

Investigating Alternative Assessment In Two-Way Bilingual Immersion Programs

FINAL REPORT

**Submitted to the
U.S. Department of Education
Office of Bilingual Education and Minority Languages Affairs
by
Arlington Public Schools and the Center for Applied Linguistics**

**for the
Bilingual Education: Field-Initiated Research Program
CFDA Number: 84.292.B**

April 30, 1997

**Arlington Public Schools
Education Center
1426 N. Quincy Street
Arlington, VA 22207
703-358-6097**

EXECUTIVE SUMMARY

This field-initiated research project served to create a more coordinated process across schools involved in the Arlington Public Schools Spanish Partial-Immersion Program for assessing student language ability. The project provided alternative assessment instruments in schools that previously were not using any, and enhanced measures in schools where alternative assessment was already being implemented. The information provided by these alternative assessment instruments also illuminated areas of the participants' instruction that needed support so that they could better meet their students' language needs. The collaborative inquiry research approach constituted a coordinated effort to provide all limited English proficient and English proficient immersion students with better educational service and their parents with a better understanding of their children's progress, skills, and abilities. Due to the early termination of this project, however, the direct impact of this project on students and parents could not be assessed. The material products resulting from this project, however, have been and will continue to be widely disseminated to two-way immersion program administrators and staff by the Arlington Public Schools and the Center for Applied Linguistics.

OVERVIEW

The purpose of this research project was to explore alternative assessment of language proficiency in two-way bilingual immersion programs through the use of portfolios and other innovative assessment measures. The project focused on 1) how teachers in two-way bilingual immersion programs could use alternative means to assess oral and written language skills in Spanish and English, and 2) how the use of alternative assessment measures could influence instructional practices. The project was implemented in the belief that the development and use of alternative assessment measures could lead to improvement in the education of limited English proficient students by giving immersion teachers a more complete picture of their students'

language abilities and progress in language development. Further, this information could be made available to parents, who would, in turn, benefit from more detailed information about their children's language learning. Finally, this project aimed to develop a standard approach to more authentically assessing the language development of students in the three county elementary schools in which the immersion program operates.

This project was originally awarded a grant for two years, but saw the Year 2 funding eliminated from the federal budget along with all other research projects administered by OBEMLA. The project received a five month no-cost extension through March 1997, but even this provided insufficient time for all of the proposed goals and objectives to be achieved as thoroughly as originally proposed.

The project was a collaborative effort between the Arlington Public Schools (Arlington, VA) and the Center for Applied Linguistics (CAL). The research team included teachers from three schools involved in Arlington County's Spanish Partial Immersion Program, a representative from the middle school immersion program, a special education teacher, administrators from the school district office, and researchers from CAL (Appendix 1). A team of teachers from each school worked with researchers on developing, piloting, field-testing, and evaluating alternative means for assessing language ability. As a result of this project, rubrics for oral language and written language assessment in both Spanish and English were developed, piloted, and field-tested (Appendix 2). In addition, the research team explored ways in which using alternative assessment might inform and otherwise influence instructional practices. Teachers from all participating schools had many opportunities to share their assessment strategies, impressions of the effectiveness of the language rubrics, and classroom experiences throughout the brief duration of the project. The information derived from the collaborative process and the rubrics designed have been and will continue to be disseminated to those interested or involved in two-way programs throughout the United States.

BACKGROUND AND RATIONALE FOR THE PROJECT

Two major findings have led many educators of limited English proficient students away from teaching language in isolation and toward teaching language through content. These findings are: 1) native language development and academic achievement are not negatively affected as a result of participating in an immersion program (Genesee, 1994), and 2) language is learned best when it is learned in context (Brinton, 1989; Crandall, 1987, Genesee, 1987, Mohan, 1986). Consequently, many two-way immersion programs have adopted a pedagogical approach that makes language the *medium* of instruction rather than the *focus*. As a result, much of the student language development in these programs occurs implicitly—as the students master content, they also learn to use language for specific purposes. Authentic communication occurs in authentic settings, where there is little explicit focus on language itself.

Studies of immersion programs in Canada, however, have shown that progress in language development is not continuous and does not directly correspond to the amount of exposure to the second language the students receive (Genesee, 1994). In these studies, students who received more content-area instruction in the second language early in the program (early immersion) attained the same language proficiency level as those students who had less early exposure (late immersion). Further, Lyster (1994) has observed that, although students in French immersion programs received a great deal of input in their second language through their content courses, their second language competencies tended to continue to be non-native-like with respect to grammatical and sociolinguistic competencies.

To improve students' second language accuracy, some immersion programs have begun to incorporate a more explicit focus on the second language itself (Räsänen, 1994). European immersion models emphasize the explicit teaching of language as a subject

along with content instruction in the second language. This has been credited for promoting higher levels of linguistic accuracy in the second language in Europe (Räsänen, 1994). Lightbown (1989) states that programs need to devote systematic attention to the way in which language is used to express meaning and teach content. Swain & Carroll (1987) add that adopting systematic plans for language development will help prevent immersion teachers from adopting strategies that are less than optimal for developing second language proficiency. Genesee (1994) recommends that the *implicit* language instruction which occurs in immersion programs be coupled with an *explicit* focus on language within the academic curriculum.

This project was undertaken in the belief that a focus on language throughout the curriculum in a systematic and coordinated manner can improve second language development and proficiency. Elementary school structure is conducive to such explicit focus on language across the curriculum, since the teachers of content are, in many instances, the same who teach language arts. In the middle school, this approach becomes more difficult to implement, since the content teachers are most often not language arts teachers. For this reason, it was thought that coordination between middle school and elementary teachers on language proficiency issues could prove useful to both elementary and middle school teachers. Feedback from the middle school could help elementary teachers by informing them of the language development areas needing attention for students entering middle school. Middle school teachers, on the other hand, could benefit from participation in elementary school plans for language development by acquiring a more in-depth understanding of the instructional experiences students were receiving before entering middle school.

A second rationale for the project was that standardizing the assessment of oral and written language abilities across all three schools participating in the immersion program would engender a more cohesive and stronger program. Moreover, it would provide the county with similar data on language development from all schools in the program, which the district could then use to assess the program's effectiveness and areas of weakness. Bringing teachers from all three schools together to explore these

issues and develop the rubrics would create closer bonds among the three schools. Further it would allow teachers to share with and learn from each other in ways that would not have otherwise occurred.

Prior to the inception of the project in the fall of 1995, formal instruments used in the Arlington Public Schools had provided some useful information about student language ability. However, such information was not complete. Evaluators of the Key School program found that the *Language Assessment Scales* (LAS) test being used had provided useful information on student language ability in the early years, but by third grade most students reached the highest level rating and little further information could be yielded by the test in subsequent years. Further, the *Student Oral Proficiency Rating* (SOPR) scale, which has been used by the Arlington Public Schools, rated oral language ability from the teacher's perspective, but the instrument did not account for or indicate grade-level development and provided information on student's speaking ability alone. Standardized tests, such as the *Iowa Test of Basic Skills* assessed reading and writing, but only in English. Also, as stated above, such tests only gave a limited picture of a student's abilities and did not reflect the learning process the student had undergone nor the breadth of ability the student was capable of (i.e., using language in a variety of ways).

Middle school teachers of English and Spanish language arts had also noticed certain limitations in the language ability of the students who had come out of the elementary school program. These observations mirrored those made of students who have participated in other immersion programs. The Spanish language arts teacher, for instance, had commented that while the immersion students spoke very fluently, their speech was less redundant (morphologically) and less grammatically accurate than the speech of native speakers. The English language arts teacher had observed that some immersion students had more spelling problems and more limited vocabulary in English than students who had not been in immersion.

Researchers from the Center for Applied Linguistics (CAL) had been conducting a case study of Key Elementary's immersion program for the previous two years, under the auspices of the National Center for Research on Cultural Diversity and Second Language Learning (an OERI research center). As part of this study, the researchers had observed the students' reading, writing, listening, and speaking abilities at all grade levels. Through interviews with program teachers, the researchers also discerned a need for more strategies for improving accuracy in the oral and written use of language. Further, in the two years prior to this project, the program had adopted a more explicit approach to teaching language arts in Spanish. Consequently, teachers who were accustomed to only teaching language implicitly (through content) had begun to formally teach language arts. During case study interviews, these teachers expressed a need to learn how to assess language proficiency and teach language in more effective ways than through traditional grammar instruction (Christian et.al., 1996).

In this context, then, the current project attempted to explore the use of alternative language assessments to provide teachers, administrators, parents, and students more accurate information on English and Spanish language development. Given the reduction in the time the project was funded, project directors decided to focus on the development of assessments of the productive skills—speaking and writing—only.

PARTICIPANTS

At the time this project began, in late 1995, Arlington County Public Schools (ACPS) consisted of 19 elementary schools, 6 middle schools, and 4 high schools serving a total of 17,031 children. Approximately 40% of students enrolled in Arlington Public Schools were of diverse ethnic backgrounds, with the largest single group being Hispanics, who made up 29.8% of the total student population. The district had 3,203 limited English proficient (LEP) students, or 19% of the total district elementary school population. Many of these students' second language development needs were served through English as a Second Language (ESOL) and High Intensity Language Training (HILT)

programs at all educational levels. Students at three schools (listed below) were and are currently involved in two-way bilingual education programs. These programs were educating approximately 600 students in two languages—Spanish and English.

The Spanish Partial Immersion Program in Arlington County operates in three elementary schools: Francis Scott Key, Oakridge, and Abingdon. The program feeds into and continues at one middle school and one high school.

Francis Scott Key Elementary School's immersion program exists in a highly diverse context, both culturally and linguistically, within the school and the community in general. The program began in 1986 and grew out of a desire to combine services to limited English proficient students with foreign language instruction for native English speaking students. The first two-way immersion class consisted of one first grade class of 18 students—half native English speakers and half native Spanish speakers—to be educated half the day in English and half the day in Spanish.

In 1991, ACPS received a Title VII Developmental Bilingual Education grant from the Department of Education to strengthen and expand the Key Elementary School program's capacity to serve a greater number of students, fully develop the curriculum units for all grade levels, improve instructional strategies, and provide increased teacher training. Title VII funds also contributed to the program by adding a half-time Project Specialist, adding a supplemental developmental bilingual education program at the kindergarten level, providing a Spanish language arts summer school component, establishing a Parent Advisory Committee, and offering Spanish language and bilingual literacy classes to increase parent involvement.

In the seven years prior to the inception of the current project, the program had grown to include 10 teachers and 319 students in grades K-5. At the start of the project, over 50% of the school and immersion student population were of non-Anglo background, and 53% of those students were of Hispanic origin. While students from the Key School

neighborhood had priority for enrollment in the program, a number of students were bused in from other neighborhoods within the district.

The program is currently involved in integrating some students from the English as a Second Language (ESOL) and High Intensity Language Training (HILT) programs into classes taught in Spanish. The immersion program also serves special education children by including them for instruction in some multi-age classes.

In 1991, the developmental bilingual program in Arlington County expanded to two other schools—Abingdon Elementary and Oakridge Elementary. Since that time, Key Elementary has provided guidance, assistance, and support to the administration and staff at the newer sites. The program was also extended to the middle and high school levels. At the time of this research project, 50 students continued receiving instruction in Spanish and English in grades 6-8 at Williamsburg Middle School. The first class of immersion students is now in 11th grade, continuing their Spanish language education at Washington-Lee High School.

Abingdon Elementary School's immersion program began in 1992 in first grade and has since expanded one grade level per year to fifth grade. At the time of this project, there were two classes at each grade level educating a total of 113 students. Almost 60% of the school's population was non-White, of which 22.2% was of Hispanic origin. Within the immersion program, 40% of the students were designated LEP. The program selected students by parent interest and residency within the school neighborhood. The program attempted to provide a balance of 50% native English speakers and 50% native Spanish speakers in the classrooms.

Oakridge Elementary School's immersion program also began in 1992 in first grade and has since expanded one grade level per year to fifth grade. Currently, there are two classes at each grade level educating a total of 121 students. At the beginning of this project, over 75% of the school's population was non-White, of which 39.9% were of Hispanic origin. Within the program, 30.5% were designated LEP. The program also

selected students by parent interest and residency within the school neighborhood and has attempted to provide a balance of 50% native English speakers and 50% native Spanish speakers in the classrooms.

NEEDS ADDRESSED BY THE PROJECT

The Arlington immersion program's goals have always been high academic achievement and bilingual proficiency for all of its students. The program benefited greatly from the three years of Title VII funding which came to an end in 1995. During that time, the goals for expansion of the program, development of curriculum, and improvement of instruction were largely met. While there had been much progress on improving the instruction of content, there was still a need to focus research on the language development process and how best to assess student achievement in this area. In an instructional environment where students have primarily learned the second language through content, it is often difficult for teachers to separate out language knowledge from content knowledge and ability in order to assess the student's language development and ability. Further, teachers felt that the use of a range of assessment tools would give them greater and more detailed information on the abilities of their students, which would, in turn, enable them to provide more constructive assistance to students in their particular areas of need.

This project, then, was to serve the immersion teachers at all schools by allowing them to craft their own assessment measures to be used in a more uniform manner than was being done. Prior to this project, alternative assessment tools, such as portfolios, were used widely at Key Elementary, but were used to a lesser extent by teachers at Abingdon and Oakridge. Moreover, it was often the case that while student portfolios were kept, teachers were not as aware as they could be of what should be included in a portfolio or of the diverse ways in which portfolios could be utilized to improve assessment and instruction. In the absence of assessment measures, such as scoring rubrics, teachers had no consistent means of evaluating the information they had collected in the portfolios.

Finally, as the immersion programs began to realize the need for more explicit focus on language to achieve higher levels of proficiency and accuracy, teachers needed to be aware of the developmental stages of language acquisition, how to best identify students' strengths and weaknesses in the language, and how to adapt language instruction to meet the students' needs. An alternative assessment package such as the one developed through this project was intended to meet these needs.

NARRATIVE OF PROJECT EVENTS

This section describes the process of the research project. It should be kept in mind that the meetings described in this section were preceded and supplemented by many hours of planning and follow-up between the project coordinator and the project specialist.

Initial Meeting: Establishing the Goals

The first project meeting was held on November 21, 1995 at Key Elementary School. Teachers from all participating schools (except Williamsburg) attended and were provided by Project Co-Director (PD) Ullrich with outlines of the project, its goals and objectives, and a timeline.

After introductions of the research team members, a discussion was held of the existing instruments used to assess language proficiency at the participating schools. It was observed that most of the standardized assessments administered in the schools took place at the end of the school year. This, the teachers noted, did not allow them to receive feedback on student progress that could inform their instruction during the academic year. Further, student test results on oral and written proficiency from one year were not provided to the teachers who would have the students the next year. The teachers, then, were not receiving information on the development of student language proficiencies until the end of the school year, if at all.

As a first step toward designing alternative forms of assessment that could be used throughout the year to yield the kind of information the teachers needed, the teachers were provided with frameworks for language objectives for English and Spanish that had been previously developed by Key Elementary staff and Arlington County. Since these objectives were numerous, developing alternative assessment instruments for all of the objectives would be impractical, it was agreed, since it would be too time-consuming to use them to assess each student's abilities in every area. Therefore, the teachers decided it would be beneficial to choose a subset of these objectives for which they would like to develop instruments. The teachers decided to work with peers from their schools in December and January to choose approximately three objectives for oral proficiency and three objectives for written proficiency for their grade level for which they would like to develop assessment instruments. These objectives would be drawn from the county's Immersion Spanish Language Arts framework for language objectives. Gaps would be filled in by other teachers, or at a later date (possibly over the summer) by the research team. As an incentive for enlisting wider support for the project, Project Co-Director Ullrich offered continuing education credits to teachers at the three participating schools who were not participating officially in the project in exchange for their assistance in project work.

Second Meeting: Exploring Assessment and Setting Objectives

The second meeting was held on January 30, 1996 at Key Elementary. The participating teachers were joined at this meeting by: the coordinator of the county's ESOL/HILT program, the Project Co-directors (PDs), the Project Coordinator (Project Coordinator), the Project Specialist (Project Specialist), and the principal of Key School.

To engage the team in thinking about assessment and its place in instruction, the meeting began with a pair of activities meant to elicit discussion about why and how teachers assess. In the first activity, teachers divided into groups of three and were presented with a scenario in which their principal had asked them to give a five minute presentation at a staff development meeting on "Why We Assess." The groups brainstormed ideas about what they would say. The answers included:

- to diagnose student strengths and weaknesses;
- to provide feedback on student learning;
- to provide a basis for instructional placement within the classroom, (e.g., reading groups);
- to inform and guide instruction;
- to provide practice applying knowledge and skills;
- to provide a basis for student evaluation;
- to obtain data for site-based management;
- to gauge program effectiveness; and
- to teach students how to self-assess.

In general, the teachers expressed the need for assessment to provide information to them about student skills and abilities that could be used to improve instruction.

The next activity involved the groups imagining they were in an evening adult Japanese foreign language class, and discussing how they would like to be assessed. The teachers felt strongly that each student's learning style should be ascertained and considered when assessing the student so that assessment tasks could either be tailored to or incorporate different learning styles. Further, when evaluating the results of an assessment, knowing a student's strengths and weaknesses as a learner could provide supplemental information about why that student performed well or not on certain tasks (e.g., why a visual learner performs poorly on the aural aspects of a task).

Other ways the teachers said they would like to be assessed included:

- authentic/performance assessment;
- through assessments tied to student interests;
- assessed in a non-stressful manner;
- assessments of four basic language skills; and
- assessments in areas of strength and weakness.

In general, the teachers seemed to realize the importance of providing multiple forms of assessment in order to receive greater and more varied information about student skills and abilities. Project Co-Director Ullrich then reviewed some of the trends in assessment that the county was attempting to implement. These included:

- integrating reading, writing, speaking, and listening;
- moving away from paper and pencil and into real-world assessment;
- moving away from single-occasion testing and into more broad-based, portfolio assessment to show change over time;
- moving away from single-attribute testing and into using more holistic, multi-dimensional assessment; and
- moving away from exclusive emphasis on individual assessment and moving toward more collaborative assessments (e.g., assessment of group work).

After covering why and how one should assess, a brief discussion was held about for whom the assessment results were intended, in the particular case of the research team in the Arlington Public Schools. The teachers identified the following audiences:

- teachers (themselves);
- students;
- parents;
- grade-level/department team members;
- school administrators;
- curriculum supervisors;
- business community (sponsors);
- general public; and
- special interest groups.

The teachers emphasized the importance of tailoring assessment to the needs of the parents. Parents, they stated, are not only interested in seeing what progress their child is making as an individual, but also how their child is doing relative to his or her peers. It was stated that the forms of assessment developed by the project should be able to provide this type of information for the parents.

At this point, the teachers presented the grade-level objectives for writing and speaking that they had identified during their individual school meetings. Most teachers found it difficult to limit themselves to only three objectives and instead listed numerous ones. The objectives ranged from task-oriented (e.g. tell a story) to skill-oriented (e.g., ask and answer questions) ones. Some were more grade-level specific than others. For example, "recognize and use past tense verb forms" was specific to third grade, but it was also pointed out that such an objective might be inappropriate when considering language development differences for first-language and second-language learners.

This led to a discussion of whether the project should be addressing such differences or concentrating on objectives that would be broad enough to apply to all students, regardless of language dominance. It was decided that it would be too time-consuming to develop separate English language objectives for native English speakers and non-native English speakers, or separate Spanish language objectives for native Spanish speakers and non-native Spanish speakers. If students were evaluated on broad components of writing and speaking, however, the same assessments could be used for both groups. To demonstrate this, Project Coordinator Montone shared several kinds of rubrics with the teachers on overhead transparencies. These and others were also distributed to all teachers in the form of a packet. After examining some of these rubrics, the teachers decided that tying objectives to broad components of the language skill, such as composing, style, and usage for writing, would work well to assess all students. Additionally, it was observed that such a rubric could also provide a means of comparing native and non-native students' development, if an analytic scoring criteria were used (such as a graded scale). If, however, a holistic scoring criteria were adopted, the teachers might only be able to gain information on individual student proficiencies and progress, and not be able to easily make comparisons between different language-dominant groups. The choice between a more analytic or more holistic grading scale was not made at this time. The discussion was postponed until teachers could decide what kind of information (e.g., individual and comparative, or just individual) they wish to gain from the assessments.

The teachers were then presented with other rubrics that included scales covering all levels of proficiency across grades. The teachers agreed that they would rather develop rubrics that were specific to each grade level. At this point, the ESOL/HILT coordinator pointed out that such rubrics had already been written for English, and

Project Co-Director Ullrich announced that Spanish rubrics had also been created, but not yet disseminated. It was agreed that if adequate rubrics existed, they should be used, rather than re-invent the wheel. However, it was also agreed that the existing rubrics could possibly be separated by grade level—since they existed only as a composite covering all grade levels—and adapted so as to include the agreed upon language skill components of speaking and writing.

With regard to the skill components, the teachers agreed to use those writing components employed by Virginia Passport to Literacy, since all students must be able to pass this test in order to graduate from high school. These components were:

- Composing
- Style
- Sentence Formation
- Usage, and
- Mechanics

For the speaking components, they decided to use those employed by the Student Oral Proficiency Rating (SOPR):

- Comprehension
- Fluency
- Vocabulary
- Pronunciation
- Grammar

The teachers decided to meet with grade-level colleagues on the research team before the next full team meeting in March to determine grade-level objectives for each component for both writing and speaking. Project Specialist Fernández, Project Coordinator Montone, and Project Co-Director Kenyon (when available) would also attend these interim meetings to assist the teachers in writing these specific objectives. It was agreed that the Arlington Public Schools' existing English and Spanish language rubrics would be used to the extent possible to aid in this task.

Finally, dialogue journals were handed out to the teachers with written instructions on how they were to be used to provide the Project Directors, Coordinator, and Specialist

with information and insight on the impact the project was having on the teachers. Other readings and resources were also provided to the teachers, related to the topics of assessment in general, alternative assessment, and portfolio assessment.

Crafting the Rubrics

Between the second and third meetings of the research team, grade-level teachers met to write out objectives and skill descriptors for each of the component areas for writing and speaking that were identified during the second meeting. Project Co-Director Kenyon, Project Coordinator Montone, and Project Specialist Fernández worked with these teachers when possible to craft the new rubrics.

The third meeting of the research team was held on March 19. Teachers presented their draft grade-level rubrics for writing and speaking. At this meeting, most of the discussion was related to issues of what and what not to assess.

There was debate as to whether to assess students' "academic" language only or "social" language as well. It was pointed out that research in immersion programs has shown that students tend to develop good levels of proficiency in "academic" language in their second language, but lower levels of "social" language. Questions were raised regarding how to elicit "social" language for the purposes of evaluation. It was decided that the rubrics would be able to assess students' ability to use language in both social and academic contexts with the understanding that whatever tasks were used to elicit the language to be assessed would have to provide opportunities for the student to demonstrate both social and academic language abilities (or it was to be noted which type of language the task elicited).

Teachers also debated whether to include pronunciation as a component area in the oral rubric. After some discussion it was decided not to include pronunciation for the following reasons:

- 1) the developmental nature of two-way immersion does not necessitate teaching pronunciation, since it develops naturally in young children;
- 2) in such programs, teachers rarely teach pronunciation explicitly, and therefore should not assess pronunciation formally;
- 3) little information is gained by assessing pronunciation; and

- 4) in many two-way immersion programs, the focus is on ability to communicate meaning rather than on form.

It was further agreed that if a teacher wanted to note observations about a student's pronunciation, s/he could do so in the "Comments" area of the "Fluency" component section.

Additionally, in the area of writing, it was decided that the "Style" component should be omitted from the first grade rubric since it was felt students at that grade level should not reasonably be expected to incorporate style into their writing.

Also at this meeting, several scoring options were presented for discussion, such as numerical or letter-based. It was decided that a numerical system would be developed, which could be used for each component area and that would be anchored to the grade-level objective for that component.

Finally, it was noted that the rubrics should be checked across grade levels and component areas for consistency. Interesting questions and comments from teacher journal entries were also presented and discussed.

Fourth Meeting: Revising the Rubrics, Preparing for Piloting

The April meeting focused on further revision of the rubrics, scoring options, and planning for collection of student language samples to use for piloting the rubrics.

After making further revisions to the rubrics, based on continued review for consistency, appropriateness, and practicality, the team revisited the issue of scoring options. A scale of 1-5 was agreed upon, with level 3 representing the end-of-year expectation the teachers had for their students' abilities at each grade level (Appendix 3). Below level 3, a "1" indicates that the student does not approximate the end-of-year expectation, and a "2" indicates that the student approximates the end-of-year expectation. Above level 3, a "4" indicates work exceeding the end-of-year expectation,

while a "5" indicates work *far* exceeding the grade level expectation. It was agreed that scoring with these numbers would not be required, but would be an option for those teachers who might want to have a quick visual reference of their students' proficiency levels.

The team then turned its attention to the criteria for selecting tasks that would be appropriate for eliciting enough oral and written language to enable the rubrics to be utilized fully and effectively. Teachers were given a matrix with each skill component listed for either writing or speaking (Appendix 4). The teachers brainstormed tasks that students could undertake that would yield information for each component area. If the task did not permit a student to demonstrate his/her ability in every component area, the task could not be included. Teachers also had to indicate whether a task elicited social language, academic language, or both, since the rubrics measured student ability in both areas.

Fifth Meeting: Final Preparation for Piloting

For the May meeting, teachers met for a full day to revise and finalize translations of the writing rubrics prior to piloting, to review and revise the instructions for the rubrics (Appendix 5), to develop oral and written elicitation tasks (Appendix 6), and to plan for the summer session and the field-testing that would take place in the fall. During this meeting, the development of a class matrix was proposed (Appendix 7), so teachers could see all their students' scores at a glance.

Summer Work: Elicitation Tasks and Field-Testing the Writing Rubrics

The intensive summer session took place from June 24-28 with a core of four teachers, the Project Specialist, Project Coordinator, and occasional visits by Project Co-Director Kenyon. The nuclear research team met every day from 9 am to 3 pm. During this week, all project products were reviewed once again. English and Spanish language rubrics were reviewed for glaring errors or inconsistencies. The class matrix proposed in the May meeting was developed by the Project Coordinator and presented to the teachers for their comments. The writing rubrics for all grades were field-tested with

samples of student writing taken from the spring county writing assessment. As a result of the field-testing, many revisions were proposed to the writing rubrics. In addition, the Project Specialist proposed collecting oral language samples from among the summer school students that could be used for similar field-testing of the oral rubrics in the fall. Also during this week, grade-level oral and written elicitation tasks were designed. However, it was felt that when field-testing the oral rubrics, it would be helpful to use a uniform instrument for eliciting appropriate and sufficient language from students. To this end, a unique elicitation instrument was created to collect oral language samples in a uniform manner from the summer school students (Appendix 8). Toward the end of the week, the Project Specialist led discussion of other assessment issues, such as the use of language arts portfolios and a new county report card that would incorporate multiple forms of assessment.

Piloting and Field-Testing the Rubrics

At the start of the new school year in September, the project was informed that, despite the Congressional elimination of funding for the project, a five month no-cost extension was granted.

Also at this time, Arlington Public Schools Foreign Language Department decided to have all teachers in the immersion program at all three elementary schools use the writing rubrics for the Fall county-wide assessment of writing in October. A rater training session was conducted by the Project Coordinator and Project Specialist for teachers from all three schools. The next day a research team meeting was held to field-test the oral rubric with the audiotaped language samples the Project Specialist had collected over the summer. Teachers listened to speech samples of students at all grade levels in English and Spanish and used the oral rubrics to score them. As a result, changes in the oral rubrics were suggested and implemented prior to having the teachers pilot them with their students in October.

In the first week of October, the teachers collected writing samples and scored them using the rubrics. Each teacher served as the first rater, scoring her own students'

writing. In November, the research team met for the purpose of conducting a second rating of the students' writing, in order to monitor the teachers' rating reliability. For this activity, a subset of the first rated writings were selected by the Project Coordinator and given to each teacher. No teacher received her own students' writing samples for the second rating. Due to some ratings being incomplete and others not being turned in, this second rating procedure and subsequent analysis was extended throughout the fall.

Due to the many tasks being asked of the teachers in October, the field-testing of the oral rubrics was postponed until November. During the November research team meeting, the oral rubrics were reviewed and more changes were suggested. Later in November, the project teachers used the oral language elicitation instrument developed over the summer to pilot the oral rubrics. Each project teacher pulled individual students out of the class for about ten minutes, elicited the student's speech, and then used the oral rubric to assess the speech sample before assessing the next student. In each case, the Project Specialist attended and served as the second rater. Analysis of the rater reliability was conducted in the winter of 1997.

Classroom Visits and Instructional Impact

During November the Project Coordinator and Project Specialist visited project teacher classrooms at Key and Abingdon to experience the teachers' local settings and instructional contexts. Due to the premature ending of the project, more classroom visits for this purpose as well as to form basis for suggesting ways in which the rubrics could be integrated more into instruction could not be undertaken. Neither could the planned teacher visits to one another's classrooms.

Meetings in January and February gave teachers the opportunity to discuss assessment practices more in depth. During these meetings, teachers shared their experiences using the rubrics and the impact the project has had on their awareness of the students' language abilities and their own (and the district's) assessment and instruction practices. The teachers also continued to review and revise the rubrics, and revised the

classroom matrix and introduction to the rubrics. The Project Specialist also continued to revise the Spanish rubrics and completed the translation of the introduction to the rubrics during this period.

Final Touches and Continuation

A final project meeting was held on March 17 during which the teachers reviewed all project products one more time prior to completion of the final report and widespread dissemination of the products. Project Co-Director Kenyon presented his analysis of the reliability of the writing and speaking ratings undertaken in the fall. (See Appendix 9 for a more detailed analysis.) During this meeting, final revisions were made to the rubrics.

Later in April, the teachers reviewed and commented on a draft of the final report, which was written by the Project Specialist, Project Coordinator, and Project Co-Director Kenyon. Although the formal project was brought to a close, Arlington Public Schools remains dedicated to continuing work on the rubrics. The Spanish language rubrics that were developed by the project have provided the county with its only instruments for measuring student progress in speaking Spanish, and for having grade-level objectives for speaking and writing in Spanish. For this reason, there is strong incentive for the county to continue using these rubrics.

MEETING THE NEEDS

Through a collaborative research approach, teachers, administrators, and researchers worked together to meet the immersion program's need to evaluate extant assessment instruments and determine what information these instruments were providing, what further information about student language development would be useful to obtain, and what other assessment tools and measures could be used to provide such information. During period of the grant, teacher/researchers developed and field-tested an assessment tool that would meet the program's needs and fit into the program's larger portfolio assessment plan.

Over the course of the 17 months of the project, the research team designed scoring rubrics for speaking and writing in Spanish and English that were consistent with the immersion program's and the district's language development objectives. Through the field-test process, the research team observed the efficacy of the language rubrics as measures for assessment, making numerous revisions. The research team compiled data from the use of rubrics to determine their validity and reliability as well as practical aspects related to their use in the classroom.

In addition, regular teacher/researcher reflection on experiences utilizing the rubrics revealed that teachers, as a result of this project, had begun to focus more on the language development of the students. These impressions were recorded in dialogue journals, by which means the participants communicated their questions, concerns, and insights to the project coordinators. Information from the journals were shared and discussed during meetings of the research team throughout the year. These discussions provided the teacher/researchers with valuable opportunities to share insights, successes, and failures with their colleagues so that they could improve on their assessment procedures and suggest new instructional strategies and techniques.

The participation of CAL researchers assisted the program in a number of ways. CAL researchers were very familiar with the Arlington immersion program, having provided it with guidance, assistance, and evaluation since the program's inception in 1986. This familiarity with the program allowed CAL researchers to help guide the teacher/researchers in crafting assessment measures that would meet their program's unique needs. CAL's knowledge and experience in the field of linguistics, language development, and language testing were used to help teachers identify the language needs of the students, develop and field-test, and evaluate the rubrics.

GOALS ACHIEVED

The following are the goals and objectives of the project as laid out in the original project proposal. It should be remembered that these goals were written for a proposed three-year project. The original project was cut back to two-years after review by OBEMLA. As mentioned above, it was funded for only one year. ACPS and CAL were given a five-month no cost extension, which meant the duration of the project was 17 months, less than half the time envisioned when these goals were originally written. Nevertheless, through the hard work the project team met these goals and most of the objectives.

GOAL ONE

Develop sets of language objectives by grade level for a two-way immersion program.

Objective 1.1. Increase the research team's awareness of the spectrum of abilities and skills underlying language proficiency and communicative competence in general.

Objective 1.2. Determine overarching, general objectives for the language goals of the two-way program appropriate to each grade focused on in the study.

Objective 1.3. Examine the implications of such objectives for instructional practice.

Objective 1.4. Develop and adopt general language rubrics as part of an ongoing process informed by insights from the assessment and instructional components of the research effort.

RESULTS FOR GOAL ONE:

GENERAL: The project integrated the grade-level objectives of the ACPS Spanish and English language curriculum frameworks into the assessment rubrics. In doing so,

language objectives have become fully incorporated into assessment.

Objective 1.1. Through discussion at large group meetings, grade-level meetings, and between the Project Specialist and teachers, the teaching members of the research team became more fully aware of issues in language proficiency and communicative competence. In addition, the Project Specialist and Project Coordinator provided many pertinent outside readings to the teachers on the topic. An increased awareness and sensitivity to these issues was present in the teachers' discussion and also in the journals they kept. Teachers have gained a better understanding of assessment by reading and sharing articles, and other research and references that support the use of alternative assessment. Teachers now suggest strategies and changes in instructional practices that include:

- more conversational opportunities to improve fluency;
- more emphasis and opportunities for writing; and
- awareness that teachers need to give students more time and to help them develop oral language skills.

Objective 1.2. This objective was met as the teachers selected grade-level language objectives as the core of the rubrics and then further discussed and refined them together as the rubrics were drafted, revised, piloted, and field-tested. Examining the application of the rubrics to actual oral and written samples helped teachers gain more insight into the appropriate language objectives for each grade. Thus, the process of developing the scoring rubrics encouraged teachers at all grade levels to focus more specifically on the language abilities of their students. With the help of some assessment instruments already being utilized in the district (e.g., for the ESOL/HILT Program, and the state literacy requirements for graduation from high school), the research team identified grade-level objectives for component areas in speaking and writing in both English and Spanish. The language objectives for Spanish constituted the first time the district has had such objectives, filling a gap in the curriculum design.

The project also required teachers to focus on the process of language development over time and across grades. Teachers reviewed and revised the rubrics many times, striving to make them consistent across grades and to demonstrate progressive and realistic advancement in proficiency from one year to the next within every skill component area. As a result, the immersion teachers now have a more complete picture of what their students' abilities are and should be and a means to measure their progress in language skill development on an ongoing basis.

Objective 1.3. The teachers involved in the project reflected on how their instruction might improve as a result of having explicit language objectives. The teachers have expressed their opinion that having clear language objectives will give them uniform guidelines to follow and that there will be more consistency across grade levels. One teacher remarked, "I think my analysis is less subjective because I have guidelines to follow. I also have a clearer set of expectations for each grade level." In the future, as teachers internalize these objectives, they expect them to have a greater impact on the delivery of their instruction. As the same teacher added, "I have not yet internalized the objectives. I think that with more frequent use, I will be more comfortable with them, and I will incorporate them into my assessment."

During the process of developing the rubrics, the research team teachers had opportunities to talk and debate over how to apply rubric-based criteria to actual students' performances, and they recognized that they would have to struggle to find time to do a more in-depth evaluation of their students. Overall, the teachers found that oral and writing rubrics provide them with the following:

- guidelines, thus making the observation of students' learning less subjective;
- a clear set of expectations for each grade level;
- information to report students' progress to parents at conferences; and

- a means to help determine grades for report cards.

Some teachers have used the rubrics during parent/teacher conferences to report on student progress. These parents stated that it was very helpful to have a set of objectives on hand in order to be able to talk about the progress of their child.

Objective 1.4. In electing to use the writing rubrics developed by the project to score state-mandated assessments during the fall and spring of the 1996-97 school year, ACPS has demonstrated that this objective has been reached. While there is no mandated oral testing, we expect that there will be a widespread use of the oral rubrics in the county's immersion programs as well. Thus, as a direct result of this project, the Arlington immersion program now has a more complete assessment process for meeting the language development goals they have for their students. There is now a more uniform and informative assessment of student language abilities across program sites.

GOAL TWO

Develop assessment procedures appropriate to the two-way immersion context and that reflect students' linguistic and communicative competence accurately and comprehensively.

Objective 2.1. Explore and assess ways for teachers to receive a more complete view of student language proficiency and communicative competence and to receive a valid picture of students' progress in language learning and use.

Objective 2.2. Examine the compatibility and congruence between forms of assessment and instructional activities that advance the goals of two-way immersion.

Objective 2.3. Field-test assessment procedures for their applicability, validity, and quality of information they yield.

Objective 2.4. Evaluate the alignment of assessment procedures with both language objectives and instructional activities.

Objective 2.5. Critique assessment procedures and their implications for other aspects of the instructional process.

RESULTS FOR GOAL TWO:

GENERAL: The project clearly met this goal for the productive skills of speaking and writing. Rubrics were developed to assess performances on oral and written tasks, and teachers brainstormed lists of grade-level tasks appropriate for assessment (Appendix 6).

Objective 2.1. Through group meetings, grade-level meetings and individual conversations with project personnel, teachers became more focused on assessing language ability than previously. This was reflected in their discussions, journals, and responses on questionnaires. Teachers talked about how they could get a valid picture of their students' language development. They often mentioned that they had not been very focused on it in the past. They thought of creative ways of eliciting spoken and written language performance, and how to use the rubrics developed in the project and creative ways to assess it.

As teachers reflected on the impact that this project has had on their knowledge and views of assessment methods, they realized the changes in their perspectives of assessment. As a result of this project, teachers in the research team have a wider perception of oral and written assessment, and this has had a positive effect in the delivery of their instruction.

In particular, this study has helped teachers to become more aware of their students' *oral* language development and the importance of providing more oral activities so that students can be more actively engaged in the classroom. One teacher noted, "I never concentrated very heavily on oral language until this study. I find oral assessment to be an important part of first-grade instruction." It has also made teachers more conscientious of alternative assessment methods and has caused them to look at students in a developmental manner; that is, as moving along a language continuum.

Objective 2.2. Learning, teaching, and assessment are inextricably linked, and project teachers felt that assessment procedures should derive from existing classroom practices. As the teachers developed performance tasks, they saw this link and opportunities to measure performance with different tools. Thus, designing performance tasks that were in tune with their instructional strategies was an integral part of the curriculum.

Before this study, assessment tried to reflect what was taught, but now teachers use assessment as a tool to tell them what to teach. For example, one teacher remarked, "I try to incorporate assessment by using the information to see what needs to be reviewed and what has been mastered." Teachers now understand the need of integrating assessment and instruction, as assessment helps gather the information that will guide the teacher in providing the instruction that students need, both collectively and individually. For example, if many students were consistently using the improper placement of adjectives, the teacher might plan lessons on specific placement of adjectives to make their writing more effective. The teacher could then observe who successfully applied this skill in his or her writing. It would therefore become a cycle—first, assess by observing students work, then instruct, and then evaluate students' ability to apply the instruction. In this way, teachers intend to use assessment to plan their teaching, and group students of similar levels for instruction, as they conscientiously work to improve specific areas of writing such as composition, grammar, and so forth.

In this way, teachers felt that the rubrics could be used as both formal and informal assessment tools to evaluate oral and writing language development. As a more formal instrument, teachers plan to use the information gathered from the rubrics to report to parents, as part of the student's portfolio, to substantiate students' needs whether it is the gifted learner or the learning-disabled student, and also to help them determine grades for report cards.

Overall, the teachers' own classroom teaching experiences support the research literature's statements that assessment plays a key role in instructional and learning experiences. Including several opportunities for assessment and evaluation in the cycle assures keeping one's instructional goals and objectives "on track." As one teacher pointed out, "I don't see myself using numbers to evaluate my students, but look at what they are doing and what they aren't, according to our criteria of language development." In balance, this was a most useful research project, as it yielded

precisely the information needed to help teachers chart a realistic course of action for the students in the two-way bilingual immersion programs in Arlington Public Schools.

Objective 2.3. As described in the report narrative, the rubrics were field-tested in all grade levels for both English and Spanish. (A detailed description of the field testing and the empirical results are presented in Appendix 9.)

Due to the reduction in the length of the project and the inability to examine the rubrics used operationally, it is difficult to address their validity in detail. The notion of validity of an assessment is ultimately related to the use that is made of the scores resulting from the assessment. The main goal of the project was to provide classroom teachers with an integrated approach to curriculum and assessment that would provide them with an additional measure to inform and improve their instruction. Such as classroom-based assessment may be considered "low stakes" on the part of the individual student, as it is primarily intended to inform the teacher and improve instruction. The duration of the project meant that an analysis of the implementation of the rubrics and assessments was not possible. However, teachers involved in the project felt that the rubrics and accompanying assessment provided information useful to their instruction. They felt that there was an alignment between what they more informally see in the class and the results of the assessment. Because of this, they particularly liked the possibility of more formally documenting (as in a portfolio) student work and progress in relation to the grade-level goals using the assessments and rubrics developed in the project. There was a strong feeling that this would be beneficial in communicating progress to parents and students. Thus, they felt that the quality of information yielded by the assessments was high.

The reliability of the assessment, one aspect of its technical quality, is also documented in Appendix 9. In short, the reliability appears acceptable for a classroom assessment. Nevertheless, training in the use of the rubrics with new faculty and those not involved with the project will be necessary.

Objective 2.4. Teachers felt that the rubrics should reflect the language objectives stated in the language arts curriculum and that the strategies and activities used in their instruction should facilitate gaining useful information on students mastering the objectives. As such, it was decided to have grade-level objectives be part of the rubrics. In addition, teachers felt the need to design tasks to match the grade-level objectives indicating when performance of a task was above or below the expectation for that grade level.

Objective 2.5. Through this project, teachers came to a better understanding of the developmental nature of language as a process and that instruction should account for this process. One teacher reflected, "I am now paying attention to language as a tool of communication rather than on language knowledge. I do not spend so much time trying to correct all the errors in a piece of writing."

Another result of this project was that some of the teachers realized that students could also assume responsibility for the evaluation of their work. For example, they could analyze pieces of writing using the rubrics, and with the teacher, determine which aspects of their language instruction required reinforcement. Some of the teachers now feel the need to observe students on a day-by-day basis and make written notes to have records of their students' development.

GOAL THREE

Achieve coordination across schools in the development, implementation, and revision of the project recommendations.

Objective 3.1. Fully utilize the combined resources and expertise of the teachers and school district members of the collaborative research team.

Objective 3.2. Ensure that all three participating schools are appropriately represented and involved in the research process and that each school's particular needs are being met.

Objective 3.3. Develop and implement professional development workshops at the research sites with a view to reaching out to a larger number of practitioners.

RESULTS FOR GOAL THREE:

GENERAL: This goal was achieved in large part by the work of the Project Specialist, Evelyn Fernández. The implementation of the results of the project in ACPS will be largely a result of her continuing efforts.

Objective 3.1. The Project Specialist position was funded through this project at 50% FTE. Arlington Public Schools district office showed its support for the improvement of the immersion program in general and this project in particular by financing the other half of Mrs. Fernández's time, so that she might serve as the county's Immersion Coordinator. Mrs. Fernández was given office space in the district office from which she administered her daily activities. This level of district and project support, along with Mrs. Fernández's many years of experience with the immersion program and her unbridled enthusiasm for it, translated into a high level of activity on her part to achieve coordination across all schools throughout each phase of this project. Under her

leadership, the resources and expertise of the teachers and school district members were coordinated and fully utilized.

Objective 3.2. As Project Specialist, Mrs. Fernández maintained almost daily contact with teachers and administrators at each of the three elementary schools, and was the point of contact for the project with the participating middle school teacher. Although each elementary school was not represented equally on the project team—due in part to teacher attrition and partly to already heavy teacher workloads, Mrs. Fernández took pains to encourage teachers from the less represented schools to participate, and personally filled in whatever gaps happened to arise.

Objective 3.3. When the district decided to use the writing rubrics for their countywide assessment in the fall of 1996, the Project Specialist and Project Coordinator provided workshops to train teachers from all three elementary schools in how to use the rubrics for the writing assessment. Unfortunately, because the project was terminated after the first year, more training could not be done. Nevertheless, the district plans to continue training of this nature on its own.

DISSEMINATION AND EXTERNAL IMPACT OF THE PROJECT

In February 1997, the Project Specialist and Project Coordinator attended the annual conference of the National Association for Bilingual Education in Albuquerque, NM. They participated in a pre-conference professional development institute on assessment in two-way immersion education. As part of their presentation, the rubrics for speaking and writing in English and Spanish for all five grade levels were distributed to the over 200 participants. Project Coordinator Montone described the research project and explained the rubrics, while Project Specialist Fernández described how the rubrics would fit into the broader language arts portfolio. The presentation was well-received, and there was great interest in the rubrics.

CAL has and will continue to provide materials (e.g., final language rubrics, Key Elementary's portfolio packet, revised assessment plan) to other two-way immersion programs and interested school districts on request. The Two-Way Immersion project undertaken at CAL for the OERI-funded Center for Research on Education, Diversity, and Excellence (CREDE) is planning to use the writing rubrics to show development over several years in student writing ability at all 11 schools participating in the research project. As a result, teachers from each school will be oriented to the rubrics and trained on how to use them. One two-way immersion program in New England has even requested that Project Coordinator Montone conduct a rater training session for all two-way program staff. (This training was scheduled for May 1997.)

Further dissemination has been carried out through such channels such as the ERIC Clearinghouse on Languages and Linguistics, the National Clearinghouse for Bilingual Education, the Comprehensive Regional Assistance Centers, and conference presentations. Due to the early cut-off in funding, information derived from the project (e.g., guidelines for developing language development alternative assessment tools and rubrics) was not able to be prepared for dissemination prior to the end of the funding period. To the extent that CAL and Arlington Public Schools staff have the time and resources to develop and publish or otherwise disseminate such information, they will do so.

REFERENCES

- Brinton, D., Snow, M., & Wesche, M. (1989). *Content-based second language instruction*. New York: Newbury House.
- Christian, D., Montone, C., Carranza, I., Lindholm, K., & Proctor, P. (1996). *Two-way bilingual education: Students learning through two languages. Final Report*. Santa Cruz, CA & Washington, DC: National Center for Research on Cultural Diversity and Second Language Learning.
- Christian, D., & Whitcher, A. (1995). *Directory of two-bilingual programs in the United States, Revised*. Santa Cruz, CA & Washington, DC: National Center for Research on Cultural Diversity and Second Language Learning.
- Crandall, J. (Ed.). (1987). *ESL through content area instruction*. Englewood Cliffs, NJ: Prentice Hall Regents.
- Genesee, F. (1987). *Learning through two languages: Studies of immersion and bilingual education*. Cambridge, MA: Newbury House.
- Genesee, F. (1994). *Integrating language and content: Lessons from immersion* (Educational Practice Report No. 11). The National Center for Research on Cultural Diversity and Second Language Learning.
- Lightbown, P. M. (1989). Process product research on second language learning in classrooms. In B. Harley, J. P. B. Allen, J. Cummins, & M. Swain (Eds.), *The development of bilingual proficiency* Toronto: Ontario Institute for Studies in Education.
- Lyster, R. (1994). The effect of functional-analytical teaching on aspects of French immersion students' socio-linguistic competence. *Applied Linguistics*, 15(3), 263-287.
- Mohan, B. (1986). *Language and content*. Reading, MA: Addison-Wesley.
- Räsänen, A. (1994). Language learning and bilingual education: Theoretical considerations & practical experiences. In A. Räsänen & D. Marsh (Eds.), *Content instruction through a foreign language*, (pp. 9-31). University of Finland: Continuing Education Centre.
- Swain, M., & Carroll, S. (1987). The immersion observation study. In B. Harley, P. Allen, J. Cummins, & M. Swain (Eds.), *The development of bilingual proficiency* (pp. 190-263). Toronto: Modern Language Centre.