IEA List. Comp.	Pron.	Fluency	Gram.	Vocab.
Qual. 23	83												
No. Cr. 3	31												
Qual. 3	46	55	23										
Rep. E.I.	25	32		29									
No. Cr. E.I.	55	48	29	37	42								
No. Cr. 18	29	39		22									
Qual. 18	39	37	21	22			58						
IEA	57	52	22	23		36		30	100				
Pron.	48	43	28	30		53			39				
Fluency	54	53	22	37		41	36	33		57			
Gram.	52	43	43	26		36		21		56	53		
Vocab.	63	58	37	53		39		21		50	59	52	
Comp.	68	64		32		37	34	29	53	60	57	50	55

Table 11 (Cont'd.)

Province 2

	Item Level								Ratings			
	No. 23 Items	Qual. 23 Items	No. 3 Sits.	Qual. 3 Sits.	No. Rep. E.I.	No. Cr. E.I.	No. Item 18	Qual. Item 18	IEA List. Comp.	Pron.	Fluency	Gram.	Vocab.
Qual. 23	93												
No. Cr. 3	61	56											
Qual. 3	68	74	72										
Rep. E.I.	46	46	35	44									
No. Cr. E.I.	83	78	59	64	58								
No. Cr. 18	66	61	42	42	27	58							
Qual. 18	56	61	40	56	24	51	58						
IEA													
Pron.	21		24			21	24		46				
Fluency									41	76			
Gram.	24	20	26				22		50	78	83		
Vocab.	27	23	29	21		22	20		46	76	77	85	
Comp.	33	25	31	24		22	22		42	82	78	82	84

Table 11 (Cont'd.)

Province 3

	Item Level									Ratings			
	No. 23 Items	Qual. 23 Items	No. 3 Sits.	Qual. 3 Sits.	No. Rep. E.I.	No. Cr. E.I.	No. Item 18	Qual. Item 18	IEA List. Comp.	Pron.	Fluency	Gram.	Vocab.
Qual. 23	93												
No. Cr. 3	74	76											
Qual. 3	70	77	79										
Rep. E.I.	51	55	43	52									
No. Cr. E.I.	77	78	73	73	54								
No. Cr. 18	63	65	54	51	20	57							
Qual. 18	54	62	61	69	21	60	66						
IEA	-25												
Pron.	-32	-28	-26	-23					62				
Fluency				-20					62	78			
Gram.				-22					59	70	79		
Vocab.	-28	-29	-32	-32		-29	-27	-27	56	72	82	80	
Comp.				-20				-20	62	74	81	81	80

Note: Values less than 0.20 and decimal points have been omitted.

other two provinces, correlations of ratings with IEA are high, but correlations of ratings with French quality scores are low, even negative (all correlations in the table are significantly different from zero). Interviewers in Provinces 2 and 3 certainly appear to have meant something different when they marked each item than they did when they made global ratings at the end of the interview. Factor analysis of the three correlation matrices in Table 11 (see Appendix D, Table D3) revealed a sharp and clear separation between the French quality variables and the others in Provinces 2 and 3, but substantial overlap, and one more factor, in Province 1.

By now it is clear that there is no simple answer to the question, did students in contact with Monitors perform better than Controls in the French interview? In an attempt to bring all the results from the measures of French competence together, one final summary analysis was performed according to the model in Table 5.

Four variables were involved: the global ratings were aggregated into a single score. The sum of the quality/completeness ratings for all 30 items of the French interview was calculated (see pages 13-14: variables 3, 7, 10, and 11). Also, a total was obtained for the number of correct responses to all 30 items (variables 2, 6, 9, and 12). The IEA Listening Comprehension Test was also included.

The summary multivariate analysis is presented in Table 12, and the parallel univariate analyses in Appendix D, Table D4. Neither the multivariate test of all four variables nor the univariate test for the sum of the ratings indicated a significant Monitor effect or Monitor by Province interaction. The significant Monitor by Class interaction is evidence that the analysis was quite sensitive. It was possible to trace this interaction to the classes in one school in Province 2. In the Monitor school, one class was rated very well and one very poorly on both quality/completeness and Number of Correct Responses (see Table D4). The two classes did not differ significantly on either IEA or the sum of the five global ratings, and the Control classes in the matched school were both quite average.

Table 12

Multivariate Analysis of Variance Results for Three Aggregate Variables Derived from the French Interview[1] Plus IEA Listening Comprehension

Sources	df	F	p
Province Effects			
Province	8,16	5.66	.00**
School/Prov.	44,44	4.71	.00**
Class/Sch. + Prov.	56,1033	2.12	.00**
Monitor Effects			
Monitor vs. Control	4,8	2.12	.17
Mon. x Prov.	8,16	2.10	.10
Mon. x Sch./Prov.	44,44	1.49	.10
Mon. x Class/Sch. + Prov.	56,1033	1.66	.00**[2]
Within Classes	1261	-	-

[1] Quality/Completeness, Number Correct, and Sum of Five Ratings

[2] Traced to classes in one school in Province 2 (see Appendix D, Table D4 for class means).

** Highly statistically significant

9. CONCLUSIONS

In the attempt to sum up what has been learned, the first task is to describe the limitations on the generality of the results--to estimate how far one can generalize the findings beyond the sample of schools, classes, and students who were involved. Conclusions are then offered under the three main headings listed under the Purpose of the Evaluation--students' competence in understanding and speaking French, knowledge of French culture, and attitudes toward French and French culture. As noted above (p. 16), it was decided to concentrate on attitudes to learning French. Observations in classrooms detected no differences between Monitor and Control classrooms.

Limitations to the Generalizability of Results

By the design of the study, the conclusions are limited to core-French classes, here defined to be classes of English-speaking students receiving instruction in French about 20 minutes per day, i.e. not regularly more than 100 minutes per week. Again by design, results apply only to beginning students, those in their first one or two years of French study. These first two restrictions and a limited budget for travel meant that the sample for most analyses was confined to grades 5, 6, 7, and 8 in three provinces, Nova Scotia, Prince Edward Island, and Saskatchewan. The 28 schools from these provinces were atypical in at least one sense--they were all within 65 km of a university.

Substantial variability was observed on all measures *among* provinces and *within* all provinces among schools and classes (see Tables 6-9 and Table 12, 'Province Effects'). The absence of significant Monitor by Province effects suggested consistent results across provinces, within the exception

of the rating differences discussed at length in the section "Inconsistent Results on Competence." That effect, though very weak, seemed important to explore, and a case was made for Interviewer differences. In other words, generality was limited by choice of Interviewer, not by characteristics of the sample. Wherever Monitor by Class or Monitor by School effects suggested inconsistent results, hence less generality, the effect was traced to one or two schools in one province--not a large enough threat to generality to justify limitations on the conclusions.

To sum up, the conclusions reached in this study may be applied with reasonable confidence to all classes like those in our sample. If generalizations are made to different classes within the same three provinces, then these generalizations should be made cautiously.

Competence in Understanding and Speaking French

When all analyses are taken into account, no strong case can be made for a difference in listening or speaking competence between Monitor and Control groups. Controls showed higher global ratings than Monitors; the difference was statistically significant with the ratings considered in isolation, but not so when they were analysed jointly with the other competence measures. The latter analysis was seen to be very sensitive in at least one respect. Evidence was discovered that can be interpreted to mean that reasonable Interviewers disagreed on the meaning attached to the global ratings.

Monitors showed higher quality/completeness and Number of Correct Scores when individual item ratings from the same Interviewers were combined, but that difference did not reach significance. Scores derived from the elicited imitation tasks (items 28-30), situations (items 25-27), and reading numbers aloud (item 18) were positively correlated with each other and with scores derived from the simple questions (items 1-24, except 18). Therefore, they were reduced to two measures, a sum of the quality/completeness ratings in one and a total count of number correct in another.

In a final summary analysis (Table 12) with global ratings, the standardized test of listening comprehension, the overall quality/completeness ratings, and the total number correct, no Monitor-Control differences were significant. If there is a strong case, it is for no difference.

Knowledge of French Culture

Validity of and limitations to the direct measure of cultural knowledge have been noted (pages 15-16). The score used was a count of items of knowledge, and in Figure 3 one can see that the median of the school means for both Monitor and Control schools is less than one. In very few schools (all Monitor schools) was the average number of items of knowledge greater than two.

On the other hand, students in Monitor schools recorded significantly more experiences in French and these were judged to be of higher quality. Seventy-five percent of Monitor schools recorded quality means higher than all but 25% of the Control schools (Figure 3). Assigning a Monitor to a school is very likely to make a big difference in the number and quality of experiences students have in French. An increase or decrease in average knowledge of discrete facts is unlikely to be found.

Attitude to Learning French

Monitor schools had mean attitudes 4.5 or higher. The Control schools were not far behind, and the difference was not significant. Attitude scores were not significantly correlated with competence scores (hence correlations were not supplied). Again, as with competence scores, the data support a conclusion of no difference between Monitor and Control schools, with the comment that students in core French in these provinces (all five) seem to have had a positive attitude, whether or not they enjoyed the assistance of a Monitor. Other studies where influences might have been more powerful (e.g. Hanna et al.,1980) have found no increase in already positive attitudes to learning French.

Summing Up

Comparisons were made in five provinces between beginning core-French students who had contact with part-time French Monitors and students who did not. Measures of French speaking and listening competence, cultural knowledge, quantity and quality of experiences in French class, and attitude toward the learning of French were derived from interviews conducted in French and in English. Students who had Monitor contact reported many more experiences in French class, and their experiences were judged of higher quality than those of Control students. In other respects, however, no differences of note were found between Monitor and

Control students. By virtue of the fact that they enhance the quality and quantity of experiences in French class, part-time French Monitors deserve to be called cultural ambassadors.

REFERENCES

Buhler, S., & Buhler, R. P-STAT 78 user's manual. Princeton, N.J.: PSTAT, 1979.

Campbell, D. T., & Stanley, J. C. Experimental and quasi-experimental designs for research. Chapter 5 in N. L. Gage (Ed.), Handbook of research on teaching. Chicago: Rand McNally, 1963.

Carroll, J. B. The teaching of French as a foreign language in eight countries. New York: John Wiley, 1975.

Cronbach, L. Coefficient alpha and the internal structure of tests. Psychometrika, 1951, 16, 297-334.

Finn, J. D. Multivariance: Univariate and multivariate analysis of variance, covariance, regression and repeated measures--A Fortran program. (Version VI, Sept. 1977). Chicago: National Educational Resources, 1977.

Genesee, F. A comparison of early and late immersion programs. Montreal: McGill University, 1979 (Mimeo)

Hanna, G., Smith, A. H., McLean, L. D., & Stern, H. H. Contact and communication: An evaluation of bilingual student exchange programs. Toronto: OISE Press, 1980.

Lapkin, S., Swain, M., Kamin, J., and Hanna, H. Late immersion in perspective: The Peel study. Canadian Modern Language Review, 1982, in press.

Lepicq, D. Aspects théoriques et empiriques de l'acceptabilité linguistique: le cas du français des élèves des classes d'immersion. Unpublished doctoral dissertation, University of Toronto (OISE), 1980.

McLean, L. D. Separating the effect of more hours of instruction from general language development on achievement in French as a second language. Toronto: OISE, Department of MECA, 1980.

Morrison, F., Walsh, M., Pawley, C., & Bonyun, R. French proficiency of grade 8 primary and late-entry immersion students. Ottawa and Carleton Boards, Spring 1979. French Working Paper No. 131. Ottawa: Research Centre, Ottawa Board of Education, 1979.

Naiman, N., Frohlich, M., Stern, H. H., & Todesco, A. The good language learner. Toronto: OISE, 1978.

Stern, H. H. Of honeymoons, hangovers, and bootstraps in foreign language teaching: The Modern Language Association Twentyman Lecture. Modern Languages, 60(1), March, 1979, pp. 1-9.

Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill, 1962.

APPENDIX A

OISE MONITOR EVALUATION

1980

Interviewer's Checklist

The primary purpose of your visit to each school is to interview 12 students selected by the teacher and to observe each of two selected classes in the school at least once. In addition, you are to meet with the French teacher to verify certain details of the French program in the school and to check on the progress of the testing phase of the study.

You should try to contact the principal of each school the day before your visit starts to remind him you are visiting the school. Leave a message, if necessary. Each principal and French teacher has been advised of the dates of your visit. Arrive at the school *at least* a half hour before classes start in order to have time to introduce yourself to the principal and French teacher. Ascertain (a) the location you will be using for the interviews, (b) the time of the French classes you are to observe, and (c) whether the students have been advised of the times and location for their interviews.

Be sure you have the following *before* going to each school:

a. Copy of the interview questions (1)
b. Copies of Interview Coding Sheets (12 + 2 spare)
c. Copies of Class Observation Form (2 + 1 spare)
d. Cassette recorder (1)
e. Cassette tape (1 + 1 spare)
f. Class Lists

At some time during your visit to each school you should arrange a short appointment with the French teacher.

Ask (a) If the French Listening Comprehension test has been administered to the selected classes. If it has not, please remind them to do so before the end of May (or if it is after the end of May as soon as possible).

(b) Ask if the Student Background Questionnaires have been completed--if not ask them to do so ASAP.

(c) Ask (a) the number of days in the school cycle (if other than a week) and the number of minutes of French each selected class receives during each cycle; (b) the number of students in each of the selected classes.

(d) Ask, "How do you feel about teaching French in this particular school?"

(e) Then ask, "How do you feel about the students in your French classes this year?"

(f) Ask, "How were the 12 students chosen for interview?"

Before you begin in the next school, write up a brief description of your experiences. Include the information requested above in your report. In your description, focus on:

(a) Exceptions to the plan, if any

(b) Unique features of the school.

Before you leave a school, use the summary checklist below to be sure you have remembered everything:

Interview 12 students, recording two a.m., two p.m.

Observe in two classrooms and complete the forms.

Meet French teacher and pose questions.

- Listening tests completed?
- Student background questionnaire?
- Length of school cycle and French period?
- Size of French class?
- Feeling about teaching French in this school?
- How were students chosen for interview?

Write description of visit.

INSTRUCTIONS TO INTERVIEWERS

Each interview should take about 20 minutes. Ask the school to schedule the interviews 30 minutes apart to give you time to jot down any notes and to prepare for the next student.

Be sure to mark in the identifying information at the top of each coding sheet. We must be able to correlate the interview with the other tests.

The answers to the English part of the interview are mostly Yes, No or numeric answers. Simply mark the student's response on the coding sheet.

Under the comments section, note your perceptions of the student, in particular his ability and willingness to answer the questions and how verbal and communicative he was. Also note any other occurrences of importance.

Any question that asks, "How often?" is coded as follows:

0 = never
1 = once per month
2 = twice per month
3 = 3 times per month
4 = once per week
5 = more often than once a week

The French section of the interview is more complex to code. Each response is coded on 2 scales.

The first scale 0 1 2 X codes the response as follows:

0 = no response
1 = incorrect response
2 = correct response
X = student was prompted or question was rephrased.

The second scale codes the quality or completeness of the response as follows:

0 = no response
1 = response was one word
2 = response was a phrase
3 = response was simple sentence
4 = response was very complete

As an example the response to question 1

Bonjour - ça va

could be any of:

a) Bonjour
b) Très bien merci
c) Je vais très bien merci
d) Je vais très bien merci, et vous

All of the above responses would score a 2 on the first scale but on the second scale, answer a) would rate a one, b) would get 2, c) would get 3 and d) would rate a 4.

Questions 28, 29, 30 are scored as follows. On the first scale give one point for each "part" of the sentence the student attempted to repeat, for a maximum of 3. On the second scale give one point for each part of the sentence the student repeats absolutely correctly.

As an example if a student said

"La grand chien /mange /le petit poule"

the student would score a 3 on the first scale, but only a 1 on the second scale "la grand" and "le petite" are incorrect.

At the end of the interview rate the student on his Pronunciation, Fluency, Grammar, Vocabulary, and Comprehension as follows.

Pronunciation
0 = often unintelligible
1 = obviously foreign but rarely unintelligible
2 = foreign but always intelligible
3 = sometimes foreign, always intelligible
4 = Native-like

Fluency
0 = every utterance requires enormous obvious effort
1 = very hesitant, often long pauses before answering
2 = hesitant, must pause before answering
3 = rarely hesitates
4 = responds as well in English

Grammar
0 = no control of syntax, often conveys wrong information
1 = fair control of syntax, can convey meaning in simple phrases, although there are frequent errors
2 = good control of syntax, conveys meaning well, can formulate simple sentences, there are still errors
3 = very good control, formulates more complex sentences, relatively few errors
4 = very good control, complex sentences, almost no errors

Vocabulary
0 = almost no vocabulary at all
1 = adequate for basic courtesy requirements
2 = simple social and school needs
3 = adequate for simple general conversations
4 = very good vocabulary, can handle routine social and general conversations

Comprehension 0 = did not understand most of questions asked, required frequent repetition, slow speech
1 = understood most of what was said, required slow speech, some repetition
2 = understood almost all of what was said, could follow slow speech with almost no repetition
3 = understood most all of what was said at normal rate of speech. Very few repetitions
4 = could follow all of what was said, no repetition required

In the comments section note in particular any differences in the students willingness to respond as compared to his performance in English.

The Ontario Institute for Studies in Education
Department of Measurement, Evaluation and Computer Applications

French as a Second Language Monitor Study

Interview Questions--English--10 minutes

1. Do you enjoy learning French at school?
2. Can you tell me one thing that you like about your French class?
3. Can you tell me one thing you don't like about your French class?
4. On the days you have French at school, is there one special French class that you like better than the others? A French class that you like more than the others?
5. Why do you like this particular class more?
6. Has your French class learned any French songs this year?
7. How many French songs did your class learn?
8. Can you name (up to) 3 of them for me?
9. Can you name (up to 3) singers who sing songs in French?
10. Do you speak to your French teacher in French when you are not in French class?
11. At school do you ever have a chance to speak to someone other than your French teacher in French?
12. How often? 0. never, 1. 1/month, 2. 2/month, 3. 3/month, 4. every week, 5. more
13. Do you have a French Club in your school?
14. Did you learn anything about how French-speaking people celebrate Christmas and New Year in your French class this year?
15. Did you learn anything about other special holidays or festivals that French-speaking people have?
16. Can you name one?
17. Do you play any special games in your French class?
18. How often?
19. Do you ever get to see a movie in French or about French-speaking people at school?
20. Do you listen to French stories in your French class?
21. How often?
22. Do you ever act out a little play in French at school?
23. How often?
24. Do you ever see a slide show in French or about French-speaking people at school?

25. Do you ever have a chance to do something in French at school other than in French class (cooking, sports)?

26. How often?

27. Is there someone at school other than your French teacher that you can ask for help in learning French?

28. Do you ever have a chance to work alone or in a small group with your French teacher?

29. Do you ever get a chance to work alone or in a small group with a French-speaking person other than your French teacher?

30. How often?

31. Do you ever listen to a French record in your free time?

32. Do you ever watch French programs on T.V.?

33. Do you think you would like to learn to speak French well?

34. Do you think it is important to learn French?

35. Do you think you will study French in high school?

Interview Questions--French--10 minutes

1. Bonjour - ça va?
2. Comment t'appelles-tu?
3. Quel jour est-ce aujourd'hui?
4. Dit moi (name), quel temps fait-il?
5. Quelle est la date aujourd'hui?
6. Quelle heure est-il?
7. Qui est ton professeur de français?
8. Où est ton professeur de français maintenant?
9. Combien de crayons y-a-t-il sur la table? (3)
10. Où est le livre? (sur)
11. Où est le stylo? (dans)
12. Où est la fille? (derrière)
13. Où sont les crayons?
14. Combien de frères as-tu?
15. As-tu des soeurs?
16. Quel age as tu?
17. Combien font 5 et 7?
18. Lisez les numéros suivants 13, 15, 20, 21, 31, 50, 80, 81.
19. Que fais-tu le samedi matin?
20. Quelle est la date de Noel?
21. De quelle coleur est la balle? (blanche)
22. De quelle couleur est le papier? (blanc)
23. Quel temps fait-il pendant l'hiver?
24. Quel temps fait-il pendant l'été?
25. Listen to the following situation and then answer in French.
    - You are in a McDonald's in Quebec City; the person behind the counter speaks only French. You want to order some milk. What would you say?
26. You are visiting a school in Quebec where the kids speak only French. You want to find a washroom. What would you say?
27. You are on a camping trip in Quebec. You stop in a small town to buy some groceries. The clerk speaks only French. Ask him for eight apples.

You are going to hear a sentence read in French twice; after you have heard the sentence the second time I want you to repeat it.

28. Le grand chien / mange / la petite poule.

29. Les enfants / regardent / le cheval noir.

30. Hier / j'ai / rencontré / mon ami François.

CLASSROOM OBSERVATION

The purpose of the classroom observation is to observe the behaviour of the selected classes in their normal classroom setting. The teacher should be asked not to do anything special for your visit but rather to carry on a normal class.

The class is observed on the basis of teacher-student interactions. Each interaction is coded in one column on the code sheet.

Responses

Class - Indicates the interaction was between the teacher and the class. A choral response.

Correct - Indicates the student's response was judged correct or accepted as correct by the teacher.

Incorrect - Indicates the student's response was judged incorrect by the teacher.

No response - self-explanatory.

English - Indicates the student's response was in English.

Quality - is used to indicate the quality of the student's response. Use the same criteria as for the French interview.

Prompts

To Elicit - indicates the student was further prompted by the teacher in an attempt to get the student to respond.

To Correct - the teacher prompted the student to correct a previously wrong answer.

English - indicates the prompt was in English.

Goes to other - indicates the teacher gave up on this particular student and went on to another, usually with the same or a similar question.

Teacher English - indicates the use of English by the teacher for explanation purposes.

Under comments and observations: Note the physical layout of the class rows, circle informal seating on floor etc. A general outline of the lesson. The French "ambiance" of the classroom, e.g., posters, charts, etc., for use in French class.

THE ONTARIO INSTITUTE FOR STUDIES IN EDUCATION
Department of Measurement, Evaluation and Computer Applications

French as a Second Language Monitor Study

STUDENT BACKGROUND INFORMATION QUESTIONNAIRE

School Number _____ (1 - 4)

1. What is your Student Number _____ (5 - 6)

2. What is your date of birth ____ _____ _____ (7 - 12)
year month day

Please put a circle around the answer for the following questions:

3. What grade are you in now? 4 5 6 7 8 9 (13)

4. Are you a boy or a girl? 1. boy 2. girl (14)

5. Do you take French at school this year? 1. Yes 2. No (15)

6. In what grade did you start to study French at school?
1 2 3 4 5 6 7 8 (16)

7. Do you study French anywhere other than at school? 1. Yes 2. No (17)

8. If you answered YES to Question 7 where do you study French? _____ (18)

9. Does anyone at home speak a language other than English?
1. Yes 2. No (19)

10. If you answered YES to Question 9 what is the language they speak? _____ (20)

11. Do YOU speak any language other than English at home?
1. Yes 2. No (21)

12. If you answered YES to Question 11 what is the language you speak at home? _____ (22)

THANK YOU FOR YOUR HELP

APPENDIX B

Reliability Estimates and Related Statistics for the Variables Derived from the French and English Interviews

(N=384)

	n	Σs_i^2	s_x^2	α^a
Variables Pertaining to Compétence in French				
26-Item Subtest of French Interview (includes situations)				
1. Number of failures to respond	26	5.52	42.50	0.90
2. Number of correct responses	26	5.51	43.71	0.91
3. Rated quality of responses	26	34.31	312.94	0.93
4. Number of prompts from interviewer	26	4.60	21.20	0.8[illegible]
4-Item Subtest of French Interview				
5. Number of units pronounced correctly	4	5.52	15.97	0.67
Variables derived from the English Interview				
6. Attitude toward the learning of French	7	0.81	1.29	0.43
7. Knowledge of French culture	3	0.98	1.21	0.28
8. Quality of experiences in French class	17	3.24	7.33	0.59
9. Quantity of experiences in French class	7	12.65	25.11	0.57

[a]Coefficient alpha (Cronbach, 1951) is defined as follows:

$$\alpha = \frac{n}{n-1}\left[1 - \frac{\Sigma s_i^2}{s_x^2}\right]$$

APPENDIX C

Means and Standard Deviations of Age Distribution for Students in Monitor and Control Schools Separately by Province

(Note: Age is Measured in Months)

SCH TYPE	PROVINCE					ROW TOTALS
	SASK	MAN	P.E.I.	ONT	N.S.	
MONITOR	48	24	66	12	48	198
	161.542	131.625	138.288	161.583	143.813	145.869
	7.973	3.797	13.930	8.339	15.000	16.073
CONTROL	48	12	66	12	48	
	161.792	132.167	137.152	162.667	144.708	146.785
	8.705	5.750	13.026	10.899	14.347	16.219
TOTAL N	96	36	132	24	96	384
MEAN	161.667	131.806	137.720	162.125	144.260	146.313
S.D.	8.304	4.465	13.446	9.507	14.607	16.129

Cross-Tabulation of Students by Sex and by Type of School (Monitor vs. Control) Within Province

SEX

PROV	SCH TYPE	MALE	FEMALE	TOTALS
SASK	MONITOR	25	23	48
	CONTROL	24	24	48
MAN	MONITOR	12	12	24
	CONTROL	6	6	12
P.E.I.	MONITOR	31	35	66
	CONTROL	32	34	66
ONT	MONITOR	4	8	12
	CONTROL		12	12
N.S.	MONITOR	24	24	48
	CONTROL	20	28	48
TOTAL N		178	206	384

Cross-Tabulation of Students by Grade by Type of School
(Monitor vs. Control) Within Province

		GRADE				
PROV	SCH TYPE	5	6	7	8	ROW TOTALS
SASK	MONITOR			24	24	48
	CONTROL			24	24	48
MAN	MONITOR	24				24
	CONTROL	12				12
P.E.I.	MONITOR	45	9	6	6	66
	CONTROL	45	9	6	6	66
ONT	MONITOR			6	6	12
	CONTROL			6	6	12
N.S.	MONITOR	24	12	6	6	48
	CONTROL	24	12	6	6	48
TOTAL N		174	42	84	84	384

Cross-Tabulation of Students by Ability Rating by Type of School (Monitor vs. Control) Within Province

PROV	SCH TYPE	ABIL. RATING LOW	AVERAGE	HIGH	ROW TOTALS
SASK	MONITOR	15	17	16	48
	CONTROL	16	16	16	48
MAN	MONITOR	8	8	8	24
	CONTROL	4	5	3	12
P.E.I.	MONITOR	22	22	22	66
	CONTROL	22	22	22	66
ONT	MONITOR	4	4	4	12
	CONTROL	3	4	5	12
N.S.	MONITOR	15	17	16	48
	CONTROL	15	17	16	48
TOTAL N		124	132	128	384

Cross-Tabulation of Students by Grade in Which the Study of French Was Begun by Type of School (Monitor vs. Control) Within Province

STRT. GRD. FRNCH

PROV.	SCH TYPE	GRADE FOUR	GRADE FIVE	GRADE SIX	GRADE SEVEN	ROW TOTALS
SASK	MONITOR				46	48
	CONTROL	1			47	48
MAN	MONITOR	24				24
	CONTROL	12				12
P.E.I.	MONITOR	40	9	6	11	66
	CONTROL	38	11	5	12	66
ONT	MONITOR		12			12
	CONTROL		12			12
N.S.	MONITOR	25	12		11	48
	CONTROL	24	12		12	48
TOTAL N		164	68	13	139	384

Cross-Tabulation of Students by Whether the Student Does Not Speak English at Home by Type of School (Monitor vs. Control) Within Province

PROV	SCH TYPE	Is A Language Other Than English Spoken by Student at Home? YES	NO	ROW TOTALS
SASK	MONITOR	3	45	48
	CONTROL	4	44	48
MAN	MONITOR	4	20	24
	CONTROL		12	12
P.E.I.	MONITOR	2	64	66
	CONTROL	3	63	66
ONT	MONITOR	5	7	12
	CONTROL	2	10	12
N.S.	MONITOR	5	43	48
	CONTROL	4	44	48
TOTAL N		32	352	384

Cross-Tabulation of Students by Language the Student Speaks at Home (If Not English) by Type of School (Monitor vs. Control) Within Province

(Note: Three More Students Responded to This Question Than Said They Did Not Speak English at Home.)

PROV.	SCH TYPE	LANGUAGE STUDENTS SPEAK AT HOME: FRENCH	OTHER	ROW TOTALS
SASK	MONITOR		4	4
	CONTROL		4	4
MAN	MONITOR	2	2	4
	CONTROL		1	1
P.E.I.	MONITOR		3	3
	CONTROL	3		3
ONT	MONITOR		5	5
	CONTROL		2	2
N.S.	MONITOR		5	5
	CONTROL	2	2	4
TOTAL N		7	28	35

Cross-Tabulation of Students by Whether Anyone at Home Does Not Speak English, by Type of School (Monitor vs. Control) Within Province

DOES ANYONE AT HOME NOT SPEAK ENGLISH?

PROV	SCH TYPE	YES	NO	ROW TOTALS
SASK	MONITOR	23	25	48
	CONTROL	26	22	48
MAN	MONITOR	11	13	24
	CONTROL	2	10	12
P.E.I.	MONITOR	9	57	66
	CONTROL	11	55	66
ONT	MONITOR	7	5	12
	CONTROL	3	9	12
N.S.	MONITOR	14	34	48
	CONTROL	12	36	48
TOTAL N		118	266	384

Cross-Tabulation of Students by The Language Spoken by Someone at Home (Not English) by Type of School (Monitor vs. Control) Within Province

(Note: Two More Students Responded to This Question Than Reported That Someone at Home Speaks a Language Other Than English.)

LANGUAGE SPOKEN BY SOMEONE AT HOME

PROV	SCH TYPE	FRENCH	OTHER	ROW TOTALS
SASK	MONITOR	7	16	23
	CONTROL	3	23	26
MAN	MONITOR	4	7	11
	CONTROL		2	2
P.E.I.	MONITOR	5	6	11
	CONTROL	10	1	11
ONT	MONITOR		7	7
	CONTROL		3	3
N.S.	MONITOR	7	7	14
	CONTROL	7	5	12
TOTAL N		43	77	120

APPENDIX D

APPENDIX D

Table D1

Pooled within-cell Correlations (Class/School/Prov.) of Rating Variables and IEA Listening Comprehension

(N=324, df=296, $p < .01$)

	IEA	Pron.	Fluency	Grammar	Vocab
Pronunciation	0.46				
Fluency	.47	.70			
Grammar	.53	.68	.72		
Vocabulary	.49	.69	.74	.75	
Comprehension	.54	.72	.74	.73	.76

APPENDIX D

Table D2

Values of Univariate Mean Squares for Attitude Variables, Based on Data from Sask., P.E.I. and Nova Scotia. Arrows Indicate Error Term to be Used in F-ratios

	Prov.	S:P	C:SP	T	TxP	TS:P	TC:SP	Within Class
Attitude to French	21.91	3.75	2.17	3.78	0.50	1.08	0.53	1.15
Knowledge of French Culture	1.05	5.87	0.86	1.78	0.28	3.28	0.93	.79
Quality of Experiences	128.38	46.97	9.35	245.44	19.99	20.67	4.64	3.01
Quantity of Experiences	785.54	149.69	19.91	699.31	71.74	73.31	5.39	9.74

W

Rating Variables: Univariate MS

	Prov.	S:P	C:PS	T	TxP	TS:P	TC:SP	Within Class
Pronunciation	17.57	2.36	0.87	7.72	0.47	2.11	0.90	.60
Fluency	25.86	2.31	0.68	2.60	2.05	1.66	1.18	.66
Grammar	24.97	1.90	0.72	4.00	1.80	1.56	0.76	.53
Vocabulary	15.00	3.20	1.31	7.72	2.00	1.97	0.71	.55
Comprehension	37.21	5.72	1.33	4.00	2.77	3.19	0.84	.65

W

-97-

APPENDIX D

Table D3

Varimax Rotated Factor Loadings for Significant Factors[1] Extracted from the Correlation Matrices in Table 11. Factor Loadings Less than 0.40 and Decimal Points Have Been Omitted. Variables Are Listed in Order of Decreasing Loading on Factor 1

Province 1		Factors			Province 2		Factors		Province 3		Factors	
	Variable Label	1	2	3		Variable Label	1	2		Variable Label	1	2
1	Grammar	74			1	Fr. Cr. 23 Items	92		1	Fr. Qual. 23 Items	93	
2	Vocabulary	69			2	Fr. Qual. 23 Items	92		2	Fr. Cr. 23 Items	88	
3	Pronunciation	68			3	Fr. Pronun No. Cr.	86		3	Fr. Pronun. No. Cr.	86	
4	Fr. Cr. 23 Items	68			4	Fr. Qual. 3 Sits	79		4	Fr. Qual. 3 Sits	84	
5	Comprehension	63			5	Fr. Cr. 3 Sits	67		5	Fr. Cr. 3 Sits	83	
6	Fluency	59			6	Fr. Q18 Pronun.	67		6	Fr. Q18 Pronun	70	
7	IEA List. Comp	53			7	Fr. Q18 No. Cr.	65		7	Fr. Q18 No. Cr.	67	
8	Fr. Qual. 23 Items	53	45	50	8	Fr. Pron. No Rep.	51		8	Fr. Pronun. No. Rep.	52	
9	Fr. Pronun. No. Cr.	51		51	9	Grammar		91	9	Fluency		90
10	Fr. Cr. 3 Sits	51			10	Comprehension		90	10	Grammar		90
11	Fr. Q18 No. Cr.		88		11	Vocabulary		89	11	Comprehension		89
12	Fr. Q18 Pronun.		60		12	Pronunciation		86	12	Vocabulary		86
13	Fr. Pronun. No. Rep			80	13	Fluency		86	13	Pronunciation		82
14	Fr. Qual. 3 Sits			40	14	IEA List. Comp.		51	14	IEA List. Comp.		70

[1]Iterative factor analysis procedure in PSTAT: FACTOR, ROTATE, GROUPCOR, BPRINT/0.40 (Buhler & Buhler, 1979, pp. 248-252).

APPENDIX D

Table D4

Values of Univariate Mean Squares for Table 12, Based on Data from Sask., P.E.I., and Nova Scotia. Arrows Indicate Error Term to be Used in F-ratios. Numbers in Parentheses are Degrees of Freedom

	Prov.(2)	S:P(11)	C:SP(14)	M(1)	MxP(2)	MS:P(11)	MC:SP(14)	Within Class (268)
IEA Listening Comprehension	563.86	63.50	50.85	1.36	8.97	46.35	21.06	17.46
RIGHTS	472.50	1175.27	119.10	1.00	158.94	123.75	55.03*	35.92
COMPLETES	821.44	5752.90	649.21	210.25	6.48	430.51	370.96*	173.23
Sum of Five Ratings	573.98	63.51	21.23	124.69	38.93	46.35	18.87	11.14
	↑	↑	↑ W	↑	↑	↑	↑	

*Mean scores for classes in two schools in Province 2. The significant Monitor by Class interaction was traced to these schools. One Monitor class in School A had particularly low scores and the other Monitor class had particularly high scores.

RIGHTS

	School A (Monitor)	School B (Control)
Class 1	17.5	22.7
Class 2	34.8	21.5

COMPLETES

	School A (Monitor)	School B (Control)
Class 1	40.3	58.3
Class 2	84.7	51.7

APPENDIX E

FREQUENCY DISTRIBUTIONS OF DIFFERENCES BETWEEN PAIRED MONITOR AND CONTROL CLASSES ON THE OBSERVATION VARIABLES

APPENDIX E

Frequency Distributions of Differences between Paired Monitor and Control Classes on the Observation Variables

TABLE 1

TOTAL NUMBER OF INTERACTIONS

LOW	HIGH	N	PCT	CUM
65.00	65.00	1	5	5
43.00	50.00	3	14	19
26.00	37.00	7	33	52
21.50	25.00	3	14	67
9.00	10.00	2	10	76
-12.50	-8.00	3	14	90
0.00	0.00			
0.00	0.00			
0.00	0.00			
0.00	0.00			
-75.00	-75.00	1	5	95
-91.00	-91.00	1	5	100
GOOD N				21.
MEAN				14.8810
S.D.				38.1454

TABLE 2

PROPORTION OF TEACHER-CLASS INTERACTION

LOW	HIGH	N	PCT	CUM
0.65	0.65	1	5	5
0.46	0.46	1	5	10
0.40	0.40	1	5	14
0.26	0.26	2	10	24
0.16	0.16	1	5	29
0.09	0.10	2	10	38
-0.00	0.04	4	19	57
-0.12	-0.05	4	19	76
-0.17	-0.16	2	10	86
-0.30	-0.30	1	5	90
0.00	0.00			
-0.52	-0.50	2	10	100
GOOD N				21.
MEAN				0.0176
S.D.				0.2901

TABLE 3

PROPORTION OF CORRECT RESPONSES

LOW	HIGH	N	PCT	CUM
0.25	0.25	1	5	5
0.00	0.00			
0.14	0.14	2	10	14
0.00	0.00			
0.08	0.08	1	5	19
0.01	0.02	3	14	33
-0.03	0.00	7	33	67
-0.06	-0.04	3	14	81
-0.10	-0.10	2	10	90
-0.15	-0.15	1	5	95
0.00	0.00			
-0.23	-0.23	1	5	100
GOOD N				21.
MEAN				-0.0112
S.D.				0.1045

TABLE 4

PROPORTION OF INCORRECT RESPONSES

LOW	HIGH	N	PCT	CUM
0.38	0.38	1	5	5
0.00	0.00			
0.00	0.00			
0.22	0.22	1	5	10
0.16	0.17	2	10	19
0.10	0.10	1	5	24
0.06	0.09	6	29	52
0.00	0.01	2	10	62
-0.02	-0.01	3	14	76
-0.05	-0.05	1	5	81
-0.13	-0.10	2	10	90
-0.19	-0.18	2	10	100
GOOD N				21.
MEAN				0.0374
S.D.				0.1323

TABLE 5

PROPORTION OF TIMES NO RESPONSE WAS GIVEN

LOW	HIGH	N	PCT	CUM
0.10	0.11	3	14	14
0.00	0.00			
0.00	0.00			
0.00	0.00			
0.05	0.06	3	14	29
0.04	0.04	1	5	33
0.00	0.00			
0.01	0.02	2	10	43
0.00	0.01	7	33	76
-0.01	-0.01	1	5	81
-0.02	-0.01	3	14	95
-0.04	-0.04	1	5	100
GOOD N				21.
MEAN				0.0217*
S.D.				0.0428

*This means is significantly different from zero ($p < .05$)

TABLE 6

PROPORTION OF RESPONSES IN ENGLISH

LOW	HIGH	N	PCT	CUM
0.21	0.25	3	14	14
0.00	0.00			
0.00	0.00			
0.10	0.10	1	5	19
0.06	0.06	1	5	24
-0.01	0.02	10	48	71
-0.03	-0.02	2	10	81
-0.08	-0.08	1	5	86
-0.14	-0.10	2	10	95
0.00	0.00			
0.00	0.00			
-0.27	-0.27	1	5	100
GOOD N				21.
MEAN				0.0118
S.D.				0.1177

TABLE 7

PROPORTION OF INTERACTION DURING WHICH THE TEACHER PROMPTED THE STUDENT OR CLASS IN FRENCH IN ORDER TO ELICIT A RESPONSE

LOW	HIGH	N	PCT	CUM
0.18	0.21	4	19	19
0.15	0.15	1	5	24
0.10	0.13	3	14	38
0.09	0.09	1	5	43
0.02	0.04	3	14	57
-0.00	0.00	3	14	71
-0.03	-0.03	1	5	76
-0.11	-0.10	2	10	86
0.00	0.00			
-0.18	-0.18	1	5	90
-0.20	-0.20	1	5	95
-0.28	-0.28	1	5	100
GOOD N				21.
MEAN				0.0250
S.D.				0.1367

TABLE 8

PROPORTION OF INTERACTIONS DURING WHICH THE TEACHER PROMPTED THE STUDENT OR CLASS IN ENGLISH IN ORDER TO CORRECT A RESPONSE

LOW	HIGH	N	PCT	CUM
0.40	0.40	1	5	5
0.31	0.31	1	5	10
0.00	0.00			
0.00	0.00			
0.14	0.14	1	5	14
0.09	0.12	3	14	29
0.03	0.06	5	24	52
-0.01	0.03	2	10	62
-0.06	-0.03	2	10	71
-0.10	-0.10	1	5	76
-0.16	-0.13	3	14	90
-0.23	-0.19	2	10	100
GOOD N				21.
MEAN				0.0167
S.D.				0.1553

TABLE 9

PROPORTION OF INTERACTIONS DURING WHICH THE TEACHER PROMPTED THE STUDENT OR CLASS IN ENGLISH

LOW	HIGH	N	PCT	CUM
0.06	0.07	2	10	10
0.00	0.00			
0.05	0.05	1	5	14
0.00	0.00			
0.00	0.00			
0.00	0.00			
-0.01	0.00	12	57	71
-0.02	-0.01	4	19	90
0.00	0.00			
-0.05	-0.05	1	5	95
0.00	0.00			
-0.07	-0.07	1	5	100
GOOD N				21.
MEAN				0.0002
S.D.				0.0318

TABLE 10

PROPORTION OF INTERACTIONS DURING WHICH THE TEACHER MOVED TO ANOTHER STUDENT TO GET A RESPONSE

LOW	HIGH	N	PCT	CUM
0.13	0.13	1	5	5
0.00	0.00			
0.08	0.08	1	5	10
0.00	0.00			
0.04	0.04	1	5	14
0.02	0.03	3	14	29
-0.00	0.01	6	29	57
-0.02	-0.02	3	14	71
-0.03	-0.02	4	19	90
-0.05	-0.05	1	5	95
0.00	0.00			
-0.10	-0.10	1	5	100
GOOD N				21.
MEAN				0.0017
S.D.				0.0459

TABLE 11

PROPORTION OF INTERACTIONS DURING WHICH THE TEACHER USED ENGLISH

LOW	HIGH	N	PCT	CUM
0.26	0.26	1	5	5
0.00	0.00			
0.14	0.14	1	5	10
0.10	0.10	1	5	14
0.00	0.00			
0.02	0.05	3	14	29
-0.02	0.00	9	43	71
-0.07	-0.04	4	19	90
-0.08	-0.08	1	5	95
0.00	0.00			
0.00	0.00			
-0.24	-0.24	1	5	100
GOOD N				21.
MEAN				0.0024
S.D.				0.0954

TABLE 12

AVERAGE QUALITY RATING OF RESPONSES

LOW	HIGH	N	PCT	CUM
1.11	1.11	1	5	5
0.77	0.81	2	10	14
0.61	0.68	3	14	29
0.39	0.44	2	10	38
0.16	0.29	5	24	62
-0.10	-0.03	4	14	76
-0.25	-0.22	2	10	86
-0.39	-0.39	1	5	90
0.00	0.00			
0.00	0.00			
-0.96	-0.96	1	5	95
-1.31	-1.31	1	5	100
GOOD N				21.
MEAN				0.1538
S.D.				0.5792