

Section 3 What Data to Collect

Language Proficiency, Academic Achievement, Attitude, and Other Assessments

- Identify Kinds of Data Needed
- Decisions about Measurement Instruments
- Locating and Gathering Data
- Collecting Data on Program Improvement
- Why Develop a Timeline?
- SUGGESTIONS & EXAMPLES: Develop Your Data Collection Plan and Timeline

Identifying what data to collect is an important step in program evaluation.

Now that you have determined what your research questions and your goals and objectives are, you need to figure out how to collect the appropriate data to answer those questions. Of course, what you might discover is that you were a little overly ambitious in your research questions because they require much more data collection than your staff is willing to do. So, this is where you get realistic about identifying what measures you can use or develop and what your timeline will be for collecting data.

You might want to form a committee that will look into some of these issues and report back to the staff at a later meeting where you can discuss your evaluation plan.

Identify the Kinds of Data You Need

It is essential to identify early in the process what kinds of data will be needed to answer each of your evaluation questions. There are three types of data you might want to think about:

- **Demographic Data:** This refers to student background data—see more about this in Section 4 – Setting up the Database. You need to think about how much background data you want. There are some standard items that you probably want: student name, student ID, grade level, ethnicity, language background, maybe SES (e.g., parent education level), language proficiency background (ELL, IFEP, EO) at entry. You may also want to consider date of birth, retention in grade, special education designation, or participation in GATE/gifted program.
- **Student performance data (language proficiency and achievement test data):** You need to consider how much language proficiency and achievement test data you need. Some things to consider:
 - You need to meet the national, state and possibly district requirements for achievement testing—typically testing only in English. Sometimes this includes writing assessments as well.

- You should use some achievement measure in the other language of the program as well. In Spanish, you can use Aprenda3, SABE, or Supera. Beginning in 2007, California will produce a criterion-referenced test in Spanish that assesses the California State content standards for second through fourth grades (STS-Standards Test in Spanish). Other states, like Texas, may also have tests in Spanish. Assessments in languages other than Spanish are a little more difficult to obtain, as you have probably already discovered. (Check with other two-way sites to see what they are using and whether they have developed something you might be able to use or purchase or adapt for your needs. You can find out about other two-way programs from the CAL website at www.cal.org. Be sure you use the appropriate measure if there is one required by your state.
- Most states, like California, have developed an oral language proficiency test in English to assess English oral proficiency and literacy. These tests are developed for ELL students and are not usually appropriate for native English speakers.
- NCLB requires that states administer an English language proficiency test to assess English listening, speaking, reading and writing skills. These tests are developed for ELL students and are not usually appropriate for native English speakers.
- You should collect some data measuring oral language proficiency in the second language of the program (e.g., Spanish, Korean, Cantonese, Russian). Several possible examples are given below.

We suggest that you include at least the following types of student performance data:

- Academic achievement measures in both languages
- Measurement of oral and written proficiency in both languages

- **Attitudinal data:** You may want to use surveys and/or questionnaires to examine attitudes in students, parents, and/or staff at your site. These are also discussed in more detail below.
- **Additional or Post-Program Data:** Dual language programs have long-term goals, and many people will be interested in the successes and experiences of those students who have finished the program, whether a K-5 or K-8 program, and are now in high school or even have finished high school. Some students who are in high school now and have completed the program at an earlier grade level may be continuing the study of Spanish content classes or Advanced Placement courses. Others may have had other experiences that were enriched by their dual language program participation, such as travel abroad. High school graduation and college enrollment rates would be of particular interest among students traditionally underrepresented in high school graduation or college enrollment. For an example of a follow-up study with types of information one could collect, you can check out the following articles by Kathryn Lindholm-Leary. These studies collected attitudinal data, grades, achievement scores, and self-ratings of achievement and language proficiency:

Lindholm-Leary, K. J., & Borsato, G. (2005). Hispanic high schoolers and mathematics: Follow-up of students who had participated in two-way bilingual elementary programs. *Bilingual Research Journal* 29, 641-652.

Lindholm-Leary, K. J. (2003). Dual language achievement, proficiency, and attitudes among current high school graduates of two-way programs. *NABE Journal* 26, 20-25.

While it may seem obvious that evaluation data should be used for something, many times they are only used to write reports to send to some administrator somewhere. Evaluation data can be very helpful in further designing the program to better meet the needs of students. If there are multiple measures used, as suggested above, then these data can provide a rich storehouse of information for making instructional decisions within and across grade levels to better articulate the program. Survey, interview and observational data with teachers can be used to determine how well the program is implemented, what strengths and weaknesses teachers have, and what additional training is necessary. It should now be apparent that whatever data are collected can and should be used to enhance student performance, program development and teacher training.

Decisions about Measurement Instruments

Experts recommend using multiple measures as part of an effective assessment and evaluation plan. There are four important qualities of effective assessment in two-way and developmental bilingual programs (Cloud, Genesee, & Hamayan, 2000):

- Assessments are linked to instructional activities and objectives
- Assessments are authentic
- Assessments optimize student performance
- Assessments are developmentally appropriate

National Center for Research on Evaluation, Standards and Student Testing suggests that a strategy to design and implement assessments should meet three criteria:

- Lead to coherent, sustained learning.
- Support a spiral form of teaching, each enhancing and linking what has come before.
- Direct students to knowledge and skills that can be transferred or applied to new or unforeseen situations.

The most common evaluation instruments are standardized tests, which measure student accomplishment in reading, language arts, and content areas (social studies, science, math). On standardized tests, the test items are developed by specialists, and norms are provided to help interpret test scores. The validity of standardized tests for ELL and other students can be enhanced with appropriate accommodations (see Abedi, 2001¹). These tests provide data that can be used to compare achievement across different student groups (e.g., TWI vs. non-TWI; DB vs. SEI), different sites, even different states.

More recent assessments at the state levels tend to be criterion-referenced tests (see Section 5 for information about norm- and criterion-referenced tests) that assess whether students have met the reading and content area requirements that the state has established. While it is possible to make some comparisons across student groups or school sites within a particular state, these measures cannot be used to compare the achievement of students in TWI or DB

¹ Abedi, J. (2001). *Assessment and accommodations for English language learners: Issues and recommendations* (Policy Brief 4). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing

programs in one state to achievement of TWI/DB students in another state. One reason is that the measures are very different and another reason is that the expectations on which the assessments are based may differ (i.e., it may be easier in some states than in other states to meet grade-level expectations).

Besides achievement tests, a variety of measures are available that provide data on different aspects of student performance, learning, and attitudes in both languages; also parent and teacher understanding of and attitudes toward the TWI or DB program:

- Teacher-made tests
- Rubrics, rating scales that measure student ability or behavior
- Teacher observations
- Portfolios of student work
- Collective group work
- Questionnaires/surveys

You might also want to consider **authentic assessment**. Authentic assessment refers to a “form of assessment in which students are asked to perform real-world tasks that demonstrate meaningful application of essential knowledge and skills”². It includes various forms of assessment used to understand student learning, achievement, motivation, and attitudes:

- Oral interviews
- Story or text retelling
- Writing samples
- Projects/exhibitions
- Experiments/demonstrations
- Teacher observations
- Portfolios

Portfolio assessment, which is a form of authentic assessment, can be defined as the "systematic, longitudinal collection of student work created in response to specific, known instructional objectives and evaluated in relation to the same criteria" (National Capital Language Resource Center, <http://www.nclrc.org/portfolio/modules.html>).

Portfolios are advantageous in two-way and bilingual programs because they can be used to document student progress over time in two languages. Carefully designed portfolios provide additional information to monitor student progress and to diagnose particular student concerns. They also provide teachers with good samples of work and growth to show parents at parent-teacher conferences.

Essential elements of portfolios include:

- Samples of student work (e.g., writing samples, audio or video tapes, math or science problems, reports, experiments)
- Student self-assessment (e.g., student self reflection on content or selection of work for portfolio)
- Clearly stated criteria (criteria and standards for grading student work).

² <http://jonathan.mueller.faculty.noctrl.edu/toolbox/whatisit.htm>

In some school districts, two-way programs have employed language development portfolios that include various measures of students' oral and written language over the course of the elementary grade levels. These portfolios included annual measures of:

- Teacher ratings of students' oral proficiency in each language
- Written language samples in both languages
- Reading logs
- Parent survey on use of library & literacy sources at home
- Student surveys of literacy and attitudes

These portfolios provided an excellent source of information to document students' progress in developing language and literacy skills in the two languages. *Be careful—too much documentation can be overwhelming* (e.g., having teachers rate student writing samples four times a year may be too time consuming, though twice a year is manageable).

PORTFOLIO RESOURCES:

- Books on assessment and assessment design:
<http://www.cal.org/k12nflrc/printresources.htm>
- Portfolio Assessment in the Foreign Language Classroom, NCLRC & CAL:
<http://www.nclrc.org/portfolio/modules.html>
- Center for Advanced Research on Language Acquisition (CARLA):
<http://www.carla.umn.edu/resources/index.html>
- O'Malley, J.M, and Valdez Pierce, L. (1996). *Authentic assessment for English language learners: Practical approaches for teachers*. New York: Addison-Wesley.

Locating and Gathering Data

Having identified the kinds of data, the next step is to determine whether the data already exist and can be readily retrieved, whether they exist in paper form and need to be transferred to an electronic database, or whether new kinds of instruments need to be obtained or developed.

Many kinds of data might be readily obtained from your district assessment office, so you should check with them to see about this (see Section 4—Setting up Your Database for more information).

Oral Language Proficiency

1. Individualized Testing

- Many individualized tests are available to measure students' oral proficiency. Examples include:
- Student Oral Proficiency Assessment (SOPA) and Early Language Listening and Oral Proficiency Assessment (ELLOPA) are language proficiency assessment instruments designed to allow young students to demonstrate their highest level of

- performance in oral fluency, grammar, vocabulary, and listening comprehension.
- Simulated Oral Proficiency Interview (SOPI) is a performance-based, tape-mediated speaking test. SOPIs are currently available in Arabic, Chinese, French, German, Hausa, Hebrew, Indonesian, Japanese, Portuguese, Russian, and Spanish, and are intended for adolescent or adult learners of foreign languages.
- Woodcock Language Proficiency Battery—Revised (English and Spanish versions): assesses oral language skills in learners of any age.
- IDEA Proficiency Test (IPT): English & Spanish in young students
- Language Assessment Scales (LAS): English & Spanish in young students.
- Minnesota Language Proficiency Assessments (MLPA): Spanish, French, German (assesses within range of Intermediate level on ACTFL) secondary-level students

Resources for further information:

- Center for Applied Linguistics (CAL) -- www.cal.org
- Center for Advanced Research on Language Acquisition (CARLA) – <http://www.carla.umn.edu/assessment/MLPA.html>
- Directory of K-12 Foreign Language Assessment Instruments and Resources, Center for Applied Linguistics -- <http://www.cal.org/CALWebDB/FLAssess/>

2. Ratings & Rubrics

Many ratings and rubrics have been developed to assess students' oral proficiency. Most of these can be used in any language, including adaptations to uncommonly taught languages. To increase reliability, teachers or other raters should be trained in their use of these measures:

- Student Oral Language Observation Matrix (SOLOM) -- 5 categories (Listening comprehension, Fluency, Grammar, Vocabulary, Pronunciation). Scale ranges from 1-5 for each category. The SOLOM is available in the Appendix.
- Stanford Foreign Language Oral Language Skills Matrix (FLOSEM) -- based on SOLOM, with same 5 categories. Scale ranges from 1-6 for each category. [Rubrics are more objective & positive than SOLOM.] The FLOSEM instrument and scoring guide is available for download at <http://www.asdk12.org/Depts/Sped/policy/part2/FLOSEM.pdf>
- ACTFL Proficiency Guidelines define 4 levels of proficiency (Novice, Intermediate, Advanced, Superior). These were developed for assessing secondary-level foreign language students.

Oral Proficiency Resources:

- Center for Applied Linguistics (CAL) -- www.cal.org
- Center for Advanced Research on Language Acquisition (CARLA) – www.carla.umn.edu/resources/index.html
- Chicago Public Schools – http://intranet.cps.k12.il.us/Assessments/Ideas_and_Rubrics/ideas_and_rubrics.html
- Fairfax County Public Schools – PALS: Performance Assessment for Language Students (lots of rubrics for different tasks) – <http://www.fcps.k12.va.us/DIS/OHSICS/forlang/PALS/>
- O'Malley, J.M, and Valdez Pierce, L. (1996). Authentic assessment for English language learners: Practical approaches for teachers. Longman, Addison-Wesley

Attitudinal Data

Collecting attitudinal data can be very helpful in providing information about the program.

- Students -- ratings of academic competence, satisfaction with program and school or classroom climate, motivation, cross-cultural attitudes, language use outside of school.
- Teachers/Administrators/Staff -- satisfaction with program (assessment of problems), attitudes toward program and other staff, level of and additional needs for training.
- Parents -- satisfaction with program (assessment of problems), frequency of literacy activities in home, training needs, use of target language in home, participation in cultural activities.
- Some sample questionnaires for students, teachers/staff, and parents are available in the Appendix.

Collecting Data on Program Implementation

In Section 1, we recommended preparing an Evaluation Notebook that would, among other things, describe the program design and selected instructional strategies. In Section 2, we looked at some examples of implementation questions and goals that a program might want to investigate. The evaluation process can help ensure that the program actually operates in the way it was described and in the way program staff agreed upon. Implementation data can be collected in a variety of formats—classroom observation protocols to document the use of immersion or bilingual strategies, notes from teacher interviews and focus groups, surveys of teachers regarding their practices and their perceptions of how effective the implementation is.

Implementation data can play an important part in interpreting your outcomes. If outcomes were not as good as expected, and the program was not consistently implemented, you have clear implications for change and possibly professional development as well. If the program was consistently implemented as planned, and the outcomes were still not as good as expected, you might need to rethink the design of your program. Implementation information is a key element in a comprehensive program evaluation.

A good tool for examining the quality of your implementation is the *Guiding Principles for Dual Language Education*, as mentioned in Section 1.

Why Develop a Timeline?

A timeline is helpful so that you can visualize what data need to be collected at what time periods. You don't want to be collecting data at peak times that other things are going on or when students are a little overwhelmed with the beginning or ending of school. Some time periods you may want to avoid:

- The last couple weeks of school, holidays, just before or after a vacation
- Peak work times for staff: report cards, parent/teacher conferences
- Field trips

After filling in your timeline, you may see that you need to make some adjustments in what data you can collect or when you should collect it.

SUGGESTIONS & EXAMPLES:
Develop Your Data Collection Plan and Timeline

This is a much easier task than you might think, especially if you see an example, so check out the example below.

Data Collection Plan and Timeline – Example

This is the assessment plan for a 90:10 TWI program in California

Grade	Assessment	Assessment Period*
K	CELDT FLOSEM (Eng & Span) Preliteracy assessment	ELLs Upon Enrollment September & June Fall & Spring
1	CELDT FLOSEM (Eng & Span) Writing Prompt in Spanish Core reading assessments	ELLs Fall June Fall & Spring Nov, Feb, May
2	CELDT CST, STS & Aprenda3 FLOSEM (Eng & Span) Writing Prompt in Spanish Core reading assessments	ELLs Fall Spring June Fall & Spring Nov, Feb, May
3	CELDT CST, STS, CAT6 & Aprenda3 FLOSEM (Eng & Span) Writing Prompt (Eng & Span)	ELLs Fall Spring June Fall & Spring
4-6	CELDT CST, STS & Aprenda3 FLOSEM (Eng & Span) Writing Prompt (Eng & Span) Student questionnaires	ELLs Fall Spring June Fall & Spring March

Note. Assessment applies to all students, unless noted otherwise.

- CELDT = California English Language Development Test
- FLOSEM = Stanford Foreign Language Oral Skills Evaluation Matrix
- CST=California Standards Tests
- STS=Standards Test in Spanish
- CAT6=California Achievement Test
- Aprenda3 = Standardized achievement test of reading and math in Spanish

This assessment plan may look ambitious, but most of the assessments are required by the state or district. Only the FLOSEM, writing prompts, and student questionnaires are not required. Also, notice that they scheduled their student questionnaires at a time when students would not be doing other testing – in March.

Reflections and Worksheet on Developing Your Data Collection Plan and Timeline

There are some things you may want to consider as you develop your data collection plan and timeline.

GOOD ADVICE: When you are considering a data collection instrument, ask yourself this question: When I get the information, how will I use it? If you can't think of a way that it will help shed light on your program's effects or give information to help improve your program, maybe you don't want to collect it.

1. Go back to Section 2 and look at your goal for language proficiency.
 - What measurement instruments are you currently using at your school site(s) to measure oral language proficiency in each language, if any?
 - Look at the oral proficiency measures/ratings/rubrics presented previously in *Locating and Gathering Data* (and check out the great resources as well) if you need help in making decisions about measuring oral language proficiency in *both* languages.
2. Go back to Section 2 and look at your goal for academic achievement.
 - What measurement instruments are you required to use by your state/district?
 - If you're not measuring achievement in the non-English language, what ideas do you have for measuring achievement in that language?
 - Are you (interested in) measuring written proficiency in both languages? If so, what measures/ratings/rubrics could you use (see *Locating and Gathering Data* in this section)?
3. Do you have any attitudinal or other goals? What instruments could you use or adapt to measure these?
4. Think about your student population. Are there different subgroups (language background, Fluent English proficient, ethnicity, socio-economic status or parent education) that you should be considering separately in your interpretations of results that you have not disaggregated in the past? If so, identify the groups that should be disaggregated.

Your Data Collection Plan and Timeline

Grade Levels	For Whom? ELL, EP?	Assessment of:	Instruments	Data Collection Time Period
		Oral prof -- English Oral prof -- _____		
		Achievement – Eng Achievement - _____		