

Test Development as Transformational Process: The Case of the Maritime Oral Communication Assessment Portfolio

Sally Rehorick and Joseph Dicks

*"Few devices are as powerful, or are capable of
dictating as many decisions, as tests."*

Elana Shohamy, 1993

The role of testing in the educational system has received considerable attention over the past few years among educators, bureaucrats and members of the general public (Parkins, 1992). Demand for standards and accountability in education has increased in direct proportion to reports in the media and elsewhere comparing the poor results of Canadian students on international tests with the results of students from such countries as Japan and Germany (see, for example, The Commission on Excellence in Education, 1992). For some, testing is seen as a way to ensure this accountability; for others, a move towards more external, standardized testing has a potentially dangerous reductionist effect on student and teacher performance (Smith, 1991; Steinberg, 1988).

Howard Gardner (1988) has made the point that "most test reform has made . . . the assumption that teachers are obstacles to be circumvented; and if we could simply dictate what tests they will have to give we could force them into drumming stuff into kids' heads" (p. 5). Student competence cannot be assessed effectively "in an artificial way in an artificial setting," Gardner says; assessment needs to be engaging, contextualized and mindful of the different intelligences each individual student possesses. In successful assessment endeavours, the teacher, far from being an obstacle, is an active agent of change who plays an integral part in promoting classroom practices which contribute to the systemic validity of the testing (Frederiksen & Collins, 1989).

In this article, we will describe a French second language (FSL) test development project, the Maritime Oral Communication Assessment Portfolio (MOCAP), whose primary goal was to have a positive washback effect on classroom practice. During the course of the project, it became clear that the washback effect of the test extended beyond the microcosm of the individual classroom however; at a macro level, the project influenced processes for future interprovincial collaborations in such areas as curriculum development and human resource development. We will outline the process and products of the development within a framework of organizational change which proved effective in ensuring the acceptance of the testing system by both administrators and teachers.

The Impetus for the Project

This FSL test development project was requested by the Maritime Provinces Education Foundation (MPEF), a consortium of departments of education of the three

Sally Rehorick teaches English at Miyazaki International College where she is conducting research in second language acquisition and curriculum design. She has published three edited books, the most recent of which is *Second Language Acquisition and Teaching: Linking Theory with Practice* (1995) and was editor of the *Canadian Modern Language Review* from 1989-95.

Joseph Dicks is Assistant Professor of second language education at Université Sainte-Anne, Nova Scotia, Canada. He has conducted research in the areas of experimental and analytical features of L2 classrooms, and assessing L2 skills.

maritime provinces—New Brunswick, Nova Scotia and Prince Edward Island. The MPEF exists primarily to promote and facilitate collaborative educational endeavours such as curriculum and program development. The FSL test development project was the first effort at collaboration in the field of second languages.

The need for interprovincial collaboration reflects a country-wide trend in education, which has historically fallen under the jurisdiction of individual provinces. Dwindling resources together with the recognition that commonalties do and should exist in education have prompted provinces to redouble their efforts for cooperation (Lewington, 1994; Lazaruk, 1994).

A steering committee, composed of one person from each of the three provincial departments of education (two FSL consultants and one evaluation specialist) and the chairperson from the MPEF worked closely with the project director and project officer (the authors of this article) from the French Second Language Teacher Education Centre at the University of New Brunswick.

The initial goal of the project was two-fold:

- a. To provide a pool of items for oral interaction tests for use by individual core French and French immersion classroom teachers at the Grades 6 and 9 levels.
- b. To develop a standardized method for scoring the oral interaction test items.

The steering committee members agreed that the tests should reflect the communicative approach to language learning and teaching in order to complement the aims of their FSL programs. Although a common curriculum did not exist at that time, the steering committee felt that the underlying philosophy of the communicative approach was sufficiently prevalent in all provinces that common tests could be used.

The Purpose and Design of the Tests: A Negotiated Process

In spite of the general agreement on the nature of the tests (i.e. communicative), there was far less agreement on the ultimate purpose which the tests were to serve. The disagreement was not immediately apparent because the discourse of the steering committee members was framed by the perception that, because the underlying philosophy of the second language curriculum was the same, a common test could meet the needs of all three provinces. So strong was this paradigm that some months had passed before the project team (of which the authors were a part) at the University of New Brunswick (UNB) perceived that the three provinces had, in fact, very different goals for the tests.

One province's representative on the steering committee envisioned the tests as external, standardized measures of students' language competence; the tests would be summative and would be administered yearly. Thus, test results would provide feedback to students, teachers, administrators and parents, as well as inform program decisions at the provincial level. The representatives from the other two provinces viewed the tests as internal, classroom tools to be used by classroom teachers on an ongoing basis throughout the year; test results would provide diagnostic and performance feedback to students and teachers which would in turn inform teaching decisions by the classroom teacher.

All three provinces considered positive washback on classroom practice to be a primary goal. In addition, each supported the view that teachers should be involved in the development of the tests in order to increase the potential for washback. However, the more instrumental view of the potential impact of the tests held by the

one steering committee member contrasted sharply with the possibility for conceptual impact held by the other two. Shohamy (1993) characterizes instrumental impact as "short-range and goal-oriented" while conceptual impact is "long-range and meaningful" (p. 17). The different aspirations for the test were driven in part by the roles which the provincial representatives played in their own departments of education and in part by surprisingly different theoretical perspectives reflected in the second language programs of the three provinces.

These different perspectives were not evident in the curriculum documents provided to the project team (i.e. project director and project officer) by each province. On the contrary, a comparison of the official curricular documents revealed striking similarities, a fact which comforted the project team in the initial stages of defining the test specifications. Two examples of statements contained in these documents are the following:

Students should be encouraged to use their new language in a context that is relevant and of interest to their age group. . . The experiential aspect of learning will be emphasized. . . This active, experiential approach fosters the positive attitudes so necessary to second language learning. (Province of New Brunswick, 1988, pp. ix, 2)

Communicative language teaching should be organized according to themes which are important and of interest to the learners. . . . The learners must contextualize what is to be learned (Province of Nova Scotia, 1988, pp. 9, 16).

In official statements about the role of evaluation, these similarities seemed to persist:

The emphasis on . . . evaluation both oral and written must be consistent with the objectives and learning activities. . . . The evaluation of linguistic skill should always be linked to a communicative objective. (Province of Nova Scotia, 1988, p. 75)

Tests and evaluation procedures are inextricably tied to teaching methods and goals. . . . It is important to consider language functions, authentic speech acts, real-language activities and the like. . . . (Province of New Brunswick, 1988, p. 26)

It was only during the initial discussions of the test specifications with the steering committee members that the project team realized the differing theoretical perspectives of the provinces. In New Brunswick, teachers had been influenced by another test which had been in use for over a decade in that province. The Oral Proficiency Interview is administered to all students registered in a French course who are graduating from New Brunswick high schools. This testing endeavour influenced second language teaching in the province in two major ways. First, it succeeded in ensuring that more classroom time was spent on oral interaction; second, it redirected teachers' notions of what constitutes language competence from a largely structural view of language (for example, grammar, pronunciation, vocabulary) to a framework of communicative competence in which speech functions played a major role (Rehorick, 1991). Thus teachers organized classroom and evaluation activities around such functions as "narrate and describe," "give instructions," "support opinions" and "persuade and counsel." The test design envisioned by New Brunswick for the MPEF project was for a bank of items, functionally-based, from which could be drawn a series of parallel tests to be administered province-wide on a given day during the school year.

In Nova Scotia and Prince Edward Island, joint projects of curriculum reform based on the National Core French Study (NCFS) were underway. This study (see Rehorick & Edwards, 1990, for a summary of the NCFS) established the framework for a multi-dimensional curriculum composed of four syllabi: communicative/experiential, language, general language and cultural. One of the driving principles of this approach is that all units of study are based on themes and fields of experience of

the student population. Language functions, so prevalent in the New Brunswick classrooms, were defined in the Nova Scotia/Prince Edward Island curriculum only within the overall thematic units. For these two provinces, therefore, the test design for the MPEF project needed to be based on the various themes developed during the year in each program.

Since the purpose which a test is to serve largely determines its design (Bachman, 1990; Shohamy, 1985), progress on the development of these FSL tests came to a standstill. The project team could not proceed with test specifications, training of teachers to assist with the test development, or scheduling of pilot tests. The impasse created by the differing needs of the three provincial representatives seemed for a time to be insurmountable. The need for the public accountability which a standardized test could serve seemed to conflict with the need for classroom evaluation tools which would be part of every teacher's arsenal.

The Transformed Role of the Project Team

When the MPEF contracted the French Second Language Teacher Education Centre to develop the French as a second language tests, it did so with the understanding that there was already agreement on the purpose of the tests. As project leaders knowledgeable in test development, we were to design and implement communicative FSL tests. When the discrepancies in direction were detected by the project team, our role as project leaders underwent an evolution from expert consultants to process consultants. Up until that time, we had participated as information specialists who, "through [our] special knowledge, skill, and professional experience, [are] engaged as . . . outside consultants to provide special-knowledge services" (Lippett & Lippett, 1978, p. 33). This role was transformed into that of the process consultant who, according to Schein (1969), "seeks to give the client 'insight' into what is going on around him, within him, and between him and other people" (p. 9).

Before continuing, we knew we had to facilitate a dialogue among the steering committee members in such a way that a new collective vision could be established for the tests. In order to ensure that new vision would be shared, this dialogue had to proceed in a risk-free environment. Watkins and Marsick (1993) note "the power of talk in bringing about dialogue—by telling what is on one's mind, asking questions about its impact, listening for the reasoning in people's answers, and keeping open to new viewpoints" (p. 13). Because of the physical distances separating committee members, much of the dialogue which ensued between the steering committee and the project team took place over the telephone, a situation which challenged us all.

As process consultants, we were able to draw out what seemed to be the principal commonalties on which the steering committee could base their revised ideas for the tests. The need for public accountability expressed by one province was in fact an underlying theme for the other two provinces as well. Environmental scans provided to the project team revealed that this accountability was a desirable goal for many aspects of public education in the three maritime provinces. The main difference among the three provinces was the time frame for this goal to be accomplished. Nova Scotia and Prince Edward Island viewed the process to be long term and only possible once they were assured that teachers had had sufficient time to reach a comfort level with the testing methods and the corresponding teaching methods; in other words, the washback on classroom practice had to occur before standardized testing could be implemented. New Brunswick, on the other hand, expressed a need for a somewhat shorter timeline, but recognized the merits of ensuring that teachers were prepared in an appropriate manner.

Major systemic change such as that represented by this testing system needs to be preceded by careful assessment of the readiness on the part of the teachers for the change to occur. If this assessment of the climate for change does not occur, the washback on the classroom is likely to be both short-lived and negative (Shohamy, 1993). Change brought about by and for only one part of the education system (as indeed with any system) cannot have the impact of systemic change (Watkins & Marsick, 1993). Systemic change recognizes that the various components of an organization interlock and interact in numerous ways. The learning required to effect this change in a positive way depends on continuous collaboration and, when the collaboration is successful, this learning can transform the organization to "work in new, fundamentally different ways [to give it] wholly new capacities" (Watkins & Marsick, 1993, p. 11).

The discussions about accountability and the gradual acknowledgment that it was indeed a mutual goal lead to a consensus concerning the role of teachers in developing and implementing the testing project. To avoid any potential conflict between teachers and administrators regarding the format and use of the tests, the steering committee agreed that involvement of teachers throughout the project was an essential component. Teachers were thus viewed as pedagogical experts whose opinions and expertise were valued. This belief set the foundation for a level of commitment from teachers which proved to be a key factor to the success of the project, particularly with respect to the washback effect.

New Guiding Principles of Test Development

As a result of the learning which occurred during these initial months of discussions with the steering committee, the project team revised the two goals for the project. Because of the shift from product to process, the revised goals were stated as "guiding principles":

1. to devise an instrument on the basis of a theoretical framework of communicative competence which is compatible with the approaches to French as a second language instruction in the provinces involved.
2. to develop a test which reflects the curriculum aims and objectives shared by the three provinces. These aims and objectives would emanate from a common theoretical perspective as described in #1 above.
3. to develop this test such that the administration and scoring could be easily carried out by classroom teachers.
4. to develop this test such that the administration and scoring would be consistent from one testing situation to another allowing meaningful comparisons to be made among classes in different areas.
5. to involve teachers in the process of test development so that
 - a) the items written reflect a consensus among teachers from the three provinces regarding language forms, functions and context;
 - b) the test be piloted by these teachers in their classrooms so that, if necessary, items could be revised accordingly.
6. to develop a test instrument such that the development process and the final product ensure that the theoretical framework of communicative language ability and the aims and objectives which stem from it are reflected in both teaching and testing practices at the classroom level.

By articulating these principles in this way, we were able to make our new learning and agreement visible. The discourse of these principles demonstrated the key notions of accountability (#4), teacher involvement (#3,5), interprovincial collaboration (#1,2) and washback (#6) which had been leitmotifs of the discussions among the members of the steering committee and project team.

Teacher Selection and Involvement

The decision to involve teachers in every aspect of test development and implementation had major implications for the role of the project team. As well as having the role of process consultants and information specialists, we would now also assume the role of consultants as trainers/educators. As noted by Lippett and Lippett (1978, p. 34), "the trainer/educator role of the consultant may be essential in designing and facilitating a learning process or an organization-change process." In the case of the testing project, our knowledge of the theoretical differences in FSL learning and teaching among the three provinces (i.e. functionally-based versus thematically-based) lead us to believe that these differences in philosophy needed to be acknowledged and dealt with early on in the process. Failure to do so could result in disagreements which could conceivably block the development of the tests themselves.

Each province selected ten teachers to participate in the project. The project team provided guidance to the steering committee by designing a competency profile (Table 1) outlining the knowledge, skills and attitudes required by the resource teachers.

The competencies outlined in the profile underscore the role played by the teachers in the change process. These were teachers who were not only subject-matter experts recognized for their excellence in the classroom but, in addition, were proactive and adaptable individuals, comfortable with working with new ideas. Each of the teachers selected was considered to possess most of the competencies on the profile. It goes without saying that teachers who have these characteristics are sought out frequently by administrators for a variety of projects and problem-solving endeavours. Because of the magnitude of the involvement required for the FSL testing project, we were concerned that the teachers might feel overburdened and would leave the project prematurely. As Barth (1990, p. 135) has pointed out, "a tired teacher is a tired teacher." Our dual challenges were clear: to provide the intensive training needed by the resource teachers to create the tests, and to do so in a positive climate which would ensure the ongoing commitment and feeling of ownership of the test project.

The Item Writing Workshop: Setting the Climate for Change

The four-day workshop in which the resource teachers participated provided us with the necessary scaffold to develop sufficient expertise among the teachers for creating and piloting test items. The workshop was also the only time for us to work with the full group present and thus we needed to establish quickly a learning environment appropriate for developing a positive climate for the change process.

To accomplish these two goals, we integrated some key principles for adult learning and adopted the multiple roles of facilitators/trainers/team builders. Each teacher had received a package of information prior to the workshop which described the project and their role in its development. Although we had already developed a set of test specifications based on our prior work with provincial curriculum documents, we decided not to distribute this prior to the workshop; teachers were asked, however,

RESOURCE TEACHER FSL TEST DEVELOPMENT PROJECT COMPETENCY PROFILE

The resource teachers will be the most important component of the test development, ensuring that the final product will be appropriate to their needs.

RESPONSIBILITY STATEMENT

To participate in the development and implementation of oral interaction tests of French as second language for Grades 6 and 9 for use in Nova Scotia, Prince Edward Island and New Brunswick.

GENERAL OBJECTIVE

Under the guidance of the Project Director and Project Officer, provide tests which reflect the communicative approach currently being used in school curricula.

COMPETENCIES REQUIRED

A. KNOWLEDGE

- possess thorough knowledge of the philosophy and practice of the communicative approach to learning a second language.
- possess native or near-native French language skills.
- basic communication ability in English.

B. INTERPERSONAL COMMUNICATION

- ability to work as an effective team member.
- demonstrate commitment to project and fellow Resource Teachers.
- communicate effectively with project team (Resource Teachers, Project Director, Project Officer) to ensure that ongoing feedback concerning the project is maintained.

C. PROBLEM SOLVING/DECISION MAKING

- interpret information accurately and consistently, seeking appropriate clarification when necessary.
- identify problems by determining the basic issues and formulating a strategy for dealing with the problem.
- take action promptly after a problem has been identified to clear up the situation in an effective and timely manner.

D. CHANGE MANAGEMENT/LEADERSHIP

- recognize that the test items may represent an entirely new way to thinking about the question of student evaluation and be prepared therefore to be innovative and imaginative in developing test items.
- adapt quickly and willingly when new information or changing priorities affect the project.
- keep the overall vision for the new tests firmly in mind so that the final result achieves the objectives of the project.
- be recognized as a competent teacher by colleagues in order that the piloting and implementation of the tests are facilitated.

Table 1

to bring classroom materials including authentic documents which they felt represented their work.

As trainers, we saw our role to articulate the specifications for the test items, to demonstrate how to develop items from within the chosen themes while using the classroom materials which the teachers had brought, and to respond to the teachers' progress as the items were developed. As facilitators, we assisted the teachers with making decisions about such areas as the selection of themes to develop and with staying on task when energy levels were low.

Our role as team builders, however, was arguably the most crucial aspect to the success of the workshop. The thirty teachers, most of whom did not know each other prior to the workshop were required to work effectively in teams grouped according to grade level and program taught (i.e. either core French or French immersion). The process of learning together, or the "team process", evolved through four predictable phases: fragmented, pooled, synergistic, and continuous (Watkins & Marsick, 1993, p. 97). These phases can be attributed to the naturally volatile and developmental nature of teams, a characteristic succinctly summarized by Watkins and Marsick (1993, p. 97): "Teams are crucibles through which opposing ideas can be

brought together and confronted—ideas that otherwise would remain within the heads of individuals and not be linked together in new combinations."

As project leaders, we did not take for granted our resource teachers' ability to work effectively in teams, especially since the team members did not know each other prior to the workshop. Teachers are often isolated within their own schools, frequently so overwhelmed with the day-to-day business of their own classrooms that reflecting upon and sharing their experiences with colleagues is burdensome at best and invasive at worst. Roland Barth (1991, p. 33) quotes one teacher's assessment of the school classrooms as "adjoining caves" from which teachers emerge occasionally to chat about administrative or discipline issues but rarely about the craft of teaching. Since the resource teachers were to be the catalysts for the successful development and implementation of the tests, we conducted the workshop with the goal of creating a community of teacher/leaders who would gain the ability to work interdependently and to share their insights with others in their school environment.

Team Development: Four Phases of Learning

Fragmented phase

After some icebreaker-type activities, we began the workshop with an overview of the testing project and a discussion of the teachers' views of evaluation. This discussion proved very instructive to the project team because it revealed the suspicion which the teachers held about the purpose of the tests. Many participants felt that the ultimate reason for the interprovincial collaboration was to make teachers conform to a common curriculum and methodology. In other words, these teachers perceived "washback" as manipulative and negative. These perceptions seemed to be framed by prior experiences which some participants had had with the use of tests. In order to resolve these difficulties, the project leaders facilitated a discussion between representatives of the provincial departments of education and the workshop participants which lead to the following points of mutual understanding:

1. The teachers had been selected for the resource team because of their subject-matter expertise and their ability to be educational leaders.
2. The project team leaders were facilitators who would be relying on the expertise of the teachers.
3. The teachers would be involved in ongoing decisions about test design and redesign, piloting, and training of other teachers in the use of the test.
4. The tests would be a tool for teachers to use in the classroom, as a source of pedagogical information; standardization of scoring and reporting of overall test results would only be attempted after teachers had received sufficient time to be trained in the testing techniques and to adapt their pedagogical practices accordingly.

These initial discussions served one major purpose in the team learning process: The workshop participants knew that their opinions counted and that they could influence the decision-making process. The feeling of empowerment which flowed from the discussions enhanced ownership and commitment of the project as a whole.

The first task for the teams was to select themes and content which the teachers deemed to be common to FSL classes at the particular grade level taught. There were four teams, each representing a grade level and program: Grade 6 core French, Grade 6 immersion, Grade 9 core French and Grade 9 immersion. The teachers experienced a degree of fragmentation in this phase because many of them

felt that defining common themes and objectives was not feasible given the individuality of each FSL classroom and the different curricula of the three provinces. Their discussion remained at the level of abstraction, however, and they did not share the specifics of their own practices.

Pooled phase

As the nature of the team interaction became apparent, the project leaders realized that feedback on the team learning which was occurring was needed. We pointed out that they needed to orient their discussions about themes to specific topics and content. In order for them to begin this phase productively, we provided them with the lists of fields of experience from *The Communicative/Experiential Syllabus* of the National Core French Study (Tremblay, Duplantie & Huot, 1990). This list includes such themes as food-related experiences, experiences with travel, self-protection experiences, family-related experiences and experiences with conservation. Once the teachers had perused these documents, they began to share their own practices and classroom themes. The discussions thus progressed from the abstract to the specific although, typically, the degree of specificity remained fairly superficial. The team members did not, for example, move beyond the mere naming of themes used in their classrooms.

Synergistic phase

The project leaders coaxed the participants to select one theme to which everyone on their team could agree and to specify the scope of learning objectives and activities which would be appropriate for pupils at that level. Watkins and Marsick (1993, p. 107) say that "at the level of synergy, the team jointly constructs shared meanings, assumptions, and language, which leads to consensually developed solutions, positions, and recommendations." By this time, our teams had progressed enough in their learning that they were able to capitalize on their mutual strengths and to discuss points of difference openly. Within a relatively short period of time, the teachers were ready to develop some test items based on specific thematic content.

Continuous learning phase

According to Watkins and Marsick (1993), this phase of team learning is rarely reached. It involves the exportation of synergistic learning to other teams thereby creating a positive network of continuous learning throughout an organization. In other words, the learning of one team has a positive washback on the interaction of other groups. In the case of our resource teachers, this continuous learning was demonstrated by their willingness to assist the other teams and to share their progress and results with them. We wanted to encourage ongoing sharing with team members and other colleagues after the workshop was concluded as well since we knew that this was to be a key component of the organizational change process which we were promoting.

As the workshop progressed, the teachers went through the four phases of team learning continually. As each new task, such as the drafting of test items, was presented, the new learning required periods of fragmentation, pooling, synergy and continuous learning. The length of time for each phase, however, changed substantially. By the end of Day 3, the teachers had developed their abilities to work productively together to such a degree that the fragmented and pooling phases were virtually nonexistent. This is not to say that there were no disagreements. Rather the divergent views which did exist were integrated into new ways of thinking among the teachers. They dealt with any conflicts with the conviction that they were capable of solving them.

The Product: The Components of MOCAP

As a result of the collaborative effort of the thirty resource teachers over the course of the four day workshop, test items were developed by each of the four teams for their particular grade and program. The resulting package comprises four evaluation modules each representing a different theme and a different grade and program. Because MOCAP is thematically based and because teachers typically incorporate the study of numerous themes in the course of one year, the four modules represent in fact a prototype for this kind of oral testing. Student evaluation is meant to be continuous and contextualized within the thematic unit. Assessment is integrated into the normal activities of the classroom; in fact not all students might be tested at the same time and every student might not be asked to perform each evaluation task. The results of each student's testing makes up the individual portfolio of that student. The performances of the students are thus tracked over time during the year. Using the principles of portfolio assessment, we ensured that the testing would be task-based, positive and continuous (Belanoff & Dickson, 1991; Dicks & Rehorick, 1995; Thorogood, 1992; Underhill, 1987; Wesche, 1987).

As noted above, there are in effect four versions of MOCAP—one version for each grade and program: 6 core, 9 core, 6 immersion, 9 immersion. Three of the evaluation techniques are basically the same for all four levels—role plays, information gaps, discussion. Two other techniques vary from core to immersion with forms and describing pictures being used in core French and questionnaires and oral summaries being used in French immersion. The thematic content, language tasks, and linguistic content varies from one version to another. The following is a listing and description of the various evaluation techniques.

Describing pictures

This technique is designed for core French at both grades 6 and 9. It consists of a series of illustrations that students are required to look at and talk about. This may take the form of description or narration or some combination of both. Essentially, students are instructed to look at a sequence of illustrations and to describe and/or narrate the content of these drawings. The teacher's involvement should be minimal in such situations. The student is allowed two minutes to examine the illustrations before beginning the activity. It should be recognized that pauses during the activity could be meaningful and eventually lead to productive expression, and teachers should only intervene where necessary to keep the activity going. The extent to which the teacher has to provide clues, ask questions, and help the student will be reflected in the evaluation.

Oral summary

This technique is destined for use in French immersion classes at both grade levels. Students are required to listen to an oral, tape-recorded passage twice, and then, in their own words, give an account of the main points of the passage. Before listening to the passage which is recorded on audio-cassette (supplied in the assessment package), students are told that the purpose of this activity is for them to understand the principal elements of the oral passage and to reformulate these in their own words. Students are told that they are not expected to remember secondary details. They are also reminded that the passage will be played twice. Again in this situation the teacher's involvement should be minimal. A question may be asked or a comment made to keep the activity going if the student is obviously stuck on a point, but otherwise the teacher's role should be passive.

Forms

This technique involves an interview between two students assisted by a form that one of the students must complete. These forms which are supplied in the assessment package may be applications for a summer camp, customs forms, and so forth that generally require factual information. The student who holds the form is the one who will be required to ask the question. It should be made clear to this student that the questions are not written in complete sentences, but rather in point form. This student's task is to formulate the required questions and to ask these in a manner that the other student can understand. The second student must answer these questions to the best of his or her ability. It is not necessary for the first student to complete the form as it is only meant as a prop and an aid to the student in formulating questions.

Questionnaire

Similar to the forms above, this technique designed for use in immersion involves an interview between two students assisted by a form. In this case, in addition to factual information, the questionnaire which is supplied in the assessment package requires students to ask and respond to questions involving opinions, attitudes, and so on. The student who holds the questionnaire is the one who will be required to ask the questions. It should be made clear to this student that the questions are not written in complete sentences, but rather in point form. This student's task is to formulate the required questions and to ask these in a manner that the other student can understand. The second student must answer these questions to the best of his or her ability. It is not necessary for the first student to complete the questionnaire as it is only meant as a prop and an aid to the student in formulating questions.

Information gap

This technique, which is used in both core and immersion, involves two students working together in order to obtain the information required to complete a specific task. One student possesses a document containing the complete or accurate set of information that the other does not have. These documents are supplied in the assessment package and are given to students at the beginning of the evaluation session. In some cases, the actual transfer of information is sufficient to complete the task while in other situations students must use the new information to answer specific questions related to it. It should be made clear to both students that their general task is to work together to provide one another with any information that may be missing from their respective documents.

Role plays

This technique involves two students in a face-to-face encounter. The students are provided with the student documents before the role play which provide them with background information and specific instructions as to the information they must request and provide. The students are given two minutes to examine the document and to ensure they understand the instructions completely. They may ask questions of clarification, but should not be allowed to rehearse their part.

Discussion

This technique involves four students in a round-table type exchange. The students are provided with a document that provides background information and specific instructions as to the language tasks that they are expected to accomplish. Students are given two minutes to read the document so they can understand the

instructions and formulate their ideas. They should not be allowed to rehearse however. In core French the exchange takes the form of a conversation or question and answer session whereas in immersion the event is much more debate-oriented. Students should be told before beginning that in order for the teacher to evaluate they all must participate in the discussion. With a group of four, the teacher may have to intervene in order to give less vocal students an opportunity to express themselves. It is also recommended that two teachers be involved in the evaluation of the group discussion as it is very difficult for one teacher to fairly evaluate four students at once.

One of the distinguishing features of all the techniques (except *describing pictures* and *oral summaries*) is that the teacher acts as an observer who evaluates student-to-student interaction. The reason for this is two-fold. First, one of the goals of evaluation is to assess authentic, meaningful communication; by removing herself/himself from the interaction, the teacher is much more likely to elicit authentic speech samples from the students. Second, most teachers are using a variety of techniques for interactive group work (for example, cooperative learning) during their classes; by evaluating students in interactive situations, a teacher ensures that the assessment methods match classroom practice.

Pilot Testing and Implementation

After the item-writing workshop the project team edited and formatted all the material developed by the resource teachers. The teachers tested all the items in their own classrooms and revisions were made by the project team according to the feedback received. A second pilot test was conducted for those items which had undergone substantive revisions. The resource team was asked to conduct the second pilot with teachers in their schools who had not been a part of the original team. This approach put the resource team teachers in the role of trainers of other teachers. Anticipating that the resource team members would eventually conduct the train-the-teacher workshops in their respective province, we were endeavouring to determine to what extent they would be comfortable in this role by trying it out in the smaller scale of their own schools.

It is perhaps not surprising that very few of the teachers were able to carry out the task of having another teacher conduct the second pilot easily. Although our resource teachers had reached an advanced level of team development among themselves which permitted ongoing interactions, they did not feel comfortable in taking a leadership role in their schools. Part of this problem can be attributed to the isolation inherent among teachers in schools which are not structured for teachers to take on leadership roles (Barth, 1991; Wasley, 1991). However a more subtle, yet equally powerful, reason is that teachers are reluctant to appear different from their colleagues. As one teacher said to the project team, "I did not want to appear to be showing off." A lack of role models plus a lack of a clear definition of the teacher leadership role within schools are at the root of this teacher's dilemma. As Wasley (1991, p.147) says, "the discussion of teacher leadership assaults the egalitarian norms that have long been in place in teaching."

Our observations of the resource team members during the pilot testing informed our recommendations for implementation of MOCAP throughout the three provinces. Although the resource teachers were now fully conversant with MOCAP, this expertise was not sufficient for them to train other trainers. In other words, in order for our teachers to take on successful leadership roles in the implementation of the testing package, another developmental stage was required. The teaching of adults is not part of either preservice or inservice teacher education programs. It was this preparation which we needed to give to our resource teachers before they would

be ready to help implement MOCAP with confidence. So we developed a train-the-teacher workshop manual (Rehorick & Kristmanson, 1994) which includes an outline of adult learning principles, a complete workshop schedule including scripts and learning activities for both one-day and two-day versions of the workshops.

Learnings and Concluding Remarks

According to Wesche (1987, p. 40) one of goals for performance testing, of which MOCAP is but one example, is "user satisfaction." Since one of the original goals for MOCAP was a positive washback effect, we had to determine who in fact the users of our testing system would be and on which part of the educational system the positive washback was meant to occur.

If we consider the project in its initial stages, the answers to those questions are straightforward: The users are classroom teachers and the washback is meant to occur at the individual classroom level. Our experience of developing MOCAP, however, showed us that changes brought about at the classroom level can only happen when corresponding organizational changes take place as well. The "users" in the case of this project include not only the teachers but the steering committee members representing each province. And because the focus was at least partly on the process of developing the tests as much as on the product itself, a third user group was the team of resource teachers.

The potential for positive washback, then, is at three distinct levels. First, at the macro level of the three provinces, the development of MOCAP gave birth to a way for collaboration on other educational projects. A promising feature of this collaboration was the recognition of positive interdependence of the three constituents. The broadened perspective which resulted ensured the necessary climate for the project team to carry out the development of the FSL tests with the resource teachers. Second, the resource teachers developed team learning processes which shaped the way they conducted their interactions. For the resource teachers, the positive washback occurred through the interactions themselves and through the self-confidence resulting from the creation of the product. By the end of the year of development and pilot testing, they had taken full ownership of MOCAP: it was *their* product. The resource teachers became the spokespersons for MOCAP.

It is too early to assess the full impact of the third level of washback—the microcosm of the individual classroom. The full implementation of MOCAP takes time and the implications will have to be tracked over time. MOCAP was not designed as a "quick-fix" solution but rather a long-term and continuous systemic change process. As the use of MOCAP becomes more widespread, we expect to see the effects of its use in the satisfaction expressed by teachers and, ultimately, by students and their parents.

Acknowledgment

The authors gratefully acknowledge the project funding provided by the Maritime Province Education Foundation. The opinions expressed in this article are those of the authors and not necessarily those of the MPEF. This article has been submitted as a chapter of *French as a Second Language in Canada: Recent Empirical Studies* (Ed. S. Lapkin). The authors wish to thank the editor for permission to print this article in *Comparative Culture*.

References

- Bachman, L.F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Barth, R. (1990). *Improving schools from within. Teachers, parents and principals can make the different*. San Francisco: Jossey-Bass Publishers.
- Belanoff, P. & Dickson, M. (Eds.) (1991). *Portfolios. Process and product*. Portsmouth, NH: Boynton/Cook Publishers. Heinemann Educational Books Ltd.
- Dicks, J. & Rehorick, S. (1995). Oral communicative assessment and the portfolio concept. In G. Duquette (Ed.), *Second language practice: Classroom strategies for developing communicative competence*. London: Multilingual Matters Ltd.
- Downey, J. & Landry, A. (Commissioners). (1992). *Schools for a new century*. Fredericton, NB: Commission on Excellence in Education.
- Fredericksen, J. & Collins, A. (1989). A system approach to educational testing. *Educational Researcher*, 18 (9), 27-32.
- French Second Language Teacher Education Centre. (1992). *Manuel des examinateurs de français*. Fredericton, NB: Department of Education.
- Gardner, H. (1988). Alternatives to standardized testing. *The Harvard Education Letter*, 4 (5), 5-6.
- Lazaruk, W. (1994). *Survey: Interprovincial collaboration in core French*. Edmonton, AB: Alberta Education.
- Lewington, J. (1994, July 12). Joining forces on the three R's. Curriculum sharing is driven partly by money woes. *The Globe and Mail*, pp. A1, A5.
- Lippett, G. & Lippett, R. (1978). *The consulting process in action*. San Diego: University Associates Publishers and Consultants.
- McLean, L. (1990). Time to replace the classroom test with authentic measurement. *The Alberta Journal of Educational Research*, 36 (1), 78-84.
- Olson, M.W. (1991). Teaching resume content: Hiring officials' preferences versus college recruiters' preferences. *Business Education Forum*, 45 (7), 13-17.
- Parkins, D. (1992). *Smart schools*. New York: The Free Press (division of McMillan Inc.)
- Province of New Brunswick (c. 1988). *French second language core program elementary/junior high school*. Fredericton: Department of Education.
- Province of Nova Scotia (c. 1988). *Elementary core French guide*. Halifax: Department of Education.
- Rehorick, S. & Dicks, J. (1991). *Maritime Provinces Education Foundation French as a second language testing project (Interim Report)*. Fredericton: University of New Brunswick, French Second Language Teacher Education Centre.
- Rehorick, S. & Edwards, V. (1990). *The National Core French study*. Welland, ON: The Canadian Modern Language Review, 47 (1).
- Rehorick, S. (1991). *New Brunswick French oral proficiency interview project*. Fredericton: University of New Brunswick, French Second Language Teacher Education Centre.

- Rehorick, S. (1992). *Maritime Provinces Education Foundation French as a second language testing project. Final report*. Fredericton: University of New Brunswick, French Second Language Teacher Education Centre.
- Rehorick, S. (1993). *French second language learning in New Brunswick schools. Paradigms, challenges and strategies*. Fredericton: University of New Brunswick, French Second Language Teacher Education Centre.
- Rehorick, S. & Kristmanson, P. (1994). *MOCAP train-the-teacher workshop*. Fredericton, NB: French Second Language Teacher Education Centre.
- Schein, E. (1969). *Process consultation: Its role in organizational development*. Don Mills, ON: Addison Wesley Publishing Company.
- Scherer, M. (1992). The loneliness of the teacher leader [Review of *Teachers who lead: The rhetoric of reform and the realities of practice*.] *Educational Leadership*, 49 (5), 55.
- Shohamy, E. (1985). *A practical handbook in language testing for the second language teacher*. Tel-Aviv: Tel-Aviv University.
- Shohamy, E. (1993). *The power of tests: The impact of language tests on teaching and learning*. Washington, DC: John Hopkins University, The National Foreign Language Center.
- Shohamy, E. & Walton, R. (eds.). (1992). *Language assessment for feedback: Testing and other strategies*. Washington, DC: National Foreign Language Center Publications.
- Smith, M.L. (1991). Put to the test: The effects of external testing on teachers. *Educational Researcher*, 20 (5), 8-10.
- Steinberg, A. Testing: Is there a right answer? (1988). *The Harvard Education Letter* 4 (5), 1-4.
- Swain, M. (1984). Large-scale communicative language testing: A case study. In Savignon, S. & Berns, M. (eds.), *Initiatives in Communicative Language Teaching*. Reading, Massachusetts: Addison-Wesley.
- Thorogood, J. (1992). *Continuous assessment and recording. Pathfinder 13*. London: Centre for Information on Language Teaching and Research.
- Tremblay, R., Duplantie, M. & Huot, D. (1990). *The communicative / experiential syllabus. National Core French Study*. Ottawa: M Editeur.
- Underhill, N. (1987). *Testing spoken language*. Cambridge: Cambridge University Press.
- Wasley, P. A. (1991). *Teachers who lead: The rhetoric of reform and the realities of practice*. New York: Teachers College Press.
- Wesche, M. B. (1987). Second language performance testing: The Ontario Test of ESL as an example. *Language Testing*, 4 (1), 28-47.
- Watkins, K. E. & V. M. Marsick. (1993). *Sculpting the learning organization*. San Francisco: Jossey-Bass Publishers.