ED 227 665   FL 013 511

AUTHOR   Perez, Ray S.; Horst, Donald P.
TITLE   A Handbook for Evaluating ESEA Title VII Bilingual
                Education Programs.
INSTITUTION   InterAmerica Research Associates, Rosslyn, Va.; RMC
                Research Corp., Arlington, Va.
SPONS AGENCY   Office of Bilingual Education and Minority Languages
                Affairs (ED), Washington, DC.
PUB DATE   Mar 82
CONTRACT   300-80-0598
NOTE   490p.
PUB TYPE   Reports - Evaluative/Feasibility (142)

EDRS PRICE   MF02/PC20 Plus Postage.
DESCRIPTORS   *Bilingual Education Programs; Elementary Secondary
                Education; *Evaluation Methods; Guides; *Program
                Evaluation
IDENTIFIERS   Elementary Secondary Education Act Title VII

ABSTRACT
        Practical guidelines and recommended approaches for
designing and conducting evaluations of ESEA Title VII bilingual
education programs are presented for program directors and
evaluators. The three volumes include a guide to evaluation basics, a
designers' manual, and a technical appendix. The first volume
provides an overview of evaluation issues and procedures. The second
volume discusses specific procedures involved in planning and
conducting an evaluation, establishing baseline data, monitoring
program operation, evaluating student outcomes, and preparing
evaluation reports. Forms and worksheets are included. The technical
appendix is a collection of articles on such topics as test
characteristics and theoretical justifications of evaluation
procedures, as well as copies of evaluation worksheets. (RW)

A HANDBOOK

FOR

EVALUATING ESEA TITLE VII
BILINGUAL EDUCATION PROGRAMS

PREPARED BY:

INTERAMERICA RESEARCH ASSOCIATES, INC.
1555 WILSON BOULEVARD, SUITE 600
ROSSLYN, VIRGINIA 22209

MARCH 1982

2

PROJECT STAFF

RAY S. PEREZ, PH.D., PROJECT DIRECTOR
INTERAMERICA RESEARCH ASSOCIATES, INC.

DONALD P. HORST, PH.D., PROJECT DIRECTOR
RMC RESEARCH CORPORATION, INC.

PROJECT CONSULTANTS

ROSS GOLDSMITH
PHILLIP GONZALES
STANLEY RATLIFF
RUDOLFO GARCIA
BEVERLY ANDERSON

TABLE OF CONTENTS

A HANDBOOK FOR EVALUATING ESEA TITLE VII BILINGUAL EDUCATION PROGRAMS

ACKNOWLEDGMENTS

5

PREFACE TO THE HANDBOOK

The 1978 amendments to Title VII of the Elementary and Secondary
Education Act (ESEA) of 1968 mandates the Secretary of Education to
develop and publish models to evaluate Title VII bilingual education
projects with respect to the progress made by project participants in
attaining English language skills (Section 731(d)(2)). Section
731(e)(3) of the Act also mandates the Secretary to develop evaluation
and data-gathering models that consider the linguistic and cultural
differences of bilingual children; the availability and operation of
State bilingual programs; and variables relevant to describing Title
VII projects, such as length of the program, hours of instruction, and
qualifications of the teachers. Section 721(b)(3)(C)(iii) also
requires that each basic grant include a plan for program evaluation.

In response to the mandate, the Department of Education initiated an'
undertaking entitled, "A Project for Developing Program Evaluation and
Data Gathering Models for ESEA Title VII Bilingual Education
Programs." The efforts of this activity produced the Handbook for
Evaluating ESEA Title VII Bilingual Education Programs. The handbook,
designed primarily for program directors and evaluators/of bilingual
programs, is comprised of three volumes: a User's Guide, a Designer's
Manual, and a Technical Appendix. The Handbook meets the requirements
of the Act and provides basic guidelines for conducting evaluations of
Title VII bilingual education programs.

1

How the Handbook was Developed

The three-volume series is intended to provide program directors and evaluators with practical guidelines and recommended approaches for determining what should be included in an evaluation and how to conduct an evaluation. The Handbook may also serve as a reference guide for other persons associated with the bilingual program, such as teachers and parents.

Two major activities were undertaken in developing the Handbook. First, information describing current evaluation practices and data gathering activities being conducted by Title VII bilingual programs was collected from programs identified by State Education Agencies (SEAs) and Local Education Agencies (LEAs). As a result of the information collected, parameters of evaluation issues such as student needs, languages served, program settings and designs, and the costs associated with an evaluation were formulated. The second activity reviewed the literature on evaluation methodology related to bilingual education and determined the potential utility of current evaluation theory and practices, as reported in the literature, to the evaluation of basic classroom bilingual programs funded under Title VII of the ESEA of 1968, as amended. Information collected from both of these activities was then utilized in developing the Handbook.

The Handbook is Organized Into Three Volumes

Volume I, A User's Guide to Evaluation Basics is intended to provide the planners with an overview of evaluation issues and a summary of procedures required to perform an evaluation. The guide presents summary information corresponding to the more detailed evaluation procedures presented in the Designer's Manual (Volume II).

Volume II, the Designer's Manual, is designed for the individual(s) actually conducting the evaluation, and contains guidelines; forms, and worksheets necessary to conduct an effective evaluation. The manual consists of five chapters addressing the following evaluation activities:

o   Planning, managing, and staffing the evaluation;

o   Establishing baseline data required for the evaluation;

o   Monitoring program operation;

o   Evaluating student outcomes; and

o   Preparing the evaluation report.

Volume III, The Technical Appendix is a collection of technical articles including topics of interest such as characteristics of specific tests, explanations of key issues in evaluation, theoretical justifications of evaluation procedures that cannot be found easily in the literature, as well as full-size copies of the various evaluation worksheets.

How to Use the Three Volumes Effectively

To benefit fully from the Handbook, the user is encouraged to read the User's Guide in its entirety. This will provide the user with a comprehensive overview of the entire evaluation process. Volume II will then direct and recommend specific actions, activities, and steps to be used in conducting the evaluation. If followed correctly, Volume II will provide the user with a systematic approach to design and conduct the evaluation. The technical information which covers different evaluation issues or methods presented in Volume III may be used by the user as reference material.

The special needs and goals of bilingual education programs require educators and administrators to continually examine the appropriateness and effectiveness of the program. This challenge can be aided through the careful planning and conducting of an evaluation designed to meet the requirements of the funding agency, to enhance the program's management and operations, and to provide useful information for the program administrators to use in improving the program.

VOLUME I

A

USER'S GUIDE TO EVALUATION BASICS

FOR

EVALUATING ESEA TITLE VII
BILINGUAL EDUCATION PROGRAMS

Prepared By:

INTERAMERICA RESEARCH ASSOCIATES, INC.

1555 Wilson Boulevard

Rosslyn, Virginia   22209

MAY 1982

TABLE OF CONTENTS

Volume I

OVERVIEW

This document represents the first of a three-volume series
constituting the Handbook for Evaluating ESEA Title VII Bilingual
Education Programs. The Handbook provides practical guidelines and
recommended approaches for bilingual education program directors and
evaluators to use in evaluating bilingual programs.

In the development of the Handbook, it was readily recognized that a
single document could not be equally suitable to all bilingual
education programs. Obviously, bilingual education programs cover a
range of languages and grade levels in a variety of settings. Some
programs have large evaluation budgets and access to teams of trained
and experienced evaluators, while others have limited budgets and
limited human resources. Additionally, the Handbook is intended to
serve different persons with different needs. Therefore, this
document, Volume I, A User's Guide to Evaluation Basics, is designed
for persons associated with the program, but not necessarily directly
involved in conducting the evaluation. The users of the guide could
be program directors, as well as other persons associated with the
bilingual program, such as teachers, parents or district
administrators.

The User's Guide to Evaluation Basics provides an overview of
evaluation issues and summarizes the procedures required to conduct an
effective evaluation. The guide provides a summary description of the
five components of a bilingual education program evaluation. These

include: planning, managing, and staffing the evaluation; establishing baseline data required for evaluation; monitoring program operations; evaluating student outcomes; and analyzing and reporting evaluation results.

Volume II, entitled The Designer's Manual for Conducting an Evaluation is designed to be used by the persons actually conducting the evaluation.   The manual describes how to implement each of the five components of the evaluation.   The Designer's Manual contains guidelines, procedures, and worksheets to assist the program director and/or program evaluator to complete the specific tasks associated with the overall program evaluation.

Volume III, entitled The Technical Appendix, contains a collection of reference materials covering different issues and topics related to evaluation practices.   These are intended to assist program directors and program evaluators in building upon or expanding the evaluation activities identified and discussed in Volumes I and II. Reproductible copies of all worksheets presented in Volume III are contained in this volume.

CONCEPTUAL FRAMEWORK FOR THE EVALUATION

Bilingual education programs represent a unique instructional approach using two languages to meet their educational goals by generally providing instruction in academic subjects using the student's first (home) language (L1) while developing the English language skills of the students. The students served by bilingual programs also reflect a wide diversity in culture, socio-economic status, and educational experiences. These aspects distinguish bilingual education programs from all other instructional approaches.

The primary goal of bilingual education programs is the development of English language skills of the students as well as the development of their home language. Teachers recruited to teach in these programs, therefore, need to have language skills in the two languages being used for instruction. Curriculum materials in the first language are also needed.

Other goals of bilingual programs often include the development of the student's self-concept by emphasizing the home culture and the improvement of his or her performance in other academic projects. In order to accomplish these goals, knowledge of the students' culture by the classroom teacher and culturally relevant curriculum materials are a necessity in bilingual programs.

1

Due to these factors, the evaluation of bilingual education programs must be performed with considerable caution. The selection of an evaluation approach must take into consideration the variety of educational services, the curriculum materials used in the classroom, the number of hours of instruction provided in English and in the first language, the language skills of the classroom teacher, as well as the educational experience and language skills of the students.

Because of the complexity of this educational context, experimental or quasi-experimental evaluation designs are often not appropriate to evaluate bilingual programs. Experimental designs usually require random selection of students. However, random selection is not realistic in a bilingual education context, because it would require that students who are eligible to receive bilingual education instruction to be placed in alternative programs for control purposes. Similarly, the unique and differing characteristics of the students and the difference in the instructional services they receive make it very difficult to find comparable comparison groups necessary for quasi-experimental designs. The consensus of the literature addressing the evaluation of bilingual programs also indicates that the use of standardized tests to evaluate bilingual student progress is of dubious value. Despite these limitations, some formal measurement of student academic achievement must be undertaken in bilingual education programs.

The Recommended Evaluation Model

The evaluation model presented in this Handbook, therefore, is solely designed to provide descriptive information about the operation of the bilingual program and on the academic performance of the students enrolled in the program. The information gathered through this process can be used to evaluate student progress and to some degree provide a barometer of program effectiveness. The model is based on the premise that an evaluation of a bilingual program should:

o   Provide descriptive information about the operations of the bilingual program; and

o   Provide information describing student performance (even if hindered from making inferences about program impact).

Therefore, the model requires the collection of student outcome data to determine if the students are making progress in their learning. It also requires the collection of information on "how" the program is operating.

The model is also practical and realistic in relation to the financial and human resources available to conduct evaluations of bilingual programs. Aside from the expertise and time of the immediate personnel of most programs, the majority of bilingual programs have limited funds (generally between $2,000 to $5,000 per year) to secure private consultants to perform or assist with the evaluation. Therefore, the model takes into consideration the amount of time and effort that can reasonably be expected to be given to the evaluation effort.

3

The recommended evaluation model consists of two components. The first component focuses on program operations (e.g., program goals, time spent on instruction, etc.) using a discrepancy evaluation design. The design places heavy reliance on descriptive data about the program, therefore requiring as an initial step, the establishment of comprehensive baseline data on the program, the students, and the community.

The actual evaluation and data collection activities needed for the evaluation of program operations are performed primarily through search and review of program documents such as the grant proposals, previous evaluation reports, student files, and related material, as well as personnel interviews and the monitoring of classroom instruction. Personnel interviews to gather information on how the program is being operated are conducted with the program director, teachers, district administrators, and parents. Monitoring of classroom instruction is performed through observation to determine if the instruction is being carried out as planned and in accordance with the original program design.

The discrepancy evaluation attempts to identify and document differences between the initial plans of the program and the actual manner in which the program is operating. Information about discrepancies between the planned and actual program activities, as identified by the discrepancy evaluation, may be used to make decisions on how to continue operation of the program and what changes might be required.

4

1

The second component of the model requires the assessment of student outcomes. The student outcomes to be evaluated are:

o   English language skills;

o   First language skills;

o   Academic achievement; and

o   Affective areas of student performance.

Because of the difficulty in conducting program impact evaluations, the recommended approach to evaluate student outcomes is simply to evaluate student performance. This approach is referred to in this Handbook as the basic evaluation on the basic evaluation design. This basic evaluation design, therefore only answers <u>the relative performance question</u>, "to what extent are the bilingual students achieving?"

The basic design has minimal requirements. These are:

o   Testing only the students enrolled in the bilingual program;

o   using adequate norm-referenced tests (NRTs) capable of measuring English language skills, first (L1) language skills, if applicable, and academic subjects (e.g., math, science, etc.); and

o   measuring performance for only one academic year.

Applying these minimal design requirements to the first student outcome component, English language performance, is all that is reqired to meet the Federal evaluation requirements. However, most bilingual programs should at least evaluate performance in two other

5

outcome areas, first (L1) language and academic subjects.
Additionally, although the basic design does not require a multi-year
evaluation design, the Handbook does recommend that bilingual programs
attempt to collect multi-year performance data. At a minimum,
programs should strive to collect data over the duration of their
grant period. It is conceivable that data showing progress over the
life of the program, can be used to argue that the bilingual program
was responsible for the outcome.

Data resulting from the analysis of student outcomes can be used as an
indicator of overall student performance. The data from this
component of the evaluation, in conjunction with the discrepancy data
can be used to determine what program changes, if any, may be required
to improve student performance.

For example, the discrepancy evaluation of program operations may
reveal a significant operational change from the original design of
the instructional program. This change could have had considerable
impact on the instructional program, to the extent that student
performance may have been affected. Knowing this, the evaluator will
be able to analyze and interpret the outcome data affected by this
change and make recommendations for changes in the program.

In summary, the purpose of the recommended evaluation model is to
describe student performance and program operations. It can not be
used as a measure of program impact. The recommended model meets all
the requirements established in the Title VII rules and regulations.

2

The regulations require that each grant have a plan to evaluate the progress and achievements of the bilingual program. The plan must include:

- o provisions for measuring the accomplishments of the instructional objectives of the program;

- o provisions for measuring the progress of the students in improving their English language skills; and

- o a procedure for using the information gained from the evaluation to improve the operation of the program.

The recommended evaluation model accomplishes this by:

- o performing an evaluation of program operations using a discrepancy evaluation approach;

- o conducting an assessment of student performance in developing English language skills, as well as first language skills and performance in academic subjects; and

- o conducting an analysis function to determine what changes may be required to improve the overall operations of the bilingual program.

The Handbook recommends that bilingual programs should not attempt to determine program impact. However, some basic guidelines for extending the evaluation to determine impact are presented as optional activities to the basic evaluation design. Extending the evaluation beyond the basic design, however may require more resources than those normally possessed by Title VII bilingual education programs. The Handbook also does not address entry and exit procedure issues. The procedures are, however, very much intertwined with evaluation of bilingual programs and should be considered when planning the evaluation.

7

CHAPTER I

PLANNING THE EVALUATION

Planning is the single most important task in conducting an evaluation. Although this point seems obvious, research indicates that many evaluations of bilingual programs, as well as evaluations of other educational programs, are not planned properly. Many evaluations occur towards the end of the program year as a last-minute thought, simply to produce a report to satisfy some external requirement, usually imposed by the funding source. As a result, they are often performed haphazardly and produce poor results.

Evaluations performed in this manner are of little use to either the program itself or the funding agency. These evaluations usually fail to address issues that program and school administrators may have about the program because the evaluation design failed to incorporate their concerns during the planning process. Likewise, these evaluations are not helpful to the funding agency since, at best, they were planned too late in the program year to capture useful information and, at worst, merely represent perfunctory efforts to fulfill a reporting requirement.

The evaluation process, to achieve its own objectives, must be approached in a serious manner and receive as much priority as other elements of the educational program. Program administrators must

realize that the evaluation process is a positive activity designed to provide information on which to base decisions for program improvement.

The planning process carefully balances the reporting requirements of the funding agency, the information needs of decisionmakers and program administrators, and the scarce resources available to conduct the evaluation. It is unlikely that any given bilingual program will have the resources needed to address all the information needs of its different audiences. Therefore, all parties concerned must realize that compromises will have to be made; otherwise, resources will be scattered, producing little useful information.

A properly conducted evaluation requires more than simply evaluating a specific component of a bilingual program (e.g., student performance). An effective evaluation plan identifies all the questions about the program that the evaluation should answer. The evaluation planning process, therefore, involves a series of carefully executed steps which identify the evaluation audience and their specific information needs, set priorities, determine which program components to evaluate, allocate scarce evaluation resources, and set timelines for the evaluation process.

Next to proper planning, effective management of the evaluation process is a must. One person must assume the responsibility and have the authority to direct and manage all facets of the evaluation. All persons involved in the evaluation process must be made aware of the

authority and be given instructions and directions on how to interact
with that person.   A clear chain of command must be delineated.   In
most Title VII programs, the program director retains and assumes that
responsibility.   For purposes of presentation, this Handbook assumes
that the program director is the person responsible for ensuring that
the evaluation is planned and conducted.

GUIDELINES FOR PLANNING THE EVALUATION

1.   Select an Evaluator and Assign Responsibilities

Proper planning and effective management of the evaluation dictate
that the person responsible for designing and conducting the more
technical aspects of the evaluation be identified as early as possible
in order to become involved in the early decisionmaking of the
evaluation planning process.   In the case of most Title VII programs,
this person is usually an independent consultant from outside the
school system.   Ideally, the evaluator should be involved in the
original design of the bilingual program itself.   In the case of Title
VII programs, this should occur during the proposal writing stage.
This enables the evaluator to begin working with the program director
in planning the evaluation before the academic period to be covered by
the evaluation commences.   The plan for conducting the evaluation, if
at all possible, should be completed prior to the first day of school
of the academic year being evaluated.

A major responsibility of the program director is to survey the available human resources in the district and, assuming he or she has the authority, decide whether to use an evaluator from within the school system or employ an independent evaluator. The possibility of contracting for the services of an evaluation specialist from a university or a private consulting firm must be weighed against the potentially lower cost to the program if the evaluation can be conducted by district personnel. The program director must decide on a course of action.

Assuming that an independent consultant or a consulting firm is contracted to perform or provide assistance in conducting the evaluation, the program director should assign clearly defined responsibilities and specific assignments to the evaluator, the program personnel assisting with the evaluation and himself.

The evaluator's function and responsibilities are usually determined by the amount of technical assistance needed by the program director in carrying out the evaluation. The evaluator's role is therefore generally narrower in scope, focusing more on technical matters such as test selection, designing data collection procedures and instruments, conducting data analyses, and reporting the evaluation results.

Listed below are some guidelines to distinguish the role of the program director and evaluator in the conduct of the evaluation. These guidelines take into consideration the fact that the majority of

the evaluation activities will actually be conducted by the program

director and program personnel.

The program director should:

o   Define program goals and objectives;

o   Describe the intended program;

o   Describe student characteristics;

o   Identify target audiences for the evaluation;

o   Determine the major areas to be covered by the
    evaluation;

o   Identify possible evaluators, and in some cases,
    select the evaluator(s) or at least recommend the
    evaluator(s);

o   Serve as a liaison with the evaluator (or appoint
    a staff member to serve as liaison;

o   Review the evaluation design prepared by the
    evaluator to make sure it meets the evaluation
    needs;

o   Arrange interviews or write cover letters to
    questionnaires to ensure timely response and
    cooperation;

o   Monitor classroom operations and observation
    activities;

o   Assign specific evaluation activities to program
    personnel;

o   Identify trained personnel and/or suggest specific
    persons who should be involved in data analysis
    and interpretation;

o   Review data and react to interpretations and
    recommendations before they are included in the
    report; and

o   Make presentations on the results of the
    evaluation.

1-5

26

The evaluator should:

o   Design the evaluation based on the information
     needs identified by the program director;

o   Select and/or review instruments to be used in the
     evaluation;

o   Monitor testing; and

o   Analyze the data and report findings.

A clear delineation of responsibilities and responsible management

will ensure that all evaluation activities are performed effectively

and on schedule.

2.   Determine the Audiences and What to Evaluate

An evaluation is designed for a particular reason and for a particular

audience. Thus, the first step is to determine who needs information

from the evaluation, what type of information is needed, and for what

purposes. In addition to the program administrators and other

personnel associated with the program, the typical users of evaluation

information usually include:

o   The funding agency;

o   District administrators;

o   School board; and

o   Parents and community groups.

Each audience has different interests and needs. Therefore, the

evaluation design must address the different needs of each audience

and provide the information desired, while remaining within the

budgetary constraints of the program. The _Designer's Manual_ (Volume II) provides suggestions and worksheets to use to accomplish this.

Evaluations of ESEA Title VII funded programs, however, must pay particular attention to the rules and regulations pertaining to these programs. Embodied in these rules and regulations are a number of provisions that should be viewed as minimal evaluation criteria. Therefore the evaluation requirements for basic and demonstration projects, as described in section 123a.22 of the April 4, 1980 Federal Register (Vol. 45, No. 67) must be considered in planning the evaluation.

These regulations require that any program funded under Title VII of the Elementary and Secondary Education Act (ESEA) of 1968, as amended must have a plan to evaluate the progress and achievement of the bilingual program. The plan must include:

   o   provisions for measuring the accomplishment of the
       instructional objectives of the program;

   o   provisions for measuring the students' progress in
       improving their English language skills; and

   o   a procedure for using the information gained from
       the evaluation to improve the operation of the
       program.

3. Set Priorities and Establish Timelines

The establishment of evaluation priorities is a must for all bilingual programs. Most bilingual programs allocate $3-5,000 of their yearly budgets for the purchase of outside consulting assistance to perform

the evaluation. This amount of money, together with the level of effort that can be devoted to this one task by the program director and the rest of the program personnel, constitute the available resources to conduct the evaluation. More than likely, the evaluation needs identified by the exercise described above will far exceed what can be accomplished by these resources. Consequently, priorities for the evaluation may have to be established.

The program director must analyze the evaluation needs identified through the planning process, assess the resources available to conduct the evaluation, and ask the following questions:

o   How much can I evaluate?

o   How much do I need to evaluate?

o   How much evaluation assistance can I afford?

Additional questions, such as the ones below, will also help to determine priorities:

o   Is information on the program's capacity to meet Title VII regulations already available? If information is available, this information should be incorporated in the evaluation.

o   What are the priority areas (e.g. parent involvement) of the program? The evaluation effort should give these areas priority.

o   How are the program resources divided among program components? Areas receiving a large proportion of program resources should be candidates for evaluation emphasis.

o   If there are insufficient resources to adequately evaluate all components, are there areas that should not be evaluated or should the scale of the evaluation be reduced in some or all areas? This

decision should be made after considering which
areas are already fairly well understood, which
areas are a low program priority, and whether the
evaluation resources are so limited that it would
be best not to evaluate them at all rather than to
conduct a general assessment of all areas.

o · Which components must be evaluated each year?

After the evaluation priorities have been determined, the program

director should establish timelines for completing the different

components, as well as the total evaluation of the program. The

program director should understand that certain elements of the

evaluation must be performed at very specific times during the

academic year and cannot be delayed or postponed. However, the

program director has to consider the other responsibilities of the

persons assisting with the evaluation. Responsiblities and

assignments may have to be modified as a result of the established

timelines.

4. Determine Level of Effort, Budget and Allocate Resources

One of the most difficult tasks in managing the overall evaluation is

deciding how best to utilize the limited resources available, and yet

meet all the evaluation needs. The assignment of responsibilities and

activities to those contributing to the evaluation process is often

difficult. Because most of the evaluation activities pertaining to

Title VII programs are usually performed by the program personnel,

coordinating time schedules to perform the evaluation with the other

program responsibilities of the personnel can be difficult, especially

if human and financial resources are limited. Nevertheless, the

timely execution of the evaluation is essential. There are activities within the evaluation process that can be rescheduled; however, others must be performed as planned in order to produce a reliable product. The effective program director must exercise initiative and resourcefulness to ensure that this is accomplished.

Determining how much of the evaluation should be conducted by program personnel, which activities should be performed by an independent contractor, and how much the total evaluation should cost is often difficult for many program directors. Districts with limited contract evaluation funds should use most of their contract funds to employ a trained and experienced evaluator to assist them in evaluating the student outcomes component of the evaluation. Other evaluation tasks, such as describing and monitoring program operation, can be performed by the program director with assistance from the program personnel as a normal part of program management. However, the evaluator should be consulted when performing these tasks. If project or district personnel are going to be employed to perform the evaluation, the program director must make specific assignments and ensure that the evaluation activities are performed on schedule.

A major step in planning and managing the evaluation, therefore, is determining the level of effort that will be required by each activity of the evaluation (e.g., evaluating student outcomes) and allocating adequate financial and human resources to the individual tasks to be performed. Evaluation resources, financial and human, will vary widely from district to district. Additionally, the level of effort

for an evaluation is affected by a number of factors, such as:

o   Size of the program;

o   What aspects of the program are evaluated;

o   The number of non-English languages represented in
     the population being served by the program; and

o   The number of evaluation questions addressed.

The Designer's Manual provides guidelines and worksheets to assist the
program director to allocate resources and budget the evaluation. The
guidelines suggest three different estimated levels of effort that can
be applied in evaluating each program component and the different
tasks within each component. These estimates are based on discussion
with persons who have conducted similar evaluation activities.

5. Plan the Data Analysis Function

The program director and evaluator should plan the specific data
analysis activities that will be required by the evaluation. The type
of analysis and techniques to be used will depend largely on the types
of data collected. Data from the first component of the evaluation
will consist primarily of narrative descriptions of program
operations, as well as responses from the interviews collected. Data
from the second component of the evaluation will be primarily in the
form of test scores.

The analysis and interpreting of program operations data is a straightforward comparison activity. The evaluator simply examines and compares the information collected on the actual operation of the program to the baseline information describing how the program was meant to operate. For example, if the goal of the program was to provide instruction in all academic subjects using the native language of the students, the analysis function, using the second set of information, simply ascertains if this indeed occurred. If the goal was met, the analysis activity documents this. If the instruction did not occur, the analysis activity also documents this and should attempt to ascertain what caused the change in the program design. Both types of findings are recorded and reported in the overall evaluation report. This type of comparison analysis is all that is needed by this component of the evaluation.

Analyzing student outcome data is also a straightforward activity, but should be performed by a trained evaluator. The analysis activities required may be performed by simply following prescribed procedures within the manuals that accompany most commercial tests. Programs using the basic evaluation design will only be required to perform basic analysis, such as frequency distributions, computation of mean scores and standard deviations. The analysis activity will also require the evaluator to estimate the degree of possible error in the results.

Interpreting the findings or attempting to find an association between the findings of the program operations component and the student

outcomes component should be performed very cautiously. The two sets of information are not meant to be "scientifically merged" in accordance with sound methodological evaluation practices. However, an alert and perceptive evaluator may be able to develop some "intelligent perceptions" about the program based on the two sets of information. For example, knowing that history was taught using the home language in the fourth grade, but not in the fifth, the evaluator may want to closely examine the student outcome data for these two grades. If the data from the fourth grade students shows significant higher achievement than that of the fifth graders, the evaluator can highlight this fact and then present a "professional opinion" suggesting that the instruction in the native language fostered this difference in achievement.

The important consideration during the planning stage is to determine how the analysis function will be conducted. Data analysis will most probably be performed by the evaluator. The time schedule for the evaluation should allow ample time to conduct the analyses.

6. Plan the Data Interpretation and Development of Recommendations

Data interpretation in bilingual program evaluation is often not a strictly empirical task. To repeat the basic premise of this Handbook, it is probably impossible to show, by employing conventional social science research methods, that children in the bilingual program did better in the program than they would have without it. Therefore, interpreting the data obtained by evaluation efforts is not

1-13

a mechanical exercise of reciting significant alphas. Rather than concluding that the bilingual program "works" better than some alternate treatment, the interpretive exercise is more likely to be in the nature of a policy question. Does the bilingual program "work" well enough? Are decisionmakers and constituents satisfied with the program and the student's progress? Recognizing the policy implication function of data interpretation, an interpretive panel may be a better alternative to perform this function. Chapter IV provides more detailed guidelines and procedures for performing the interpretation function.

Two basic approaches are therefore suggested for data interpretation and formulating recommendations for program modification. The first approach is for the evaluator to analyze, study, and interpret the results. Using informal means, the evaluator then checks the interpretations and recommendations with program staff and others as he/she deems appropriate. The second approach is to convene a panel of people with various perspectives on the program and have them interpret the results. The panel may consist of individuals that are representative of the various audiences. This decision can be made immediately before the analysis activity begins.

7. Plan the Reporting of the Evaluation

Preparation of the final evaluation report is an important activity of the evaluation. The evaluation report is the final and most visible product of the evaluation. Steps should be taken to ensure that the

report addresses the purposes and specific questions of the
decisionmakers for whom the evaluation was planned. 'In addition, the
evaluation results should be reported in a timely manner, taking care
to ensure that the technical aspects of the evaluation effort are
clearly presented. Together, these steps increase the usefulness of
the evaluation results.

Several standard elements should be included in the report. These
include:

o   Statement of purpose;

o   Program overview and background;

o   The goals and objectives of the bilingual
        program;

o   Description of the program and students;

o   Discussion of the methodology used; including
        design, sampling strategy, instrumentation, and
        data analysis procedures; and

o   Presentation of the findings, conclusions, and
        recommendations for program change.

The report should be concise and should include easily interpreted
tables, graphs, and other figures limiting the amount of narrative
material presented. Important issues should be identified and
highlighted in the report if the results of the evaluation effort are
to be maximized. Techniques such as boxing in recommendations or
using a different type face are useful to highlight the most important
points of the report. Examples of actual data collection instruments
should be included in an appendix. Chapter V provides more detailed
guidelines for developing the report.

CHAPTER II

ESTABLISHING BASELINE DATA REQUIRED FOR THE EVALUATION

The evaluation model for evaluating Title VII bilingual education programs presented in this Handbook has two components. The first component evaluates program operations (e.g. program administration, staff development, parental involvement, etc.) using a discrepancy evaluation design. The second evaluates student outcomes. Results of these two evaluation activities taken together constitute the basis for determining how the program operated and provides a description of student performance.

In order to conduct the discrepancy evaluation of program operations, information on how the program was originally designed and intended to operate must be collected and documented. This information serves as the baseline data, which are compared to the data resulting from the actual evaluation of program operations as described in Chapter III.

The information obtained from the evaluation of program operations is taken into account in developing and conducting the evaluation of student performance. Therefore, a very early and important step in conducting an evaluation of a bilingual program is the establishment of baseline information about the total program.

This description includes identifying who the program is meant to serve, what are the exact services of the program, how these services are to be provided, and what outcomes are expected from the services. Without this description, it is impossible to determine (a) whether the bilingual program meets the original intent, and (b) whether any marked achievements can reasonably be attributed to the program.

Comparing the original program design, as described by the baseline data, to its actual operation, as determined by the evaluation of program operations, will indicate areas of the program that have either not been implemented or have changed from the time that the program was originally designed. Discrepancies identified as a result of this comparison are a powerful management tool for the program director and a programmatically useful part of the whole evaluation process. This comparison can also help to determine whether the goals of the program are reasonable, and provide information about the relationship between program activities and program outcomes.

In order to accomplish this, the persons conducting the various evaluation activities must first develop proper documentation of the program context, the target students, the program goals, and the instructional program. This is not a difficult task. The information to be collected should clearly describe how the program is designed to meet its goals, as well as the total environment in which the program operates. Once this documentation is accomplished, the program director, with assistance from the evaluator, will be able to use the information to design the evaluation and to analyze and interpret the

evaluation results. The documentation does not need to be elaborate, simply informative. Most importantly, the information collected should be complete, detailed, and easy to understand. The Designer's Manual provides more detailed listings of the different information that needs to be collected. These listings are also found in the Technical Appendix.

Baseline Data Needed for the Evaluation

1. Describe the Context of the Program

Develop a simple, but accurate description of the school district and neighborhood. Data from previous evaluation reports can be easily updated, thus avoiding surveys or other time-consuming efforts. The type of information that should be covered in the description includes:

- o Community characteristics

  - Languages spoken
  - Ethnicity
  - Social economic status (SES) levels
  - Mobility and length of residence
  - Size

- o Local Education Agency (LEA) description

  - Size
  - Financial status
  - Facilities available for the bilingual program
  - General goals
  - Philosophy towards language and cultural diversity

o   School Description

　　　　　　-   number of bilingual students by language
　　　　　　　　group
　　　　　　-   number in the bilingual program
　　　　　　-   how students are assigned to classroooms
　　　　　　-   bilinguality mix in classrooms
　　　　　　-   parent involvement in school affairs

The information collected on the context of the program should be compiled immediately after the data-gathering phase. While technical analysis of the information is not required, the program director and evaluator should review the data in order to plan the evaluation of program operations and make preliminary decisions on how the data will be used during analysis to determine program outcomes. The information should be written in narrative form for inclusion in the final report. The topics and subheadings provided above may serve as an outline for reporting this information.

2. Describe the Students

Baseline information about the language proficiency and dominance, cultural background, and overall academic achievement of the students enrolled in the bilingual program is essential for designing and conducting the evaluation. The baseline data must include information on the skill level of the students in both English and their home language, as well as their level of performance in the subject areas being taught. The description should also include information on the student's learning background and school environment. At a minimum, the baseline data should include information on the following areas:

o Definition of project student

o Student selection criteria & method

   - Tests & cut-off scores used

   - Role of teacher judgment

   - Role of parent wishes

   - Method of combining criteria

o Exit criteria & follow-up

o Student turnover

o Student characteristics at beginning of year

   - Language proficiency

   - Achievement level

   - Biographic data

This information is essential for grouping students according to both
current skills and past experience during data analysis activities and
plays a major role in determining student performance.

3. Describe the Program Goals

Developing a clear and complete description of the goals of the
program is an esssential part of establishing baseline data. Goal
setting, although important, is often overlooked or ignored during the
program planning stage.

Therefore, many programs operate year-to-year with little or no set direction. Programs that fail to establish clear and measurable goals cannot expect to be able to measure program outcomes.

Programs should distinguish between short-term, intermediate goals relevant to a single-year evaluation and long-range goals that can be evaluated only over a period of several years. Failing to make this distinction creates problems for bilingual programs, since some long-term goals (e.g., improved English skills) may not be applicable and measurable until the later grades. Long-term goals are also affected by the high rate of student turnover experienced by many bilingual programs. Since long-term goals would not apply to a short-term student, two sets of goals are required. This should be clearly stated and presented in the baseline data being collected.

Defining and describing student achievement goals is another important step in establishing baseline data. While there are many important considerations to recognize when specifying student achievement goals, the baseline data must include information on:

o   Subject areas (e.g., reading, language, math);

o   Languages to be used (e.g., English, Spanish, etc.);

o   Student language proficiency category (e.g., English: limited or proficient, Spanish: limited or proficient);

o   Grade level; and

o   Student affective goals (e.g., self-concept and attitudes towards school).

Because the original needs of the program, as stated in the proposal, may have changed, the information collected should be reviewed by the program director. Changes that have occurred should be properly documented.

4. <u>Describe the Instructional Program</u>

Establishing baseline data for the instructional program requires more time and effort than any of the other three areas on which information is collected. Baseline data collection on the program context, students, and program goals basically requires the review of existing records, files, and the original project proposal. Baseline data collection for the instructional program, however, requires face-to-face interviews of persons associated with the program, as well as review of program documents.

A description of the instructional program can be divided into three categories:

o   An overview of the program as it was originally
     designed and initially implemented;

o   A description of the instructional approach used
     in the program, including (1) student selection,
     (2) self-concept and cultural emphasis, (3)
     content of instruction, (4) presentation of
     content, and (5) scheduling; and

o   A description of the management of the program,
     including (1) staff organization, (2) staff roles,
     (3) staff development, (4) parent and community
     factors, (5) communication links with different
     audiences and (6) dissemination of program
     information.

The program overview information can be collected easily from information contained in the grant proposal. It should include the grade levels and number of classrooms served by the program, the amount of instructional time devoted to dual language instruction, and a definition of the program design (maintenance, transitional, etc.).

A description of the actual instructional approach used in the classroom and the basis for that approach require the most comprehensive description of any part of the bilingual program. This information is collected from program related documents, student files, classroom observations, and interviews with program administrations, teachers, and parents. This description is also the most important element used during the data analysis and interpretation. It is therefore essential that program personnel pay particular attention to this component. A detailed listing of the types of information that need to be collected is provided in the Program Information Acquisition Form found in the Technical Appendix.

A description of the overall program organization and management is the last requirement of the baseline data collection activity. This description will provide the basis for evaluating the operational effectiveness of the program. A detailed listing of the information that needs to be collected is provided in the Designer's Manual.

5. Document and Report the Baseline Data

Once the desired information is collected, attention should be focused on the various ways it is to be used. The information:

o Will be used as baseline information during the program monitoring activities of the evaluation process;

o Will provide a partial basis for planning the analysis and interpretation of student outcomes, as described in Chapter IV; and

o Will be reported directly to various audiences as part of the evaluation reports written for them.

Immediately after the preliminary data have been collected, the data should be summarized in the form that they will appear in the Final Evaluation Report and submitted to the program director for review. An initial analysis and interpretation of the data should be conducted to determine which variables, if any, are to be used as a basis for separate analyses.

CHAPTER III

CONDUCTING THE EVALUATION OF PROGRAM OPERATIONS

The successful completion of the planning activities and the establishment of the baseline data for the evaluation enable the program director to initiate the actual evaluation of the bilingual program. As described before, the actual evaluation of the bilingual program takes two thrusts: the evaluation of program operations and the evaluation of student outcomes. These may be viewed as totally separate activities. However, the outcomes or outputs of both activities are used during the analysis function to interpret the overall evaluation results and formulate recommendations for changes in the program. This chapter presents guidelines and procedures for conducting one part of the evaluation, the evaluation of program operations.

The evaluation of program operations employs the discrepancy evaluation design described earlier. Therefore the evaluation of program operations is performed by first developing a comprehensive description of the program describing how it was designed to operate. This establishes the baseline data for the evaluation of program operations. This activity was hopefully accomplished in accordance with the recommended procedures in Chapter II. Most importantly, this activity should have been completed during the first or, at least, by the end of the second month of the program year. The second activity required to perform this facet of the evaluation is to collect another set of data similar to the baseline data on the actual operation of

the program.   Decisions on what data to collect, how and when to
collect the data, and who will collect the data will have already been
made during the planning phase of the evaluation activity (See Chapter
I).   Most of these data are collected by reviewing program related
documents, monitoring classroom activities and interviewing various
persons associated with the bilingual program.   This set of data,
decribing actual program operation (e.g., the instructional method
being used; the amount of instruction in English; the number of
teacher aides assigned to a class, etc.) is compared to the baseline
data collected at the beginning of the school year, which describes
the program design.   The comparison provides the basis for determining
if the program was operated as planned.   If this is the case, there
should be few or minor discrepancies in the two sets of data which
describe the program.   If the comparison reveals significant
discrepancies or deviations, the evaluation must document why this
occurred.

Discrepancies in the program operations should not necessarily be
viewed as a negative finding.   There are many reasons why a program
may deviate from its original design.   The important task is to
determine if this deviation influenced the instructional program.   For
example, the program may have intended to provide one hour of
instruction in social studies using the student's native language.
However, due to scheduling modifications, teacher shortage, or other
factors, a change was made during the fourth month of the program and
the instruction did not occur.   The evaluation planning process,
nevertheless, most likely identified measures for this area.   That is,
the student outcome part of the evaluation was intended to measure the

performance of the students in social studies. The resulting student outcomes data may show that progress was minimal. However, knowing that instruction in the students' native language did not occur, the program director and evaluator can explain the resulting student outcomes. The question to be addressed, then, is why the program design was changed. Should the original design be reinstated? Answers to these and other questions begin to formulate a set of recommendations for the improvement of the overall program.

While, the example above ties the evaluation of program operations to the evaluation of student outcomes, it should be clearly understood that the primary purpose of this part of the evaluation is to examine and monitor the manner in which the program is being implemented. Additionally, the discrepancy evaluation design makes no attempt to infer or determine program impact.

This chapter provides some basic guidelines for evaluating the instruction, staff development, and parent involvement components of the bilingual program. While there are other facets of the program operations that merit attention, these components are the most significant to the overall operation of the program. The level of effort allocated to the evaluation of each of these components depends upon its emphasis and/or importance to the overall program, as established during the priority setting activities of the planning process. These issues should be addressed and resolved by the program director and evaluator in planning and designing the evaluation (see Chapter I).

GUIDELINES FOR CONDUCTING THE EVALUATION OF PROGRAM OPERATIONS

1. Evaluating Program Instruction

The evaluation of the instructional program is intended to answer the following two questions:

1. Are planned instructional methods actually being used?

2. Are changes needed in the instructional methods?

Data needed to answer these questions are obtained by observing classroom activities and interviewing program teachers and administrative staff. This core of information is then compared to baseline information, obtained through activities described in Chapter II in order to determine if the program is operating as intended.

The instructional program is the core of the bilingual program. The program director must ensure that the level of effort allocated to evaluate this activity is appropriate. Information on the operations of the instructional programs is obtained by (a) conducting classroom observations, (b) interviewing the teachers whose classrooms are observed, and (c) conducting supplemental interviews with a sample of program teachers and administrative staff.

49

Conducting Classroom Observations -- Prior to observing the classroom, the program director should review the program description so that program features which satisfy the goals and objectives can be observed. The features to be observed should be identified during the planning process. Classroom observations should become a planned activity of the program director. Following each informal observation, the program director should write a summary of the classroom instruction as it was observed. These brief summaries should be synthesized into brief reports at least three times during the year -- fall, winter, and spring. Later, these brief reports should be used during the comparison activity and incorporated into the final evaluation report. Thus, over time, the program director develops a complete picture of how the classroom instruction is actually being performed.

Topical areas that should be observed by the program director will, of course, depend on how the particular program is designed. Some general categories or features to observe include:

o   Language use;

o   Content of the lessons;

o   Teaching methods;

o   Diagnosis and grouping of students;

o   Recordkeeping;

o   Staff roles in the classrooms (teachers and aides);

o   Active participation by students; and

o   Attitudes and general morale.

Conducting Teacher Interviews -- Interviews with the teachers whose classes were observed may answer questions of whether instructional methods have changed from the original planned instruction, the reasons for the changes, and what changes in instructional methods may be needed.

Supplemental Data Collection -- In establishing the baseline data (Chapter II), interviews were conducted with program personnel, parents, and district personnel. A similar set of activities need to be undertaken to identify information about the actual operation of the program. Thus, the final step in evaluating the instructional program is to interview a sample of program personnel, parents, and local and district administrators. Information obtained from these interviews becomes a direct link to the interview data used in establishing the baseline data. Comparing these two sets of data is crucial in identifying discrepancies. The program director should plan to re-interview a sample of program personnel as well as local and district administrators to elicit information about actual instructional operations.

Once the interviews have been completed, the information should be synthesized by the program director and evaluator. This information is then compared to the baseline data so that discrepancies between planned and actual program operations can be noted.

Analysis of Program Instruction Data -- A determination of whether or not the instructional component of the program is operating as intended is made by comparing baseline information about the design

I-30

and plan of the instructional program (see Chapter II) to the information acquired from the evaluation of program instructional activities. This comparison leads to the identification of discrepancies between intended and actual program operations. Noted discrepancies identify areas or issues which may require decisions to correct the discrepancies. Later, these discrepancies may also be taken into account in the interpretation of student outcome data if the changes in the instructional program are determined to have influenced student performance. The triad of intended operations/instruction data, actual operations/instruction data, and student outcome data forms the basis for identifying final recommendations for the evaluation report.

Interpretation and Use of Results -- The results of these analyses is presented to those persons responsible for decisionmaking. The program director reviews and analyzes the data to determine if either immediate or future changes should be sought in the program operations and instructional methods employed. Frequent and immediate reports to the program staff should be provided by the program director. Such reports enable staff to review the intended changes, identify means of implementing the changes, and, consequently, be a part of the program improvement process.

Additional interpretation is performed by the evaluator. Using data from the various sources, the evaluator can examine the triad of intended instruction, actual instruction, and student outcomes to recommend changes which should be sought and ways to implement these changes.

I-31

2. Evaluating Staff Development

The evaluation of the staff development activities of the instructional program compares the actual training provided to teachers to that which was planned. The comparison provides decisionmakers with information about what training actually took place and how this training is related to the intended goals of the program, as well as whether the training met the needs of the program. Specifically, the evaluation of the staff development activities answers the following questions.

1. Were the staff development activities conducted as planned?

2. Did staff training activities meet the needs identified at the onset of the program?

3. Did staff participants acquire the intended knowledge and skills?

4. Were staff satisfied with the training provided?

5. Were skills acquired through training implemented in the classroom?

Answers to these questions when compared to the baseline information will identify discrepancies between actual staff development activities and intended staff training, as well as provide information on the actual training. A variety of data collection methods can be employed to obtain the data needed to answer the above questions. Methods such as questionnaires, knowledge tests, and observations of instructional techniques can be used to provide the necessary information.

I-32

53

Questionnaires -- Information regarding satisfaction with and outcomes of staff training activities can be obtained by questionnaires completed by the program director and staff. The Designer's Manual provides a sample questionnaire which can be used to collect information on the actual staff training activities. This questionnaire provides information about the type and duration of training; numbers of program staff involved in the training; and planned and unmet expectations and objectives for the training. This data should be collected within one week following the completion of all training activities which occur throughout the program year, or at the very least, near the end of the program year.

Appropriate analytic methods for analysis of questionnaire data are determined by the form of the data. The evaluator or appropriate member(s) of the program staff should review the questionnaire responses and systematically categorize the information according to the evaluation questions posed.

Knowledge Tests -- A more immediate source of information on the impact of staff training is information derived from administering knowledge tests to trainees during or at the end of the training. These tests, devised by the instructors, should focus directly upon the instructional content of the training. Because of the specificity of such tests, no sample instruments are included in this manual. The results of the knowledge tests can be examined from one or more perspectives. The tests could be administered prior to and subsequent to training, thus allowing comparisons to be made between

1-33

pre- and post-test scores. An alternative approach would be to use a control group not involved in the training program as a basis for comparison. An additional comparison could be made between the test results and the stated objectives of the training program.

Observation of Instructional Techniques -- The classroom observation process should yield information on the instructional approaches that are actually being used by teachers. To the extent that staff training is expected to affect instructional approaches used by teachers, the data acquired from the classroom observations are also pertinent to determine whether or not the training accomplished its purposes and is being implemented as planned. For example, it may be possible to determine if staff development activities intended to provide teachers with skills that are to be used in the classroom (such as how to use new materials, or administer tests) were successful by observing the teachers in the classroom.

Classroom observation data should be analyzed according to procedures described earlier in this chapter in order to identify discrepancies between intended and actual staff development activities. Specifically, the major goals of the staff training which pertain to teachers' instructional approaches should be compared with actual classroom practices as evidenced by classroom observation data.

55

Interpretation and Use of Results -- The program director should

examine the results of the analyses described above and determine if

the goals of the staff training were met, as well as determine if

findings related to staff training can be issued periodically

throughout the program year, possibly in conjunction with recommended

changes in program instructional operations. Program personnel then

will be able to provide reactions to the recommended changes and

identify possible approaches for implementation.

3. Evaluating Parent Involvement

The evaluation of the parent involvement component should address four

questions. These questions are:

1. To what extent did the level of parent involvement
    match the planned level?

2. Were parents satisfied with their level of
    involvement?

3. Was the program staff satisfied with the level of
    parent involvement?

4. To what extent and in what ways has parent
    involvement changed over the life of the program?

Data collected and used to answer these questions when compared to

information about the planned level of parental involvement identified

in Chapter II should determine if discrepancies exist. Data needed to

answer these questions can be gathered by interviewing parents and the

person responsible for administering the parent involvement component

of the bilingual program.

1-35

4. Reporting the Evaluation Results

The information resulting from the evaluation of program operations
should be summarized, written, and presented in the format in which it
will appear in the Final Evaluation Report. The format for reporting
the results will most likely be the same used to establish the
baseline data. However, the report should contain a section on the
evaluation findings and the recommendations being made to improve the
program.

57

CHAPTER IV

CONDUCTING THE EVALUATION OF STUDENT OUTCOMES

The most important goal of any educational program is to improve the performance of the students enrolled in the program. Therefore, determining student outcomes is perhaps the most important part of a program evaluation. The purpose of this chapter is to describe procedures for evaluating student outcomes. The student outcomes to be evaluated can be divided into the following four areas: English (L2) language skills; non-English or first (L1) language skills; academic achievement (e.g., in science, social science, and mathematics); and affective areas of student performance.

Conducting an evaluation of student outcomes is neither very technical nor complicated if the evaluation is designed to simply describe student performance. A student performance evaluation is interested only in determining how the students in the program performed, rather than determining what caused the observed level of performance. An attempt to measure the latter requires a more comprehensive evaluation design than the former. These two different approaches to the evaluation of student outcomes are commonly referred to as evaluations of student performance and program impact or effectiveness. The terms program impact and program effectiveness are used interchangeably in this Handbook.

These two types of evaluations are widely confused when conducting

evaluations of most educational (bilingual and other) programs. In particular, many evaluation reports make statements about program impact or effectiveness when actually they have only measured student performance. That is, they have observed that students have done better (or worse) than some standard or comparison group and then have taken the unwarranted step of concluding that the program was responsible. The Designer's Manual presents a more detailed discussion of the distinction between these two types of evaluation.

Evaluating Student Performance

Evaluations of student performance and evaluations of program impact are both based on the same kinds of measures such as tests scores or other quantitative measures, such as attendance rates. In both types of evaluation, student scores are compared to some scale or standard to give them meaning. Evaluations of student performance usually group student standards of performance into two categories. Those are:

- o   Absolute standards of performance which compare performance such as:

  - Comprehension level (of textbooks, newspapers, job application forms, etc.);

  - Mastery of specific skills such as language, math, or science; or

  - Proportion of days present in school.

These standards of performance are measureable in absolute terms.

59

That is, they provide information on what a student can or cannot do
and are not compared to any other external criteria.

o   Relative standards of performance (typically
        reported as percentile ranks or standard scores)
        may compare student performance against:

    -    Norm-groups (National, State, and
            local);

    -    Other bilingual students (National,
            State, and local);

    -    Groups of non-bilingual students in
            the same school or district; or

    -    Bilingual program students in
            previous years.

These, of course, are only examples. There are many other comparisons
that can be made. However, the more comparisons made the more
technical the evaluation becomes, often resulting in inappropriate
comparisons and misinterpretation of results.

Measures of relative performance should be the backbone of student
outcome evaluations measuring English-language skills and academic
subjects tested in English. Performance in other languages, generally
must be measured in absolute terms because meaningful comparison
groups will be difficult to find.

Evaluating Program Impact

Although determining the level of student performance should be the
primary goal of most program evaluations, many evaluations attempt to
go beyond this to demonstrate that the program is effective and

1-39

responsible for the observed level of student performance. Explicitly or implicitly, this question of program impact underlies most evalation designs. This Handbook recommends that bilingual programs do not attempt to conduct an impact evaluation. The Designer's Manual does provide information and guidelines for expanding the evaluation to determine program impact.

The laboratory approach to answering this question would be to divide the students randomly into groups--one or more groups for each type of program--and then to compare the effects of the different programs after some reasonable amount of time. In practice, however, because of the diversity of services and the characteristics of bilingual students, this is almost never possible. The result is that the effect of a program cannot be separated from effects of other factors in a conclusive manner. An evaluation using data from a single academic year probably should not even try to prove impact. However, data collected over several years can probably be used to develop an argument that, while not completely definitive, will be reasonably convincing as to the impact of the program. Bilingual programs should attempt to collect multi-year data on student performance.

Problems Associated With Accurate Measurement

In addition to the issues described above, impact evaluations, as well as evaluations of student performance, are themselves impacted by the measurement techniques available to measure performance. The predominant factor is the ability of the evaluation design and the

evaluator to control the "noise" or more commonly, the error of measurement.

It is generally accepted that test scores include some measurement error, and that student performance is affected by many things outside of the program. Therefore, the important issues for anyone involved in evaluation are (1) how much noise is there in a carefully done evaluation? and (2) can changes be expected in students (or impacts due to the program) that are significant enough to be measured in spite of the noise factor? This issue, as well as the characteristics of bilingual programs which impact the issue, are discussed in the Designer's Manual.

Because of all the problems associated with evaluation, the Handbook strongly recommends that evaluations of bilingual programs concentrate their efforts in conducting evaluations of student performance, rather than impact. This, together with the evaluation (description) of program operations meets the Federal evaluation requirements, as well as provides the program with sufficient information with which to make informed decisions on how to improve the program.

1. Developing the Evaluation Design

The first steps in performing the evaluation of student outcomes is to determine the type of evaluation that will be conducted and what questions the evaluation is designed to answer. The type of evaluation conducted, however, must address the minimum Title VII requirements.

Title VII requires that bilingual program evaluation include provisions for measuring the accomplishments of the instructional objectives, the progress of the students in improving their English language skills and a procedure for using the information to improve the operation of the program. Meeting these requirements is relatively simple and can be accomplished by following the procedures recommended in the Handbook. In order to meet these requirements, the Handbook recommends conducting an evaluation of student performance, rather than attempting to determine program impact. This can be accomplished by using the basic evaluation design provided in this Handbook.

The Basic Evaluation Design

Because of the difficulty in conducting program impact evaluations, the recommended approach to evaluate student outcomes is simply to evaluate student performance. This approach is referred to in this Handbook as the basic evaluation on the basic evaluation design. This basic evaluation design, therefore only answers the relative

performance question, "to what extent are the bilingual students achieving?"

The basic design has minimal requirements. These are:

o   Testing only the students enrolled in the bilingual program;

o   using adequate norm-referenced tests (NRTs) capable of measuring English language skills, first (L1) language skills, if applicable, and academic subjects (e.g., math, science, etc.); and

o   measuring performance for only one academic year.

Applying these minimal design requirements to the first student outcome component, English language performance, is all that is reqired to meet the Federal evaluation requirements. However, most bilingual programs should at least evaluate performance in two other outcome areas, first (L1) language and academic subjects. Additionally, although the basic design does not require a multi-year evaluation design, the Handbook does recommend that bilingual programs attempt to collect multi-year performance data. At a minimum, programs should strive to collect data over the duration of their grant period. It is conceivable that data showing progress over the life of the program, can be used to argue that the bilingual program was responsible for the outcome.

Expanding the Evaluation

Programs wishing to extend the evaluation beyond a description of

student performance to measure program effectiveness and/or impact
will need to enhance the requirements of the basic design. At a
minimum, these evaluation designs may require three modifications.
They will have to obtain test scores for comparison purposes from
students enrolled in other bilingual or non-bilingual programs.

Single-year evaluations only serve the purpose of the basic evaluation
design and can only document if the program is effective compared to
baseline data, but cannot show year-to-year changes. Therefore,
evaluations attempting to measure effectiveness will most likely
require multi-year evaluation designs capable of tracking students
throughout their participation in the program. Multi-year evaluations
require the use of the same measurement instruments throughout the
evaluation period and strict recordkeeping.

Evaluations attempting to measure effectiveness will most likely also
need to expand their measurement instruments beyond norm-referenced
tests. These may include criterion-referenced tests (CRTs), mastery
tests, and other types of measures. The Designer's Manual presents a
detailed discussion of these issues and provides options which may be
added to the basic design in order to attempt documenting program
impact.

Preparing for the Evaluation

Because the evaluation resources are limited, the evaluation may not
be able to answer all questions. Priorities must be determined with

respect to the most useful information to be obtained from an evaluation. The evaluation does not have to provide data on each student's learning outcome. The evaluation may provide data only on the students as a group. For example, measurements may be made of changes in reading achievement of third graders but not on reading achievement of a specific student in that grade. The evaluation does r.ot have to provide data on sub-skills such as phonetic analysis but rather on general skill levels such as reading achievement.

Certain decisions must be made before any data is collected to ensure that the analyses can be conducted as desired. Program goals need to be organized according to several key student or program features such as:

o   Subject area (e.g., reading, writing, speaking);

o   Language used in instruction (e.g., English, Spanish);

o   Student language proficiency category (e.g., English: limited or proficient, Spanish: limited or proficient);

o   Grade level of students; and

o   Year of the program.

The Designer's Manual provides a worksheet and instructions for preparing the evaluation activity.

2. Evaluating the English Language Component

The English language skills to be evaluated are the fundamental components to language use. These include knowledge of the sound system for oral language and comprehension of the orthographical system for written language. While each of the four language skill areas -- listening, speaking, reading, and writing -- can be considered individually, one component of language cannot easily be isolated from another. It simply cannot be assumed that mastery of one skill area necessarily indicates mastery of a related skill area; nor can it be assumed that lack of skill in one area indicates lack of skill in another. For this reason, the model recommends that proficiency in all four language skill areas be assessed.

Three Basic Design Decisions

For practical purposes, most programs must make three basic evaluation decisions: (a) which students to include, (b) what tests to use, and (c) what period of time to include. For each decision, the Handbook recommends a choice for a basic or minimal evaluation and then offer options that will let you answer additional questions if you have the necessary evaluation resources.

o Which students to include? The basic evaluation requires only testing the students enrolled in the bilingual program. An option could be to obtain data from other students in the district for comparison purposes. Theoretically, the bilingual program staff could pick out comparison groups and test them. In practice, though, this option is realistic only where there is a district-wide

testing program, and the scores for all district students are readily available on computers or in some other easy-to-use form.

o   **What tests to use?**   The basic evaluation requires a reliable, standarized, norm-referenced test (NRT) of reading and other language skills. Usually, the test used for district-wide testing may be used. Options include criterion-referenced tests, teacher-made tests, mastery-tests, and tests included as part of commercial instructional packages. We will refer to these kinds of tests generically as "CRTs, etc."

o   **What period of time to cover?**   The basic evaluation requires covering only one academic year and testing only once in the Spring. Two options are highly desirable: (a) multi-year designs following program students from one year to the next, and (b) baseline data on program-type students obtained before the program begins. A sub-issue is whether to test once or twice a year. The first choice should be to test only once a year in the Spring. Options are (a) once a year in the Fall or (b) twice a year, Fall and Spring.

These basic choices can be summarized as follows:

		Basic Evaluation	Optional Additions
1.	Students	Program only	Comparison groups from the district
2.	Tests	NRTs	CRTs (etc.)
3.	Term of Evaluation	Single year	Multi-year Baseline data
	(Time of Testing)	Spring only	Fall only Fall and Spring

Applying the Basic Design to the English Language Component

The basic evaluation design through the use of a norm-referenced approach provides for comparing bilingual program students to a national sample of students who scored at the same pretest percentile

on a nationally-normed test. For example, if the students in the bilingual program scored at the 25th percentile on the pretest, their growth can be compared to the growth of the students in the norm group who scored at the same 25th percentile on the pretest.

The norm-referenced approach makes the equipercentile assumption that a group of similar students who are not enrolled in the bilingual instructional program will maintain the same percentile rank throughout the year. This does not mean that the group without bilingual instruction is not learning. It simply means that their learning rate keeps them at a similar position relative to other students in their grade. In contrast, the students in the bilingual program will hopefully learn faster than they would in the program.

The question therefore being addressed is, "Do the students in the bilingual program increase their percentile ranking as compared to a national norm group who began at the same percentile?"

Key Comparisons to Be Made

There are many comparisons of performance that can be made. However, the five comparisons which follow are the ones that the evaluator may find useful and can be performed without using complex statistical procedures.

1. Are the students in the bilingual program making gains?

2. Is this year's student performance an improvement over past years?

3. Are students meeting the objectives of the program?

4. Are students doing better in the bilingual program than in another program?

5. Are students doing better than they would be expected to do without the program?

The answers to the first two comparisons can be easily answered by applying the basic design and using a norm-referenced test. The other comparisons require adding one or more of the options described earlier, such as a comparison group of students from another program.

Selecting Appropriate Tests to Measure English Language Skills

The criteria for selecting an achievement test to measure English language skills in a bilingual program are the same as those used in selecting a test for any evaluation. However, some criteria are more difficult to meet because few tests have been developed with the needs and characteristics of bilingual students in mind. Note also that a major assumption is made about the measurement of the English language component -- that the students learning English language skills have enough English language facility so that testing can occur in English. If this is not true, the students are likely being instructed in their native language and they are acquiring language skills in that language.

The basic evaluation design recommends the use of a standardized, norm-referenced test (NRT) of reading and other language skills to evaluate the English language component. Most school districts now

routinely administer one of these tests to all students. If the
district does not use a norm-referenced test (NRT) and NRT scores are
not readily available, the evaluator may choose to select one of the
tests described in the Technical Appendix. These tests are
reasonable, reliable, and valid. The main concern should be that the
test content matches the program curriculum, at least on a general
level. If this basic check is not made, it may later be discovered
that the second-grade test covers third-grade curriculum, and vice
versa.

There are two major problems to consider in selecting NRTs. These
are:

o   Test level (floor and ceiling effects). In some
     bilingual programs, the at-grade-level test is too
     difficult for program students at pretest. The
     next lower level may be too easy at posttest time.
     If the mean score on a test is less than 25% of
     the items correct or more than 75% of the items
     correct, floor or ceiling effects probably exist,
     and the test cannot give an accurate picture of
     either student performance or program impact (See
     Out-of-Level or Functional Level Testing in the
     Technical Appendix).

o   Multi-year and multigrade-level requirements.
     Most bilingual programs cover several grade
     levels. Therefore, it is desirable to have
     achievement tests that can be used to compare
     progress across grades and that can be used to
     follow groups of students as they progress through
     the grades. In practice, this means using any one
     of the recognized achievement tests.

I-50

71

<u>Using CRTs (etc.) for Evaluating the English Language Component</u> -- The choice of CRTs (etc.) is more of a curriculum decision than an evaluation decision in most districts. That is, when developing objectives and curriculum materials for a bilingual program, many districts either develop or buy tests matched to their curriculum and the instructional materials. These tests are the best candidates to use in your evaluation. If you have important objectives for student performance that are not covered by any other tests, you may wish to develop or buy special tests just for evaluating student outcomes.

3. <u>Evaluating the Non-English Language Component</u>

Bilingual programs, for evaluation purposes, can be divided based on their non-English language component into three types. These are:

o   Spanish only programs;

o   Single languages other than Spanish programs; and

o   Multiple language programs.

The major differences among these three types of programs, from the evaluator's perspective, are: (a) only Spanish-English programs will find commercial tests readily available, and (b) multiple-language programs often include small groups that cannot be combined for evaluation purposes.

<u>Three Basic Design Decisions</u>

The three basic decisions made for the English language component also apply to the non-English language component: (a) which students? (b) what tests? and (c) what time period? However, the decisions are even simpler for the non-English language component, because there are fewer alternatives available to the evaluator. The basic options can be summarized as follows:

o   <u>Which students?</u> In general, only the bilingual
     program students will speak the languages in
     question and therefore the <u>only</u> students that can
     be included in the evaluation. In a few
     districts, there may be comparison groups of
     interest from other programs or other districts
     who use the same tests. However, in most cases,

only your program students will be tested in the
non-English language, making comparison groups
unavailable.

o   <u>Which tests?</u> A limited number of standarized
    tests are available in Spanish (although their
    norm groups are not analogous to those from
    English-language tests, and you should not use the
    norms as a simple standard of comparison). For
    other languages, you are limited to, at best, a
    few commercial, criterion-referenced tests, plus
    locally-made tests (CRTs, etc.).

o   <u>What period of time?</u> Here, the evaluator has the
    option of single-year or multi-year designs since
    baseline data before the start of a new bilingual
    program could be collected. However, in practice,
    few districts will do this. In general, if the
    English language evaluation is multi-year, the
    non-English language evaluation should also be
    multi-year. Otherwise, both should be single-year
    evaluations.

The decision on once-a-year (Spring) verus twice-a-year (Fall, Spring)
testing will probably also be the same for non-English testing as for
the English language testing.

The basic choices are summarized below.

		<u>Basic Evaluation</u>	<u>Optional Additions</u>
1.	Students	Program only	None from the district
2.	Tests	CRTs, etc. (NRTs for Spanish)	None
3.	Term of Evaluation	Single year	Multi-year
	(Time of Testing)	Spring only	Fall only Fall and Spring

<u>How to Select Among the Options</u> As you can see, the only real option
is whether to include the non-English language component in the

evaluation at all. If you want to know how your students are doing in this area, you will almost certainly be able to produce teacher-made tests that will serve your purposes, but you need to consider exactly which questions you can answer with such tests.

Key Comparisons to be Made

The key comparisons that can be made relative to non-English or first (L1) language skill development/performance can be the same as those made for the English language component. Performance measurement agains norms will only be possible for Spanish language performance. Therefore, answering the first comparison question for other languages will have to be made by using locally developed mastery tests.

Answering the other questions may be done by following the same procedures as before. Answering the fourth question, which requires a comparison group, should not even be attempted.

Selecting Tests for the Non-English Language Component

Selecting tests for this component is difficult because there are very few tests available. Spanish versions are available for the Inter-American Tests, the CTBS, and the ETS Circus test. However, conventional non-English language norms do not exist. The Inter-American Tests (Spanish) provide user-norms based on students in bilingual programs using that test. The norms provided with the Spanish CTBS do not represent the population of Spanish/English

bilingual students. Norms for both tests can only provide comparison standards for student performance evaluation, and these comparisons are difficult to interpret. So far as the review of literature indicated, no large-scale norm groups have been tested in any other languages.

4. Evaluating Student Performance in Academic Areas

Evaluation of performance in academic areas requires the specification of the skills to be assessed, selection of the language in which skills are to be measured, and the identification of appropriate tests in English and/or the first language of the student. The evaluator will need to determine which skill areas are to be included in the evaluation. Measurement of achievement in literacy as well as in major academic subject areas may be appropriate. This determination will have to be made on a program-by-program basis. If a student is not literate in L1 or L2, then achievement testing will not be appropriate. If the students are literate, the language in which to test the students will depend upon the language in which instruction in the particular subject has been given, as well as the fluency of the student in that language.

The Basic Design

Many bilingual programs include non-language, academic subjects, such as math, social studies, and science. The same principles that apply to the English language component apply to this component if testing is done in English. A minimal evaluation would consist of (a) testing program students only, (b) using standardized, norm-referenced tests, and (c) a single-year design. Options include local comparison groups, longitudinal designs, and baseline data.

Language of Testing

The major issue in evaluating performance in academic subject areas is whether to test in English or in the first (L1) language. The evaluation will be easier to implement and the results easier to interpret if the testing is done in English. However, as a matter of common sense, if the students are weak in English and much stronger in their native language (e.g., new arrivals or young children from non-speaking homes), then testing in the native language may be required. In such cases, the evaluation design principles for non-English language components apply (see above).

Selecting NRTs

By and large, the discussion of tests for English language also applies to tests for academic subjects tested in English. The discussion of non-English language tests applies to tests of math, science, etc. in non-English languages. The basic rule here, as it was for English language, is to utilize the test that is used throughout your district. The <u>Technical Appendix</u> contains a discussion on the selection of achievement tests, as well as a listing of these tests for testing language, mathematics, science, etc.

Using CRTs (etc.)

As in language testing, if you have test data available from your instructional program on math, science, or other subjects, you may

want to include these data in your bilingual-program evaluation. For

subjects tested in languages other than English or Spanish, you may

have to depend on teacher-made tests, and the normal cautions apply.

5. Evaluating Affective Areas of Student Performance

Affective goals, like improving student attitudes or behaviors, are mentioned in connection with many bilingual programs. If your program has specific objectives in these areas and if the program includes specific components that are intended to change student attitudes or behaviors, then you should consider evaluating the effects of these components. However, you should be aware of two problems, which are discussed below.

Affective goals must be clearly defined. In many bilingual programs, the non-acadmic goals are defined in very general terms, such as "improving self-concept." The test chosen to evaluate changes in self-concept may be some readily available commercial attitude test that bears very little relationship to the self-concept of the program students. The esults are almost certain to be meaningless.

If you wish to evaluate affective components of your program, then you must define the goals clearly, describe the components of the program that are intended to address the goals, and then identify appropriate measures, such as tests, attendance records, and so on, that match your goals. Then you can begin to consider an evaluation design to evaluate absolute student performance, relative student performance, and program impact in the areas that you have designed.

Affective goals are very difficult to evaluate. While the general evaluation design principles apply theoretically, in practice it is

very difficult and frustrating too evaluate changes in attitudes, self-concept, and so on. This is because (a) there is a great deal of noise in the measurement, (b) most measures are insensitive to change in attitudes, (c) attitudes change greatly from month to month and even from hour to hour, (d) there are few good absolute criteria available, and (e) there are seldom any very good comparison groups available.

The net result is that few evaluations can provide convincing evidence of changes in attitudes or related characteristics of the students. For this reason, we would not advise bilingual programs to invest much of their effort in evaluating these goals unless they are a major focus of the program.

Programs wishing to measure affective areas may consult the Technical Appendix. This volume contains a discussion of self-concept and a listing of different tests available.

6. Conducting the Data Collection Activity

Data collection for the first component of the evaluation, program operations, consists of obtaining student background information, interviewing teachers, program administrators, and parents; as well as observing classroom operations. Data collected for evaluating student outcomes consist of test administration, scoring, and the recording of test scores. The latter activity probably requires a higher level of effort than the former. However, data collection for the student outcome component requires strict discipline and very precise procedures.

Testing the Students

Testing in the academic program areas -- language, math, science, and so on all require the same basic procedures. The main distinction that the evaluator should make is between formal testing for evaluating student outcomes and informal testing for diagnostic or other instructional purposes, and out-of-level or functional level testing. Each type of testing and testing procedures are described in the Designer's Manual.

Scoring of Test Data -- One of the issues in scoring tests and recording the scores is whether to use computers. If the program is very large, the answer should probably be "yes," at least for norm-referenced tests. Many programs have access to district, university or state computer centers that can perform the scoring of

the tests. If these services are not available locally, the test publishers or other scoring services can provide them. Hand scoring and recording may still have to be performed for very small programs. In addition, if non-standardized tests are used, it may be necessary to score the tests by hand before entering the scores into a computer for analysis.

Recording Test Data -- Recording the scores is the final step in the data collection process. To ensure that the scores will be usable, the details of recording should be planned well before pretest time. Where a commercial scoring service is used, the evaluator may have little control over the recording process, but if the program elects to do its own scoring or wishes to transfer scores from computer printouts to a more convenient form, the evaluator must consider two important issues: (a) the accuracy of the data, and (b) the details of the data recording forms. The Designer's Manual provides more detailed guidelines for scoring and recording the data.

83

7. Analyzing Student Outcome Data

The analysis of the student outcome data should be performed or at least supervised by a trained evaluator. The analysis of student performance data should simply answer the questions which the evaluation was designed to answer and make the necessary comparisons that were established during the evaluation design phase. There are three steps in this approach:

o   Examine scores for serious mistakes or unusual results. The scores can be examined simply by drawing the frequency distributions of test scores. If two sets of scores are being compared for the same students (for example, second-grade and third-grade scores) then scatter diagrams of one test against the other should be used.

o   Compute the mean scores and standard deviations for program (and comparison) students. If the scores do not appear to reflect any serious problems or unusual program effects, then simply compute the mean score for each group of program students (and for each group of comparison students, if any). The standard deviation (a measure of how spread out the scores are) must also be calculated and reported for each group. The mean scores are used to draw comparisons or look for progress of the students.

o   Estimate the possible effect of error on your results. What may appear to be changes in student performance may only be random changes in the scores due to noise (error). Errors, in mea scores of 5 to 10 NCEs are not uncommon, especially with small groups of students.

In examining the data from the evaluation the evaluator should check to see if the distribution scores resemble a normal curve (bell shaped). If the distribution of scores is a different shape, this could indicate possible problems with the tests, testing procedures,

the scoring procedures or the data computer programs. An abnormal distribution in the data may also be attributable to the effects of the program on specific students. For example, in one bilingual program, the mean scores could show second grade students making a moderate percentile or normal curve equivalent (NCE) gain in reading. However, when individual students scores are analyzed, it may be found that only a few students in that grade have made very large gains while the rest of the students have made little or no change in their percentile standings. This information is useful to the evaluator in concluding that the program is working for some students but not for others. Using this finding, the program director may be able adjust the program for those students not showing improvement in reading.

Another problem in analyzing the data from the evaluation is the kinds of noise (error) that remain in even the best evaluation data. Consideration should be taken to ensure that change in students test scores are not due to noise but too the effects of the programs. Error in mean scores of 5-10 NCEs are not uncommon, especially for programs with small numbers of students. Tests of statistical significance provide the best way of estimating the likelihood that the results are simply examples of random error. However, tests of statistical significance do not provide information about the educational importance of results, since small gains can be statistically significant for large groups of students, while what appear to be large gains can be due too random error with small groups of students. Tests of statistical significance also will not indicate flaws in your evaluation procedures. Thus, individuals responsible

for conducting the evaluation should look for possible problems in the evaluation procedures. The Designer's Manual presents a thorough discussion of this issue.

Analyzing the Data for Program Impact Evaluations

Analyzing the data for program impact requires a demonstration that the program has had an impact on student performance, it must be shown that student performance is better than expected, and that the program and nothing else is responsible. This does not require any special analysis of the data. It requires the use of data from the program operations evaluation component and student outcomes to build a convincing argument. In addition to the three analytic steps described above, proving program impact will require three basic elements to build a convincing argument. These are:

o   Evidence that students have improved their performance. This type of information documents that similar students in the same schools had lower scores in the past. This requires compiling data from several different years.

o   Evidence that non-program students have not made a similar improvement. This type of information examines the possibility that something outside of the bilingual program, such as a new principal or a new district-wide curriculum, is responsible for the improvement in bilingual student performance. This information can only be generated by having local comparison groups -- preferably from district-wide test data.

o   Evidence that the characteristics of the bilingual students have not changed since entry into the program. In some districts, the student population can change drastically over a period of a year or two (as when large numbers of new arrivals enroll). Some evidence that changes in student population are not responsible for the

changes in student test scores must be
demonstrated.

Analyzing evaluation data, especially program impact evaluation, is
careful, systematic detective work. It consists of looking for clues
and followup of any leads that may help to explain the effects (or
lack of effects) that are observed in data. A clever and thoughtful
evaluator can often build a convincing case by assembling a variety of
evidence. Unless it is specifically required that the impact of
program be assessed, it is better to spend the effort in developing
the instructional program.

8. Interpreting the Results of the Evaluation

The analysis of student outcome data described above, provides the program director and evaluator with the quantitative information on student performance. If a norm-referenced test was used, the data will show how the bilingual students compared in achievement to a national norm group. Hopefully, the results will show that bilingual students achieved as well or better. These results, however, do not provide answers as to why the students achieved. The answer to this question may possibly be found by carefully examining the results emanating from the evaluation of program operations.

The evaluator should understand that the two components of the evaluation model, the discrepancy evaluation of program operations and the evaluation of student performance, are not methodologically linked together. As a matter of fact, each component may stand alone. The baseline data developed for the evaluation of program operation, however, does play a role in designing the evaluation of student performance. That is, the baseline data provides information to determine what outcome areas should be evaluated.

In addition, the results of the program operations evaluation can provide the evaluator with valuable information on how the program was operated, the instructional approach used, and the amount of instruction provided in the first language for each academic subject area, etc. This information can be used to "understand" the results of the student outcomes component of the evaluation. This information

is valuable to a perceptive evaluator wishing to find answers to explain student performance. For example, if the discrepancy evaluation shows that history was taught using the first language to fourth grade students, but not to students in the fifth grade, the evaluator may want to closely examine the test scores in history for those two grades. Depending on what the test scores show, the evaluator may be able to make some assumptions on what caused either the same or different level of performance. The evaluator may then want to more closely examine "how" the instruction was provided. For example, the evaluator may want to ascertain the level of language proficiency of the teacher teaching in the first language or compare the language assessment scores, if available, of the students in the two grades. All this information, when processed together, could provide clues for understanding what caused the level of performance.

Because the two components of the evaluations are not methodologically linked, there are no specific procedures that can be described for merging the two sets of data. Nevertheless, the recommended approach provides the evaluator with a significant amount of information to use in arriving at conclusions about the program. The analysis techniques required for the evaluation, as described earlier, are relatively simple and can usually be performed by following the instructions in the test manuals as well as the discrepancy procedures described in this Handbook. The other ingredient is the creativity of the evaluator and project director in their ability to use the information to better understand the program and how it might have impacted student performance.

CHAPTER V

PREPARING THE EVALUATION REPORT

Preparation of the final evaluation report is an important activity of
the evaluation. The evaluation report is the final and most visible
product of the evaluation. Steps should be taken to assure that the
report addresses the purposes and specific questions of the
decisionmakers for whom the evaluation was planned. In addition, the
evaluation results should be reported in a timely manner, taking care
to ensure that the technical aspects of the evaluation effort are
clearly presented. Together, these steps increase the usefulness of
the evaluation results.

Preparation of the final evaluation report can be a time-consuming and
burdensome process if not properly planned. However, reporting should
be a continual process occurring throughout the evaluation cycle. The
preparation and sharing of evaluation information throughout the
evaluation cycle also serves to strengthen communication between the
evaluation audiences and those conducting the evaluation, thereby
increasing the use of evaluation results.

There are a number of basic principles which pertain to the reporting
process and serve to simplify preparation of the final evaluation
report. This discussion assumes that completion of the report is the
primary responsibility of the program evaluator(s) contracted to
undertake major segments of the bilingual program evaluation.
Basically, the evaluator has three important tasks: develop an

understanding of the audiences who will use the information, select proper reporting format(s), and assist the audiences in using the results. Proper planning of the reporting requirements will make this final activity easy to complete.

The evaluator must understand that clear communication requires knowledge and understanding of the evaluation audiences. The identification of the audiences should have been completed during the planning stages. However, it is helpful to review who the audiences are at the time of reporting. The evaluator should periodically communicate with the audiences to identify their information needs and their understanding of evaluation issues, such as testing. This will help the evaluator to tailor the report specifically to the level of understanding of the audiences and to determine the best form in which to report the results. Contact with the audiences also increases the probability that evaluation results will in fact be used.

Evaluation reports can take different forms, but whatever the form, the report should be designed for a specific audience and be presented in a manner that allows for response and interaction. Although the most common format is a written report, which describes the entire evaluation, consideration should be given to alternative versions for various groups.

Oral presentations are also a major vehicle for reporting to professional audiences such as teachers and program staff. Oral presentations are particularly important for highlighting the major

findings, conclusions, and recommendations, and for establishing
two-way communication that will clarify, interpret, and influence
decisionmaking. Such presentations can be enhanced by a panel
discussion and/or small group discussions of the reported results.

Whatever reporting formats are used, the evaluator must focus on the
audience(s) and their specific needs. The amount of attention given
to the form of reporting may make the difference between a report that
is simply received and one that influences practice.

Several standard elements should be included in the report. These
include:

- o   Statement of purpose;

- o   Program overview and background;

- o   The goals and objectives of the bilingual
      program;

- o   Description of the program and students;

- o   Discussion of the methodology used; including
      design, sampling strategy, instrumentation, and
      data analysis procedures; and

- o   Presentation of the findings, conclusions, and
      recommendations for program change.

The report should be concise and should include easily interpreted
tables, graphs, and other figures limiting the amount of narrative
material presented. Important issues should be identified and
highlighted in the report if the results of the evaluation effort are
to be maximized. Techniques such as boxing in recommendations or
using a different type face are useful to highlight the most important

points of the report. Examples of actual data collection instruments
should be included in an appendix.

Once the written report is completed, copies must be submitted to the
funding agency. Plans should also be initiated to present the results
of the evaluation to specific audiences. Consideration must be given
to identify the appropriate person responsible for presenting the
results. It is recommended that this be the program director and/or
the evaluator. A decision as to which of the two will report to which
audiences will be dictated by the individual situation and deserves
careful consideration.

93

VOLUME II

A

DESIGNER'S MANUAL FOR CONDUCTING AN EVALUATION

OF

ESEA TITLE VII BILINGUAL EDUCATION PROGRAMS

Prepared By:

INTERAMERICA RESEARCH ASSOCIATES, INC.

1555 Wilson Boulevard

Rosslyn, Virginia 22209

MAY 1982

94

TABLE OF CONTENTS

Volume II

CHAPTER V: <u>PREPARING THE EVALUATION REPORT</u>

OVERVIEW

This document represents the second in a three-volume series constituting the Handbook for Evaluating ESEA Title VII Bilingual Education Programs. The Handbook provides practical guidelines and recommended approaches for bilingual education program directors and evaluators to use in evaluating bilingual programs.

In the development of the Handbook, it was readily recognized that a single document would not be equally suitable for all bilingual education programs. Obviously, bilingual education programs cover a range of languages and grade levels in a variety of settings. Some programs have large evaluation budgets and access to teams of trained and experienced evaluators. while others have limited budgets and limited human resources.

Therefore, this document -- Volume II, The Designer's Manual for Conducting an Evaluation -- is designed to provide program directors and evaluators, with specific guidelines, recommended procedures, and selected materials, such as worksheets and checklists, to use in the evaluation. The manual provides the conceptual framework for the recommended evaluation and data gathering model. The manual is divided into five chapters, each describing one of the five activities of the evaluation. These are: Planning the Evaluation; Establishing the Baseline Data for the Evaluation; Evaluating Program Operation; Evaluating Student Outcomes; and Reporting the Evaluation Results.

Each chapter presents a detailed explanation of the intended activity
and provides step-by-step procedures for using the checklists and
worksheets in conducting the evaluation. Sample evaluation
instruments, such as interview schedules and forms to gather other
types of data, are provided in reduced format. Full size copies are
provided in the Technical Appendix.

Volume I, entitled The User's Guide to Evaluation Basics discusses
evaluation issues and summarizes the procedures required to conduct an
effective evaluation. The guide provides a summary description of the
five components of a bilingual education program evaluation. The
guide is intended for program directors, as well as for persons
associated with the bilingual program, but not involved in the actual
evaluation activity.

Volume III, entitled The Technical Appendix, contains a collection of
references covering various evaluation issues. These are intended to
assist program directors and program evaluators in building upon or
expanding the evaluation activities identified and discussed in
Volumes I and II. The volume also continued full-size, reproducible
copies of the checklists and worksheets found in the Designer's
Manual.

CONCEPTUAL FRAMEWORK FOR THE EVALUATION

Bilingual education programs represent a unique instructional approach, using two languages to meet their educational goals by generally providing instruction in academic subjects using the student's first (home) language (L1) while developing the English language skills of the students. The students served by bilingual programs also reflect a wide diversity in culture, socio-economic status, and educational experiences. These aspects distinguish bilingual education programs from all other instructional approaches.

The primary goal of bilingual education programs is the development of English language skills of the students as well as the development of their home language. Teachers recruited to teach in these programs, therefore, need to have language skills in the two languages being used for instruction. Curriculum materials in the first language are also needed.

Other goals of bilingual programs often include the development of the student's self-concept by emphasizing the home culture and the improvement of his or her performance in other academic projects. In order to accomplish these goals, knowledge of the students' culture by the classroom teacher and culturally relevant curriculum materials are a necessity in bilingual programs.

Due to these factors, the evaluation of bilingual education programs must be performed with considerable caution. The selection of an evaluation approach must take into consideration the variety of educational services, the curriculum materials used in the classroom, the number of hours of instruction provided in English and in the first language, the language skills of the classroom teacher, as well as the educational experience and language skills of the students.

Because of the complexity of this educational context, experimental or quasi-experimental evaluation designs are often not appropriate to evaluate bilingual programs. Experimental designs usually require random selection of students. However, random selection is not realistic in a bilingual education context, because it would require that students who are eligible to receive bilingual education instruction to be placed in alternative programs for control purposes. Similarly, the unique and differing characteristics of the students and the difference in the instructional services they receive make it very difficult to find comparable comparison groups necessary for quasi-experimental designs. The consensus of the literature addressing the evaluation of bilingual programs also indicates that the use of standardized tests to evaluate bilingual student progress is of dubious value. Despite these limitations, some formal measurement of student academic achievement must be undertaken in bilingual education programs.

The Recommended Evaluation Model

The evaluation model presented in this Handbook, therefore, is solely designed to provide descriptive information about the operation of the bilingual program and on the academic performance of the students enrolled in the program. The information gathered through this process can be used to evaluate student progress and to some degree provide a barometer of program effectiveness. The model is based on the premise that an evaluation of a bilingual program should:

- o Provide descriptive information about the operations of the bilingual program; and

- o Provide information describing student performance (even if hindered from making inferences about program impact).

Therefore, the model requires the collection of student outcome data to determine if the students are making progress in their learning. It also requires the collection of information on "how" the program is operating.

The model is also practical and realistic in relation to the financial and human resources available to conduct evaluations of bilingual programs. Aside from the expertise and time of the immediate personnel of most programs, the majority of bilingual programs have limited funds (generally between $2,000 to $5,000 per year) to secure private consultants to perform or assist with the evaluation. Therefore, the model takes into consideration the amount of time and effort that can reasonably be expected to be given to the evaluation effort.

3

The recommended evaluation model consists of two components. The first component focuses on program operations (e.g., program goals, time spent on instruction, etc.) using a discrepancy evaluation design. The design places heavy reliance on descriptive data about the program; therefore requiring as an initial step, the establishment of comprehensive baseline data on the program, the students, and the community.

The actual evaluation and data collection activities needed for the evaluation of program operations are performed primarily through search and review of program documents such as the grant proposals, previous evaluation reports, student files, and related material, as well as personnel interviews and the monitoring of classroom instruction. Personnel interviews to gather information on how the program is being operated are conducted with the program director, teachers, district administrators, and parents. Monitoring of classroom instruction is performed through observation to determine if the instruction is being carried out as planned and in accordance with the original program design.

The discrepancy evaluation attempts to identify and document differences between the initial plans of the program and the actual manner in which the program is operating. Information about discrepancies between the planned and actual program activities, as identified by the discrepancy evaluation, may be used to make decisions on how to continue operation of the program and what changes might be required.

4   103

The second component of the model requires the assessment of student outcomes. The student outcomes to be evaluated are:

o   English language skills;

o   First language skills;

o   Academic achievement; and

o   Affective areas of student performance.

Because of the difficulty in conducting program impact evaluations, the recommended approach to evaluate student outcomes is simply to evaluate student performance. This approach is referred to in this Handbook as the basic evaluation on the basic evaluation design. This basic evaluation design, therefore only answers <u>the relative performance question</u>, "to what extent are the bilingual students achieving?"

The basic design has minimal requirements. These are:

o   Testing only the students enrolled in the bilingual program;

o   using adequate norm-referenced tests (NRTs) capable of measuring English language skills, first (L1) language skills, if applicable, and academic subjects (e.g., math, science, etc.); and

o   measuring performance for only one academic year.

Applying these minimal design requirements to the first student outcome component, English language performance, is all that is reqired to meet the Federal evaluation requirements. However, most bilingual programs should at least evaluate performance in two other

outcome areas, first (L1) language and academic subjects. Additionally, although the basic design does not require a multi-year evaluation design, the Handbook does recommend that bilingual programs attempt to collect multi-year performance data. At a minimum, programs should strive to collect data over the duration of their grant period. It is conceivable that data showing progress over the life of the program, can be used to argue that the bilingual program was responsible for the outcome.

Data resulting from the analysis of student outcomes can be used as an indicator of overall student performance. The data from this component of the evaluation, in conjunction with the discrepancy data can be used to determine what program changes, if any, may be required to improve student performance.

For example, the discrepancy evaluation of program operations may reveal a significant operational change from the original design of the instructional program. This change could have had considerable impact on the instructional program, to the extent that student performance may have been affected. Knowing this, the evaluator will be able to analyze and interpret the outcome data affected by this change and make recommendations for changes in the program.

In summary, the purpose of the recommended evaluation model is to describe student performance and program operations. It can not be used as a measure of program impact. The recommended model meets all the requirements established in the Title VII rules and regulations.

6

The regulations require that each grant have a plan to evaluate the progress and achievements of the bilingual program. The plan must include:

o   provisions for measuring the accomplishments of the instructional objectives of the program;

o   provisions for measuring the progress of the students in improving their English language skills; and

o   a procedure for using the information gained from the evaluation to improve the operation of the program.

The recommended evaluation model accomplishes this by:

o   performing an evaluation of program operations using a discrepancy evaluation approach;

o   conducting an assessment of student performance in developing English language skills, as well as first language skills and performance in academic subjects; and

o   conducting an analysis function to determine what changes may be required to improve the overall operations of the bilingual program.

The Handbook recommends that bilingual programs should not attempt to determine program impact. However, some basic guidelines for extending the evaluation to determine impact are presented as optional activities to the basic evaluation design. Extending the evaluation beyond the basic design, however may require more resources than those normally possessed by Title VII bilingual education programs. The Handbook also does not address entry and exit procedure issues. The procedures are, however, very much intertwined with evaluation of bilingual programs and should be considered when planning the evaluation.

CHAPTER I

PLANNING THE EVALUATION

Planning is the single most important task in conducting an evaluation. Although this point seems obvious, research indicates that many evaluations of bilingual programs, as well as evaluations of other educational programs, are not properly planned. Many evaluations occur towards the end of the program year as a last-minute thought, simply to produce a report to satisfy some external requirement, usually imposed by the funding source. As a result, they are often performed haphazardly and produce poor results.

Evaluations performed in this manner are of little use to either the program itself or the funding agency. These evaluations usually fail to address issues that program and school administrators may have about the program because the evaluation design failed to incorporate their concerns during the planning process. Likewise, these evaluations will not be helpful to the funding agency since, at best, they were planned too late in the program year to capture useful information and, at worst, merely represent perfunctory efforts to fulfill a reporting requirement.

The evaluation process, to achieve its own objectives, must be approached in a serious manner and receive as much priority as other elements of the educational program. Program administrators must

realize that the evaluation process is a positive activity designed to provide information on which to base decisions for program improvement.

The planning process carefully balances the reporting requirements of the funding agency, the information needs of decisionmakers and program administrators, and the scarce resources available to conduct the evaluation. It is unlikely that any given bilingual program will have the resources needed to address all the information needs of its different audiences. Therefore, all parties concerned must realize that compromises will have to be made; otherwise, resources will be scattered, producing little useful information.

A properly conducted evaluation requires more than simply evaluating a specific component of a bilingual program (e.g., student performance): An effective evaluation plan identifies all the questions about the program that the evaluation should answer.

The evaluation planning process, therefore, involves a series of carefully executed steps which identify the evaluation audience and their specific information needs, set priorities, determine which program components to evaluate, allocate scarce evaluation resources, and set timelines for the evaluation process.

Next to proper planning, effective management of the evaluation process is a must. One person must assume the responsibility and have the authority to direct and manage all facets of the evaluation. All

persons involved in the evaluation process must be made aware of the authority and be given instructions and directions on how to interact with that person. A clear chain of command must be delineated. In most Title VII programs, the program director retains and assumes that responsibility. For purposes of presentation, this Handbook assumes that the program director is the person responsible for ensuring that the evaluation is planned and conducted.

1. Select an Evaluator and Assign Responsibilities

Proper planning and effective management of the evaluation dictate that the person responsible for designing and conducting the more technical aspects of the evaluation be identified as early as possible in order to become involved in the early decisionmaking of the evaluation planning process. In the case of most Title VII programs, this person is usually an independent consultant from outside the school system. Ideally, the evaluator should be involved in the original design of the bilingual program itself. In the case of Title VII programs, this should occur during the proposal writing stage. This would enable the evaluator to begin working with the program director in planning the evaluation before the academic period to be covered by the evaluation commences. The plan for conducting the evaluation, if at all possible, should be completed prior to the first day of school of the academic year being evaluated.

A major responsibility of the program director is to survey the available human resources in the district and, assuming he or she has

the authority, decide whether to use an evaluator from within the school system or employ an independent evaluator. The possibility of contracting for the services of an evaluation specialist from a university or a private consulting firm must be weighed against the potentially lower cost to the program if the evaluation can be conducted by district personnel. The program director must decide on a course of action.

The program director should attempt to ensure that the person selected as the evaluator have a thorough understanding of the goals and objectives of bilingual education and be experienced in using measurement and evaluation techniques with limited-English-proficient students. Because it may be difficult to find a skilled evaluator who understands the special problems of bilingual programs, it may be more desirable, if affordable, to select a team of evaluators who, as a group, possesses all the experience and required skills.

Assuming that an independent consultant or a consulting firm is contracted to perform or provide assistance in conducting the evaluation, the program director should assign clearly defined responsibilities and specific assignments to the evaluator, the program personnel assisting with the evaluation, and himself.

The program director, assuming he or she is the person in charge, must take the lead in delineating these responsibilities, determine the evaluation objectives and information needed from the evaluation activity, and ensure that the evaluation is successfully conducted.

In this respect, the evaluation of the bilingual program is just
another activity managed by the program director that occurs and is
implemented as planned.

The evaluator's function and responsibilities are usually determined
by the amount of technical assistance needed by the program director
in carrying out the evaluation.   The evaluator's role is therefore
generally narrower in scope, focusing more on technical matters such
as test selection, designing data collection procedures and
instruments, conducting data analyses, and reporting the evaluation
results.   The evaluator may often serve as a technical consultant to
the program director during the planning and implementation stages.
This role of technical advisor and consultant can be valuable to the
program, since the evaluator can provide immediate, informal feedback
on how the program is being implemented.   Often, problems of program
design, implementation, and management can be identified and remedied
in the early stages of the evaluation process.   The evaluator can also
help project personnel to understand technical issues associated with
testing, diagnosis, and program design.

An independent evaluator may be able to point out instances in which
the relationship between program objectives and program activities is
tenuous or unreasonable, a relationship perhaps difficult for program
personnel to observe easily.   In this role, the evaluator can be used
as a sounding board to determine whether there is a logical and close
connection between what the program intends to accomplish and what the
program is in fact doing.   This logical nexus between program goals

and program activities will provide the most convincing evidence that the program is responsible for the outcomes observed.

Listed below are some guidelines for distinguishing between the roles of the program director and evaluator. These guidelines take into consideration the fact that the majority of the evaluation activities will actually be conducted by the program director and program personnel.

The program director should:

o   Define program goals and objectives;

o   Describe the intended program;

o   Describe student characteristics;

o   Identify target audiences for the evaluation;

o   Determine the major areas to be covered by the evaluation;

o   Identify possible evaluators, and in some cases, select the evaluator(s) or at least recommend the evaluator(s);

o   Serve as a liaison with the evaluator (or appoint a staff member to serve as liaison;

o   Review the evaluation design prepared by the evaluator to make sure it meets the evaluation needs;

o   Arrange interviews or write cover letters to questionnaires to ensure timely response and cooperation;

o   Monitor classroom operations and observation activities;

o   Assign specific evaluation activities to program personnel;

o   Identify trained personnel (and suggest specific
        persons) who should be involved in data analysis
        and interpretation; and

o   Review data and react to interpretations and
        recommendations before they are included in the
        report.

The evaluator should:

o   Design the evaluation based on the information
        needs identified by the program director;

o   Select and/or review instruments to be used in the
        evaluation;

o   Monitor testing; and

o   Analyze the data and report findings.

A clear delineation of responsibilities and responsible management
will ensure that all evaluation activities are performed effectively
and on schedule.

2.   Determine the Audience and What to Evaluate

Determining which components of the program to evaluate is obviously a
most critical decision. However, this decision is always influenced
by the different parties involved with the bilingual program.
Consequently, the decision of what to evaluate is largely determined
by the evaluation needs of these parties, as well as the financial and
human resources available to conduct the evaluation.

Thus, the first step in determining what to evaluate is to determine
who needs information from the evaluation, what type of information is
needed, and for what purposes. In addition to program administrators

and other personnel associated with the program, the typical users of evaluation information include:

o   The funding agency;

o   District administrators;

o   School board; and

o   Parents and community groups.

Each audience has different interests and needs. Therefore, the evaluation design must address the different needs of each audience and provide the information desired, while remaining within the budgetary constraints of the program.

Evaluations of ESEA Title VII funded programs, however, must pay particular attention to the rules and regulations pertaining to these programs. Embodied in these rules and regulations are a number of provisions that should be viewed as minimum evaluation criteria. Therefore the evaluation requirements for basic and demonstration projects, as described in section 123a.22 of the April 4, 1980 Federal Register (Vol. 45, No. 67), must be considered in planning the evaluation.

These regulations require that any program funded under Title VII of the Elementary and Secondary Education Act (ESEA) of 1968, as amended, must have a plan to evaluate the progress and achievements of the bilingual program. The plan must include:

o   provisions for measuring the accomplishment of the
     instructional objectives of the program;

o   provisions for measuring the students' progress in
         improving their English language skills; and

o   a procedure for using the information gained from
         the evaluation to improve the operation of the
         program.

Worksheet No. 1, which follows, is a useful tool to use to identify

the various audiences that need information from the evaluation, the

type of information needed by each audience, the reason that the

information is needed, and the time at which the information is

needed.   The Worksheet is also designed to help the program director

to plan and prepare for reporting the evaluation results to all the

audiences.   After filling out all the information required on this

Worksheet, the program director can determine the comprehensiveness

and depth that the evaluation will require, as well as how many of the

evaluation needs can be met with the resources available.

If the resources available will not permit the evaluation to assess

all the issues or program areas, the program director, evaluator, and

all other parties concerned will have to set priorities for the

evaluation.   This will inevitably require that concessions and

compromises be made by all parties.

How to Use Worksheet No. 1 -- In the first column, indicate the group

or groups of people who will need information from the evaluation and

who will receive the evaluation report in whole or in part (funding

agency, district administrators, school board, etc.).   A brief

statement indicating the type of information needed by each audience

should be written in the space provided under column two.   Indicate

under column three why the information is needed. This statement will, to a large extent, determine the type of report (column five) needed by this particular group (oral, written, executive summary, etc.). The statement will also determine which section of the report should be emphasized in the cover letter. For example, if the school board is the intended audience and it is trying to determine program impact on English language development, the cover letter should emphasize the section on student outcomes. Column four indicates when the information is needed in order to provide adequate time for the audience to react to the report.

This Worksheet, when used properly, provides a global picture of the evaluation and helps pinpoint the types of information that need to be collected during the course of the evaluation. It also helps to determine what the evaluation report will contain. It further helps to specify those points at which feedback on the evaluation report, in draft form, must be sought prior to producing the final version.

DETERMINE AUDIENCE AND INFORMATION REQUIREMENTS FOR THE EVALUATION

Audience	Type of Information Needed	Reason Information is Needed	Date Information is Needed	Type of Report and Section to Emphasize in Cover Letter

117

3. Set Priorities and Establish Timelines

The establishment of evaluation priorities is a must for all bilingual programs. Most bilingual programs allocate $3-5,000 of their budgets for the purchase of outside consulting assistance to perform the evaluation. This amount of money, together with the level of effort that can be devoted to this one task by the program director and the rest of the program personnel, constitute the available resources to conduct the evaluation. More than likely, the evaluation needs identified by the exercise described above will far exceed what can be accomplished by these resources. Consequently, priorities for the evaluation will have to be established.

The program director must analyze the evaluation needs identified and ask the following questions. Based on the information identified earlier, and the known resources:

o   How much can I evaluate?

o   How much do I need to evaluate?

o   How much evaluation assistance can I afford?

Answers to these and similar questions will assist in prioritizing the different elements of the program to be evaluated.

Additional questions, such as the ones below, will also help to determine priorities:

o   Is information on the program's capacity to meet
     Title VII regulations already available? If

information is available, this information can be
easily incorporated in the evaluation.

o   What are the priority areas (e.g. parent
    involvement) of the program? The evaluation
    effort should give these areas priority.

o   How are the program resources divided among
    program components? Areas receiving a large
    proportion of program resources should be
    candidates for evaluation emphasis.

o   If there are insufficient resources to adequately
    evaluate all components, are there areas that
    should not be evaluated or should the scale of the
    evaluation be reduced in some or all areas? This
    decision would be made after considering which
    areas are already fairly well understood, which
    areas are a low program priority, and whether the
    evaluation resources are so limited that it would
    be best not to evaluate them at all rather than to
    conduct a general assessment of all areas.

o   Which components must be evaluated each year?

Answering some or all of these questions will assist the program

director to determine what __must__ and __can__ be evaluated. Still another

exercise to help in the priority-setting process is to break down all

the different elements of the total program and prioritize each

element of the program based on the information the audience needs, as

well as the Title VII requirements. After prioritizing all the

program elements according to this criteria, a final priority listing

can be developed based on the amount of evaluation activities that can

be performed with the available financial and human resources. This

determination is based on the estimated level of effort that each task

of the evaluation will require. Estimating level of effort is

discussed in the next section.

How to Use Worksheet No. 2 -- This Worksheet is designed to assist the person planning the evaluation to gather information in order to establish priorities. This form should be completed by the program director and discussed with the evaluator. The evaluator can also use the Worksheet as a general guide for developing the evaluation design.

Depending on the answers to the questions above, the program director and evaluator will be able to·prioritize the different evaluation needs. To use the Worksheet, insert a "1" by components which should receive maximum emphasis, a "2" by those receiving moderate emphasis, a "3" by the components that will receive minimum emphasis, and an "X" by the components not to be evaluated. After completing Worksheet No. 2, the program director and the evaluator should review the information to ensure that priorities set by funding agencies as well as priorities of the program are adequately represented.

As noted earlier, planning a useful evaluation involves a careful balancing of priorities and a sensible allocation of resources. It is unlikely that an evaluation effort can address all possible components and issues in any one year. Furthermore, it is more important to do a thorough evaluation of the most important parts of the program than to do a general evaluation of all program components. Therefore, it is important that priorities be set intentionally rather than arbitrarily. A program component should not be omitted from the evaluation because of oversight oi because the evaluation resources were exhausted before that component could be addressed.

SETTING PRIORITIES

Put a "1" by components which will receive a maximum emphasis, a "2" by components receiving moderate emphasis, a "3" by components receiving minimum emphasis, and an "x" by components which will not be evaluated.

Evaluation Components	Done last year (19__)	This Year (19__)	Next Year (19__)	Following Year (19__)
A. Program Description Information				
1. Project Overview	___	___	___	___
2. Instructional Approach	___	___	___	___
3. Project Management	___	___	___	___
B. Program Operations				
1. Instructional Program Implementation	___	___	___	___
2. Staff Development	___	___	___	___
3. Parent Involvement	___	___	___	___
C. Student Effects				
1. English Language Component	___	___	___	___
2. NonEnglish Language Component	___	___	___	___
3. NonEnglish Academic Component	___	___	___	___
4. Nonacademic Student Effects	___	___	___	___

11-15

After the evaluation priorities have been determined, the program director should establish timelines for completing the different components and the total evaluation of the program. The program director needs to understand that certain elements of the evaluation must be performed at very specific times during the academic year and cannot be delayed or postponed. Additionally, the program director, in determining timelines for tasks assigned to specific individuals, has to consider other responsibilities of evaluation team members. Responsiblities and assignments may have to be modified as a result of the established timelines.

How to Use Worksheet No. 3 -- This Worksheet can be used as a bar chart to depict all evaluation activities. In the space provided under the Task heading, indicate all the major tasks and subtasks required to perform each of the evaluation activities. This exercise will help the planner to "think through" all of the steps required. The months of the "school year" are then depicted next to the tasks. To use the bar chart, simply place a line through whatever period of time each task will require. This activity will force the planner to determine which activity must be performed at what time.

A sample Worksheet, already filled out, is attached. A blank one is included in the Technical Appendix.

122

TIMETABLE FOR EVALUATION ACTIVITIES

Year_____

Tasks	MONTHS
	Aug| Sep| Oct| Nov| Dec| Jan| Feb| Mar| Apr| May| Jun| July

A. Plan Evaluation Design*
   1. Determine which goals and objectives in each component to focus on
   2. Cost out evaluation
   3. Summarize design for administrator

B. Project Description
   1. Collect data - divide up
   2. Summarize data
   3. Review & analyze data for purposes of planning its use in analyzing evaluation data

C. Monitoring of Program Operations
   1. Instructional Program Implementation
      a. Develop/select instruments
      b. Administer instruments
      c. Analyze data
      d. Interpret data
      e. Draft Report Section

   2. Staff Training
      a. Develop/select instruments
      b. Administer instruments
      c. Analyze data
      d. Interpret data
      e. Prepare report section

   3. Parent Involvement
      a. Develop instruments
      b. Administer instruments
      c. Analyze data
      d. Interpret data
      e. Draft report section

*Last possible time to do this. Ideally this would also be done the previous spring.

Year _____

Tasks	MONTHS											
	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	July

D. Evaluation of Language Components
  1. Develop/select instruments
  2. Administer instruments
  3. Analyze data
  4. Interpret data
  5. Draft report section

E. Evaluation of Non-language Academic Components
  1. Select instruments
  2. Administer instruments
  3. Analyze data
  4. Interpret data
  5. Draft report section

F. Evaluation of Non-academic Components
  1. Develop/select instruments
  2. Administer instruments
  3. Analyze data
  4. Interpret data
  5. Draft report section

G. Report
  a. Compile report sections
  b. Review report
  c. Prepare final report

**Partial analysis interpretation and reporting is done at this point.

4. Determine Level of Effort, Budget and Allocate Resources

One of the most difficult tasks in managing the overall evaluation is deciding how best to utilize the limited resources available, and yet meet all the evaluation needs. The assignment of responsibilities and activities to those contributing to the evaluation process is often difficult. Because most of the evaluation activities pertaining to Title VII programs are usually performed by the program personnel, coordinating time schedules to permit evaluation in addition to other program responsibilities can create great problems, especially if human and financial resources are limited. Nevertheless, the timely execution of the evaluation is essential. There are activities within the evaluation process that can be rescheduled; however, others must be performed as planned in order to produce a reliable product. The effective program director must exercise initiative and resourcefulness to ensure that this is accomplished.

How to Use Worksheet No. 4 -- The effective program director uses as many tools as possible. Worksheet No. 4, the Operating Checklist for Bilingual Education Program Evaluation, which follows, can be used as a checklist to ensure that the evaluation plan contains all the elements needed and that they are initiated and successfully completed.

How to Use Worksheet No. 5 -- Worksheet No. 5, the Evaluation Summary Guide, summarizes all evaluation activities by program component for easy monitoring of the entire evaluation. By using this checklist,

OPERATING CHECKLIST FOR BILINGUAL EDUCATION
PROGRAM EVALUATION

EVALUATION STEPS	Initiated	Completed
1. Planning, Managing, and Staffing the Evaluation		
1.1 Determination of audience for the evaluation		
1.2 Determine the focus of the evaluation		
1.3 Allocation of resources for evaluation activities		
1.4 Setting timelines for evaluation activities		
1.5 Develop overall management plan of evaluation		
1.6 Hire outside evaluator		
1.7 Assigning evaluation responsibilities to staff		
2. Planning Data Collection for the Evaluation		
2.1 Description of program		
2.2 Description of students		
2.3 Description of program's goals		
3. Planning Monitoring of Program Operations		
3.1 Description of program in operation		
3.2 Description of staff development activities		
3.3 Description of parent involvement		
4. Planning Evaluation of Student Outcomes		
4.1 Selection of evaluation questions		
4.2 Selection of evaluation design for English, non-English, and other areas		
4.3 Selection of assessment instruments		

	Initiated	Completed

4.4 Scheduling the testing for the evaluation

4.5 Designing procedures and scheduling data
     collection

4.6 Planning the analysis of the data

4.7 Reporting the results

5. Reporting the Results and Writing the Evaluation
    Report

5.1 Identification of audiences and reporting
     requirements

5.2 Establishing timelines

5.3 Outline for report

5.4 Analysis of the data

5.5 Selection of convening the interpretative
     panel for analyzing the data

5.6 Writing the evaluation report and planning
     presentations of results

EVALUATION SUMMARY GUIDE

Evaluation Questions	Evaluation Instruments	Source of Information	Data Collection		Data Analysis		Data Interpretation		Reporting	
			Who does it	When	Who	When	Who	When	Who	When
Program Description										
Monitoring Program Operations										
Student Outcomes										

the program director can easily monitor activities, coordinate time schedules and assignments, and continue to plan and make appropriate modifications in the management of the evaluation. All members of the evaluation team should have a copy of Worksheet No. 5, so that the entire team has an understanding of the evaluation process, the role of each person in performing the evaluation, and the deadline for each evaluation activity.

Determining how much of the evaluation should be conducted by program personnel, which activities should be performed by an independent contractor, and how much the total evaluation should cost is often difficult for many program directors. Districts with limited contract evaluation funds should use most of their contract funds to employ a trained and experienced evaluator to assist them in evaluating the student outcomes component of the evaluation. Other evaluation tasks, such as describing and monitoring program operation, can be performed by the program director with assistance from the program personnel as a normal part of program management. However, the evaluator should be consulted when performing these tasks. If project or district personnel are going to be employed to perform the evaluation, the program director must make specific assignments and ensure that the evaluation activities are performed on schedule.

A major step in planning and managing the evaluation, therefore, is determining the level of effort that will be required by each activity of the evaluation (e.g., evaluating student outcomes) and allocating adequate financial and human resources to the individual tasks to be

performed. Evaluation resources, financial and human, will vary
widely from district to district. Additionally, the level of effort
for an evaluation is affected by a number of factors, such as:

o   Size of the program;

o   What aspects of the program are evaluated;

o   The number of non-English languages represented in
     the population being served by the program; and

o   The scope of the evaluation.

The Estimated Level of Effort Worksheet (Worksheet No. 6) may be used
to estimate the amount of effort which the evaluation will require.

The Worksheet suggests three different estimated levels of effort that
can be applied in evaluating each program component and the different
tasks within each component. These estimates are based on discussion
with persons who have conducted these evaluation activities. The
three levels are defined as minimum, moderate, and major. The amount
of evaluation activity that can be performed using the minimum level
of effort may not provide adequate data for local use, but will most
likely satisfy evaluation requirements of the funding agency. The
amount of effort indicated for the moderate and major categories
represents more realistic estimates of the effort required to perform
an adequate evaluation of each program component. The major level
category does not include all of the possible evaluation activities
that could be included; rather, it establishes a level for a set of
activities which will provide adequate data for most programs. Using
worksheet No. 6, the program director can select the desired level of
effort for each component.

Using this worksheet, the costs associated with the evaluation are easily identifiable. The hourly cost of district and program personnel, including support staff, are known to the finance office. The number of hours that will be dedicated to the effort by each person multiplied by their hourly wage rate determines the direct cost to the program. This assumes no overhead for the district. Also, in some districts, trained evaluation specialists may be available at no cost to the program.

The summary section of Worksheet No. 6 enables the program director to summarize the level of effort required to evaluate each program component. It also summarizes the level of effort which will be assigned to district or program personnel and to the evaluator. After reviewing the summary and total level of effort required, the program director can decide whether the evaluation, as planned, is affordable. If not, decisions will have to be made to either streamline the evaluation effort or to seek additional resources.

A decision on the level of effort to be assigned to the independent evaluator or to the consulting firm, assuming the contract route was employed, will have to be made as early as possible. Once a contractual obligation is entered into, the district will be liable for meeting that contract. Using the worksheet, the program director can determine which of the evaluation activities will be performed by the evaluator or when s(he) will provide assistance. Adding up the total number of days that the evaluator will provide and multiplying this total by the daily rate of the evaluator will determine the cost of the service.

The costs for independent evaluation consultants are usually standard, but do vary if obtained through a consulting company. The following figures may be used to estimate their costs.

- o   Independent evaluator
      (no overhead) $100-$150 per day

- o   Evaluator contracted through
      evaluation company
      (overhead included) $250-$300 per day

- o   Senior evaluator contracted
      through major educational
      research company $300-$400 per day

By using the worksheet, the program director will be able to determine what the services of the evaluator will cost and if all the work that needs to be performed by the evaluator is within the budget allocation. When summarizing costs for the evaluation of each component, use actual cost rates. Other cost items will include purchasing tests, computer time, and report preparation.

How to Use Worksheet No. 6 -- Worksheet No. 6 may be used to estimate the number of days needed to complete each component of the evaluation. The recommended levels may be used or the program director may wish to make his or her own estimate. The number of days assigned to each task of the evaluation to be provided by program or district personnel should be circled in order to clearly differentiate the days to be provided by the evaluator.

This form should be completed by the program director and evaluator after the evaluator is selected. Ideally, Worksheet No. 6 should be

completed individually by the program director and by the evaluator in order to compare time allocation estimates. From the individual estimates, the director and evaluator should prepare a final allocation of level of effort to each task which should serve as the management tool to guide the elevation process. The worksheet provides a summary of the level of effort and cost estimates for all components of the evaluation. This worksheet can also be used to obtain bids from external evaluators.

ESTIMATING LEVEL OF EFFORT REQUIREMENTS
FOR
DESCRIBING THE PROGRAM AND THE STUDENTS

Estimates are provided for three levels of evaluation activity for a given

year: (Different activity levels may occur each year).

a) Minimum - collect information from project proposal, school records,
and project director.

b) Moderate - collect information from project proposal, school records,
project director, and a sample (one to three people in each category)
of project staff, bilingual teachers, district administrators and
parents using structured interviews or questionnaires (for estimation
purposes below, assume total number of people interviewed or
receiving a questionnaire is eight).

c) Major - same as that described for "moderate," except more people in
each category are interviewed or sent questionnaires plus classroom
observations are conducted. (For estimation purposes below, assume
the total number of people interviewed or receiving questionnaires is
fifteen and that three classrooms are observed).

Task	Minimum	Moderate	Major	Your Estimate*
Level of Effort for a Given Year (in Days)				
1. Prepare, discuss with and obtain support of project director for proposed plan	½	½	½	_____
2. Prepare data collection instruments (using samples provided in Designer's Manual)	½	1	2	_____
3. Identify specific people or records from whom to collect data and make arrangements	1	3	5	_____
4. Collect data	1½	5	12	_____
5. Analyze and organize data for use in report and analysis of evaluation data collected for later components	2	4	6	_____
Total Days	(5½)	(13½)	(25½)	(_____) Evaluator
				(_____) Project Staff

* Circle the estimate for any tasks which will be done by project staff
instead of the external evaluator. Do not include these amounts in
the total for the evaluator.

ESTIMATING LEVEL OF EFFORT REQUIREMENTS
FOR
EVALUATING PROGRAM OPERATIONS

Estimates are provided for two levels of activity to be conducted during

a given year for each of three components - instructional methods, staff

development, parent involvement (Different levels of activity may occur

each year):

Instructional Methods

    a) Minimum - Conduct observations and interviews twice/year in only
               two classrooms and have evaluator do interpretation.*

    b) Major -   Conduct observations and interviews three times/year in
               all classrooms (for estimation purposes below, assume
               total number of classrooms equal five) and have inter-
               pretative panel.*

Staff Training

    a) Minimum - Same questionnaire given to trainees following each train-
               ing session. Knowledge test not used and evaluator does
               interpretation. (For estimation purposes below, assume
               fifteen trainees and three training sessions).

    b) Major -   Same as for minimum, plus a knowledge test given pre and
               post training, an end of project summary questionnaire
               given and an interpretative panel is used. (For estima-
               tion purposes below, assume fifteen trainees and three
               training sessions).

Parent Involvement

    a) Minimum - Address only the issue of the extent to which the level of
               parent involvement matched the planned level); evaluator
               interprets data.

    b) Major -   Address all four proposed evaluation questions given on
               page 61. For estimation purposes below, assume ten
               parents and eight staff members interviewed); have inter-
               pretative panel.

---

* The alternative methods of interpreting the data are discussed in the
    staffing chapter which follows.

Task	Level of Effort (in Days)		Your Estimate*
	Minimum	Major	

Instructional Method

1. Prepare, discuss with and obtain support of project director for proposed plan	1	1	_____
2. Prepare data collection instruments (using samples provided in Designer's Manual)	1	2	_____
3. Identify who to observe and interview and make arrangements to do so	1	1	_____
4. Collect data	½	5	_____
5. Analyze data	1	5	_____
6. Interpret data	1	3	_____
7. Write report section	1	2	_____

Total days	(6½)	(20)	(_____) Evaluator
			(_____) Project Staff

Staff Training

1. Prepare, discuss with and obtain support of project director for proposed plan	½	½	_____
2. Prepare data collection instruments (using samples provided in Designer's Manual)	½	6	_____
3. Make arrangements for data collection	½	½	_____
4. Collect data - minimum (have trainer collect all data); major (have trainer collect all data except end of year questionnaire)	0	1½	_____

* Circle the estimate for any tasks which will be done by project staff instead of the external evaluator. Do not include these amounts in the total for the evaluator.

Task	Level of Effort (in Days)		
	Minimum	Major	Your Estimate
5. Analyze data	1½	7	_____
6. Interpret data and develop recommendations	;	3	_____
7. Write report section	;	1½	_____
Total days	(5)	(19)	(___) Evaluator
			(___) Project Staff

Parent Involvement

1. Prepare, discuss with and obtain support of project director for proposed plan	½	½	_____
2. Prepare data collection instruments (using samples provided in Designar's Manual)	½	1	_____
3. Make arrangements for data collection	½	1	_____
4. Collect data	½	6	_____
5. Analyze data	1.	3	_____
6. Interpret data and develop recommendations	½	2	_____
7. Write report section	1	2	_____
Total days	(4)	(15½)	(___) Evaluator
		O	(___) Project Staff

ESTIMATING LEVEL OF EFFORT REQUIREMENTS
FOR
EVALUATING STUDENT OUTCOMES

Estimates are provided for two levels of activity to be conducted during a
given year for each of four components—English language component, nonEnglish
language component, nonlanguage academic component, and nonacademic student
effects.

<u>English Language Component</u>

   a)   Minimum - Use norm-referenced evaluation design only; analyze by
       grade, subject. language used in instruction, and student
       proficiency; evaluator does interpretation.

   b)   Major   - Use time series, norm-referenced and comparison group
       evaluation designs; analyze by grade, subject, language
       used in instruction, student proficiency factors; use
       interpretative panel..

<u>NonEnglish Language Component</u>

   a)   Minimum - Use existing test and do norm-referenced evaluation design
       only; analyze by grade, subject, language used and student
       proficiency; evaluator.does interpretation.

   b)   Major   - Develop own test; use time series, norm-referenced and
       comparison designs; analyze by grade, subject, language
       used in instruction and student proficiency; use inter-
       pretative panel.

<u>Nonlanguage Academic Component</u>

   a)   Minimum - Use existing test, compare to national norms; analyze only
       by grade; evaluator does interpretation.

   b)   Major   - Use existing test, compare to national norms; analyze by
       grade and two other key factors; use interpretative panel.

<u>Nonacademic Student Effects</u>

   a)   Minimum - Use only a published self concept measure; analyze only by
       grade and student proficiency; evaluator does interpretation.

   b)   Major   - Use all proposed evaluation questions and data collection
       instruments; analyze by grade and student proficiency; use
       interpretative panel.

Task	Level of Effort (in Days)		
	Minimum	Major	Your Estimate*

English Language Component

Task	Minimum	Major	Your Estimate*
1. Prepare, discuss with and obtain support of project director for proposed plan	½	½	_____
2. Select appropriate tests	1	5	_____
3. Train test administrators and make arrangements for testing	1	2	_____
4. Supervise testing - minimum (one day each, pre- and post-testing); major (monitor all testing)	2	14+	_____
5. Analyze data - minimum (prepare achievement data for standard computer analysis); major (prepare data for standard computer analysis, for several analyses)	3	8+	_____
6. Interpret results	2	10	_____
7. Write report section	3	10+	_____
Total days	(12½)	(50+)	(_____) Evaluator
			(_____) Project Staff

NonEnglish Language Component

Task	Minimum	Major	Your Estimate*
1. Prepare, discuss with and obtain support of project director for proposed plan	½	½	_____
2. Select appropriate tests	1	5+	_____
3. Train test administrators and make arrangements for testing	1	2	_____

* Circle estimate for any tasks which will be done by project staff instead of the external evaluator. Do not include these amounts in the total for the evaluator.

Task	Level of Effort (in Days)		
	Minimum	Major	Your Estimate*
4. Supervise testing - minimum (one day each, pre- and post-testing); major (monitor all testing)	2	10+	_____
5. Analyze data - minimum (pre-pare achievement data for standard computer analysis); major - (prepare data for standard computer analysis for several analyses)	2	8	_____
6. Interpret results	2	10	_____
7. Write report section	2	10	_____
Total days	(10½)	(45½)	(_____) Evaluator
			(_____) Project Staff

Nonlanguage Academic Component

1. Prepare, discuss with and obtain support from project director for proposed plan	½	½	_____
2. Select appropriate tests - minimum (become familiar with district tests); major (review commercial achievement tests and match to curriculum)	1	5	_____
3. Train test administrators and make arrangements for testing	1	2	_____
4. Supervise testing - minumum (one day each, pre- and post-testing)	2	10+	_____
5. Analyze data - minimum prepare achievement data for standard computer analysis); major (prepare data for stand-ard computer analysis for several analyses)	2	8	_____

Task	Level of Effort (in Days)		
	Minimum	Major	Your Estimate*
5. Interpret results	2	10	_____
7. Write report section	2	8	_____
Total days	(10½)	(45+)	(_____) Evaluator
			(_____) Project Staff

Nonacademic Component

Task	Minimum	Major	Your Estimate
1. Prepare, discuss with and obtain support from project director for proposed plan	½	½	_____
2. Select or develop appropriate instruments	1	4	_____
3. Train test administrators and make arrangements for testing and other data collection	1	4	_____
4. Analyze data - minimum (prepare for standard computer analysis)	2	8	_____
5. Interpret results	2	10	_____
6. Write report section	2	8	_____
Total days	(8½)	(34½)	(_____) Evaluator
			(_____) Project Staff

SUMMARY OF ESTIMATED LEVEL OF EFFORT
REQUIREMENTS AND
ASSOCIATED COSTS

Summary of Days	Evaluator	Project Staff
Program Description	_____	_____
Monitoring Program Operations	_____	_____
Instructional Methods	_____	_____
Staff Training	_____	_____
Parent Involvement	_____	_____
Evaluating Student Effects		
English Language Component	_____	_____
NonEnglish Language Component	_____	_____
Nonlanguage Academic Component	_____	_____
Nonlanguage Student Effects	_____	_____
	(_____)	(_____)

Total days (evaluator cost per day = Total evaluator cost per year

_____ X _____ = _____

Additional Cost Items	Program Description	Costs (in Dollars) Monitoring Program Operations	Evaluating Student Effects
1. Secretary time	_____	_____	_____
2. Printing	_____	_____	_____
3. Mailing	_____	_____	_____
4. Other			
a._____	_____	_____	_____
b._____	_____	_____	_____
c._____	_____	_____	_____
d._____	_____	_____	_____
e._____	_____	_____	_____
Totals	_____	_____	_____
Total Evaluator Costs	_____	_____	_____
Total Additional Costs	_____	_____	_____
Total Costs for Evaluation	_____	_____	_____

5. Plan the Data Analysis Function

The program director and evalustor should plan the specific data
analysis activities that will be required by the evaluation. The type
of analysis and techniques to be used will depend largely on the types
of data collected. Data from the first facet of the evaluation will
consist primarily of narrative descriptions of program operations, as
well as responses from the interviews collected. Data from the second
facet of the evaluation will be primarily in the form of test scores.

The data analysis required by the recommended evaluation model is
straighforward and relatively easy to perform. Analysis data from
program operations data is analyzed by simply comparing two sets of
similar data. One set describes the program as it was intended to
operate, while the other describes how the program is actually
operating. Therefore, the only analysis required is to exaine the
information and determine if there is a difference in the two sets of
data. The analysis of student outcome-data is somewhat more
technical, but can be performed by a trained evaluator. The analysis
procedures are usually found written the test manual supplied with the
test. These procedures are usually easy to perform.

The two types of data are analyzed separately and are intertwined only
through the efforts of a perceptive evaluator. The two types of data
can stand alone and do not need to be integrated. However, a
perceptive evaluator will be able to see how the two types of data can
be used to support each other.

114

The important consideration during the planning stage is to determine how the analysis function will be conducted. Data analysis will most probably be performed by the evaluator. The time schedule for the evaluation should allow ample time to conduct the analyses.

6. Plan the Data Interpretation Function

Data interpretation in bilingual program evaluation is often not a strictly empirical task. To repeat the basic premise of this Handbook, it is probably impossible to show that children in the bilingual program did better in the program than they would have without it by employing conventional social science research methods. Therefore, interpreting the data obtained by evaluation efforts is not a mechanical exercise of reciting significant alphas. Rather than concluding that the bilingual program "works" better than some alternate treatment, the interpretive exercise is more likely to be in the nature of a policy question. Does the bilingual program "work" well enough? Are decisionmakers and constituents satisfied with the program and the student's progress? Recognizing the policy implication function of data interpretation, an interpretive panel may be a better alternative to perform this function. Chapter IV provides a more detailed discussion on the interpretation function.

Therefore, an important step in the evaluation is how the interpretation function is accomplished. An evaluation may be technically sound and well conducted, but may fail to be used by

decisionmakers because appropriate people were not involved in interpreting the data and in the development of recommendations.

Two basic approaches are suggested for data interpretation and formulating recommendations for program modification. The first approach is for the evaluator to analyze, study, and interpret the results. Using informal means, the evaluator then checks the interpretations and recommendations with program staff and others as he/she deems appropriate. The second approach is to convene a panel of people with various perspectives on the program and have them interpret the results. The panel may consist of individuals that are representative of the various audiences.

7. Plan the Reporting of the Evaluation

Preparation of the final evaluation report is an important activity of the evaluation. The evaluation report is the final and most visible product of the evaluation. Steps should be taken to assure that the report addresses the purposes and specific questions of the decisionmakers for whom the evaluation was planned. In addition, the evaluation results should be reported in a timely manner, taking care to ensure that the technical aspects of the evaluation effort are clearly presented. Together, these steps increase the usefulness of the evaluation results.

Preparation of the final evaluation report can be a time-consuming and burdensome process if not properly planned. However, reporting should

116

be a continual process or activity that occurs throughout the evaluation cycle. For example, Chapter III will recommend that following each classroom observation, a brief report should be prepared. These brief reports should be summarized at least three times during the program year--fall, winter, and spring -- and should be shared with program personnel so that they can become part of the program improvement process. Thus, these brief reports and summaries prepared throughout the evaluation cycle will all feed into the final evaluation report, thus simplifying the reporting process. The preparation and sharing of evaluation information throughout the evaluation cycle also serves to strengthen communication between the evaluation audiences and those conducting the evaluation, thereby increasing the use of evaluation results.

There are a number of basic principles which pertain to the reporting process and serve to simplify preparation of the final evaluation report. This discussion assumes that completion of the report is the primary responsibility of the program evaluator(s) contracted to undertake major segments of the bilingual program evaluation. Basically, the evaluator has three important tasks: develop an understanding of the audiences who will use the information, select a proper reporting format(s), and assist the audiences in using the results. Proper planning of the reporting requirements will make this final activity easy to complete.

Several standard elements should be included in the report. These include:

o   Statement of purpose;

o   Program overview and background;

o   The goals and objectives of the bilingual
     program;

o   Description of the program and students;

o   Discussion of the methodology used; including
     design, sampling strategy, instrumentation, and
     data analysis procedures; and

o   Presentation of the findings, conclusions, and
     recommendations for program change.

The report should be concise and should include easily interpreted

tables, graphs, and other figures limiting the amount of narrative

material presented.   Important issues should be identified and

highlighted in the report if the results of the evaluation effort are

to be maximized.   Techniques such as boxing in recommendations or

using a different type face are useful to highlight the most important

points of the report.   Examples of actual data collection instruments

should be included in an appendix.   Chapter V provides more detailed

guidelines for developing the report.

CHAPTER II

ESTABLISHING BASELINE DATA REQUIRED FOR THE EVALUATION

The evaluation model for evaluating Title VII bilingual education programs presented in this Handbook has two components. The first component evaluates program operations (e.g. program administration, staff development, parental involvement, etc.) using a discrepancy evaluation design. The second evaluates student outcomes. Results of these two evaluation activities taken together constitute the basis for determining how the program operated and provides a description of student performance.

In order to conduct the discrepancy evaluation of program operations, information on how the program was originally designed and intended to operate must be collected and documented. This information serves as the baseline data, which are compared to the data resulting from the actual evaluation of program operations as described in Chapter III.

The baseline data are also taken into account in developing the student outcomes evaluation design for the student outcomes component of the evaluation. Therefore, a very early and important step in conducting an evaluation of a bilingual program is the establishment of baseline information about the total program.

This description identifies who the program is meant to serve, the exact services of the program, how these services are to be provided, and what outcomes are expected from the services. This description enables the evaluator to determine (a) whether the bilingual program meets the original intent, and (b) whether any marked achievements can reasonably be attributed to the program.

Comparing of the original program design, as described by the baseline data, to its actual operation, as determined by the evaluation of program operations, will indicate areas of the program that have either not been implemented or have changed from the time that the program was originally designed. Discrepancies identified as a result of this comparison are a powerful management tool for the program director and a programmatically useful part of the whole evaluation process. This comparison can also help to determine whether the goals of the program are reasonable, and provide information about the relationship between program activities and program outcomes.

In order to accomplish this, the persons conducting the various evaluation activities must first develop proper documentation of the program context, the target students, the program goals, and the instructional program. This is not a difficult task. The information to be collected should clearly describe how the program is designed to meet its goals, as well as the total environment in which the program operates. Once this documentation is accomplished, the program director, with assistance from the evaluator, will be able to use the information to design the evaluation and to analyze and interpret the

evaluation results. The documentation does not need to be elaborate, simply informative. Most importantly, the information collected should be complete, detailed, and easy to understand.

Baseline Data Needed for the Evaluation

1. Describe the Context of the Program

Develop a simple, but accurate description of the school district and neighborhood. Data from previous evaluation reports can be easily updated, thus avoiding surveys or other time-consuming efforts. The type of information that should be covered in the description includes:

- o Community characteristics

    - Languages spoken
    - Ethnicity
    - Social economic status (SES) levels
    - Mobility and length of residence
    - Size

- o Local Education Agency (LEA) description

    - Size
    - Financial status
    - Facilities available for the bilingual program
    - General goals
    - Philosophy towards language and cultural diversity

o   School Description

  -    number of bilingual students by language
       group
  -    number in the bilingual program
  -·   how students are assigned to classroooms
  -    bilinguality mix in classrooms
  -    parent involvement in school affairs

The information collected on the program context should be compiled

immediately after the data-gathering phase. While technical analysis

of the information is not required, the program director and evaluator

should review the data in order to plan the program monitoring portion

of the evaluation and make preliminary decisions on how the data will

be used during analysis to determine program outcomes. The

information should be written in narrative form for inclusion in the

final report. The topics and subheadings provided above may serve as

an outline for reporting this information.

2. Describe the Students

Baseline information about the language proficiency and dominance,

cultural background, and overall academic achievement of the students

enrolled in the bilingual program is essential for designing and

conducting the evaluation. The data must include information on the

skill level of the students in both English and their home language,

as well as their level of performance in the subject areas being

taught. The description should also include information on the

student's learning background and school environment. At a minimum,

the baseline data must include information on the following areas:

o Definition of project student

o Student selection criteria & method

  - Tests & cut-off scores used

  - Role of teacher judgment

  - Role of parent wishes

  - Method of combining criteria

o Exit criteria & follow-up

o Student turnover

o Student characteristics at beginning of year

  - Language proficiency

  - Achievement level

  - Biographic data

This information is essential for grouping students according to both
current skills and past experience during data analysis activities and
plays a major role in determining student performance. For example,
a student with a low English reading pretest score might be expected
to show greater improvement if he or she were a new arrival from a
high SES background, and with no previous training in English reading,
than if he or she were from a low SES background and had been in a
bilingual program for several years.

A 'more accurate understanding of the evaluation results can be
obtained if the baseline data present a clear picture of the
environment and learning history of the students in the program.
Unfortunately, few programs collect this information during the
evaluation, and even fewer present a systematic treatment of this
information in evaluation reports.

Because most bilingual programs span several grade levels and are funded for a minimum of three years, bilingual programs should develop multi-year student profiles These multi-year profiles can increas the value of the student descriptions. Since most schools keep permanent student record files, the evaluator can easily make minor additions to the records each year to ensure that the appropriate background and information on services is readily available for each student in the program.

Many programs enroll substantial numbers of monolingual, native-English speakers, as well as students classified as limited-English-proficient (LEP), but who may be proficient in English. It is necessary to maintain the same amount of information on English language experience for these students as is required for non-proficient students. Knowledge of the different language levels of students in a class can be used to describe the linguistic environment of that class. Information on these students can be analyzed separately from that collected for students who are learning English as a second language to determine the effects of bilingual instruction on these students.

Information on the students should be compiled in narrative form for inclusion in the final report. This information differs from that collected in the previous section in two important ways: first, this information could change markedly from one year to the next (the information on the community may change but it is likely to be gradual), and second, information on the students can be modified by the program.

3. Describe the Program Goals

Developing a clear and complete description of the goals of the program is an essential part of establishing baseline data. Goal setting, although important, is often overlooked or ignored during the program planning stage. Therefore, many programs operate year-to-year with little or no set direction. Programs that fail to establish clear and measurable goals cannot expect to be able to measure program outcomes.

Most program goals are established to meet local, State, and possibly Federal guidelines in addition to other guidelines developed by parents and program personnel. Simply complying with these guidelines often determines the major goals and how they will be met. These goals, as well as those intended to meet local needs, should be included in the description. Also included should be a timetable for accomplishing the goals.

Programs should distinguish between short-term, intermediate goals relevant to a single-year evaluation and long-range goals that can be evaluated only over a period of several years. Failing to make this distinction creates problems for bilingual progra 3, since some long-term goals (e.g., improved English skills) may not be applicable and measurable until the later grades. Long-term goals are also affected by the high rate of student turnover experienced by many bilingual programs. Since long-term goals would not apply to a short-term student, two sets of goals are required. This should be clearly stated and presented in the baseline data being collected.

Defining and describing student achievement goals is another important step in establishing baseline data. While there are many important considerations to recognize when specifying student achievement goals, the baseline data must include information on:

- o Subject areas (e.g., reading, language, math);

- o Languages to be used (e.g., English, Spanish, etc.);

- o Student language proficiency category (e.g., English: limited or proficient, Spanish: limited or proficient);

- o Grade level; and

- o Student affective goals (e.g., self-concept and attitudes towards school).

Students who are exited from a bilingual program to a conventional classroom often require special follow-up services. Districts that provide such services should clearly specify and carefully describe how they are integrated into the goals of the program, along with other educational goals.

Because the original needs of the program, as stated in the proposal, may have changed, the information collected should be reviewed by the program director. Changes that have occurred should be properly documented.

A detailed description of the goals for each component of the project being evaluated -- e.g., program operations, parent involvement, staff development, and student effects -- should be developed. The baseline data collected should be used to finalize the evaluation

design and to ensure that each goal is appropriately measured by the evaluation activity. The information will also be used to interpret the evaluation results and make recommendations. The Final Evaluation Report should indicate if progress towards meeting the goals was measured, if the goals were met, and if not, what changes are necessary to ensure that the goals will be met, or what changes should be made in the goals. It is important to remember that not all goals need to be met in the current reporting period.

4. Describe the Instructional Program

Establishing baseline data for the instructional program requires more time and effort than any of the other three areas on which information is collected. Baseline data collection on the program context, students, and program goals basically requires the review of existing records, files, and the original project proposal. Baseline data collection for the instructional program, however, requires face-to-face interviews of persons associated with the program, as well as review of program documents.

A description of the instructional program can be divided into three categories:

o An overview of the program as it was originally designed and initially implemented;

o A description of the instructional approach used in the program, including (1) student selection, (2) self-concept and cultural emphasis, (3) content of instruction, (4) presentation of content, and (5) scheduling; and

o   Management of the program, including (1) staff
organization, (2) staff roles, (3) staff
development, (4) parent and community factors, (5)
communication links with different audiences, and
(6) dissemination of program information.

Thus, the description of the instructional program is the most
exhaustive of the activities associated with the establishment of
baseline data.

Information for the program overview can be collected easily from
information contained in the grant proposal. It should include the
grade levels and number of classrooms served by the program, the
amount of instructional time devoted to dual language instruction, and
a definition of the program design (maintenance, transitional, etc.).

A description of the actual instructional approach used in the
classroom and the basis for that approach require the most
comprehensive description of any part of the bilingual program. This
information can be collected from program related documents, student
records, classroom observations and interviews with program
administrators, teachers and parents. This description is also the
most important element used during the data analysis and
interpretation. It is therefore essential that program personnel pay
particular attention to this component. A partial listing of the
types of information to be collected follows. An expanded listing is
included in the Technical Appendix.

Descriptive Information on the Instructional Approach

1. Content of instruction

   a. Content areas covered
   b. Who determines content
   c. Other content features

      (1) Relationship of content to goals
      (2) Articulation of project content with existing
          district curriculum

2. Presentation of content

   a. Instructional approach

      (1) Type, e.g., concurrent, alternate day/week,
          preview/review half-day, resource room,
          and/or bilingual aide
      (2) Organizational practices, e.g., individualized,
          large group, learning centers, peer tutoring,
          small group instruction, and/or team teaching.

   b. Methodologies for bilingual education

      (1) Language of instruction

      (2) Approach to second language instruction

      (3) Approach to reading instruction

      (4) Approach to other academic instruction

4. Scheduling

   a. Grouping and regrouping

      (1) Across classes
      (2) Within classes

   b. Daily schedules

Identifying the goals of program services and describing variation in educational service is a very important part of this activity. In most programs, the services vary for different students depending on their language skills, reading and math skills, and other factors. In such cases, each different service must be described separately, and, when analyzing the evaluation data, students must be grouped according to the services they received.

In describing the bilingual program, it is essential to describe clearly what the students have experienced throughout their participation in the program. Therefore, a multi-year description of services should be developed. For example, bilingual programs that include a coordinated curriculum for grades K-6 must describe the complete program.

A description of the overall program organization and management is the last requirement of the baseline data collection activity. This description will provide the basis for evaluating the operational effectiveness of the program. The information should cover the following areas:

160

Descriptive Information on the Instructional Program

1. Staff Organization

    a. List of staff members and time commitment
    b. Organizational structure
    c. Qualifications
    d. Selection procedures

2. Responsibilities and Roles of Program Personnel

    a. Project Director
    b. Teachers
    c. Aides
    d. Other staff

3. Staff Development Program

    a. Needs assessment
    b. Structure of training
    c. Characteristics of Training
    d. Audiences Trained

4. Parents and Community

    a. Parent involvement in school affairs
    b. Community input in program planning, e.g., through
        advisory group.
    c. Community support for project
    d. Parent education
    e. Parent conferences/counseling

5. Communication

    a. Staff relations
    b. Relations with nonproject staff

6. Dissemination of project information

    a. School personnel
    b. Parents and community

5. Develop the Program Description

The amount of data to be collected will obviously vary from program to program. Once the evaluator, in consultation with the program director, makes the necessary decisions on what information to collect, the sources for the information should be identified and the proper data collection instruments selected and/or developed. Information on each program component will come from several sources. These sources may include:

o   The program proposal;

o   Student records and other files;

o   Previous evaluation;

o   The program director;

o   Program staff;

o   Bilingual teachers;

o   District administrators;

o   Classroom administrators;

o   Classroom activities; and

o   Parents.

Information from these sources is obtained by examining program documents and from interviews. Data collection should begin by September 15th of the first project year. Data should be updated by September 15th of each of the following years as needed to permit current analysis and interpretation.

162

Much of the information to describe the program context, goals, and management can be found in the grant proposal, prior evaluation reports, and other related documents. These sources may provide some information on the instructional program as well. Student records will provide information on the students characteristics, prior history, and performance. Worksheet No. 7, can be used to extract this information.

However, a significant amount of information will have to be collected from program personnel. Worksheets 8, 9, and 10 are sample interview schedules provided to assist the persons conducting the data collection activities. Worksheets 7, 8, 9, and 10 are included at the end of this narrative section.

How to Use Worksheets No. 7, 8, 9, and 10 -- These worksheets are designed to gather information from program documents, the program director, program staff, and local and district administrators. The person conducting these activities can interview the project director (Worksheet No. 8), program staff (Worksheet No. 9), and local and district administrators (Worksheet No. 10) and record their responses to questions that appear on these worksheets. These worksheets can be modified to meet the unique needs or focis of individual bilingual programs. Once all the interviews have been completed, the information should be synthesized to produce a document which provides a clear description of each of the four components of the intended program as originally described.

Each of the questions on these worksheets corresponds to one of the four program components (context, students, goals, and the instructional program). Each question is coded with a letter which identifies which component the question corresponds to. The coded letters are: C for program context; S for students; G for program goals; and P for instructional program.

Once the interviews have been conducted, the person conducting the interview can readily provide the information that describes that particular subsection of the program component.

6. Document and Report the Baseline Data

Once the desired information is collected, attention should be focused on the various ways it is to be used. The information:

o   Will be used as baseline information during the program monitoring activities of the evaluation process;

o   Will provide a partial basis for planning the analysis and interpretation of student outcomes, as described in Chapter IV; and

o   Will be reported directly to various audiences as part of the evaluation reports written for them.

Immediately after the preliminary data have been collected, the data should be summarized in the form that they will appear in the Final Evaluation Report and submitted to the program director for review. An initial analysis and interpretation of the data should be conducted to determine which variables, if any, are to be used as a basis for

194

separate analyses.   Chapters III and IV provide more detailed
information on how to conduct the data analysis and interpretation
function.

DATA COLLECTION FORM FOR INFORMATION FROM THE

PROJECT PROPOSAL AND OTHER RECORDS

The project proposal and various project or school records should be reviewed
to obtain the indicated information.

(G) 1. What are the major project goals?

Linguistically _____

_____

_____

Culturally _____

_____

_____

Academically _____

_____

(S) 2. What is the pattern of predominant languages among the student
population?

_____

_____

_____

(S) 3. What is the approximate achievement level (in languages, other
academic and nonacademic areas) of students within the various
language categories? Report separately for each language group.

Language achievement _____

_____

_____

Other academic achievement _____

_____

_____

C = refers to program context   G = refers to program goals
S = refers to program students P = refers to instructional programs

Nonacademic achievement_____

_____

_____

(P) 4. What grade levels and how many classrooms are served by the project?

_____

_____

_____

(P) 5. What portion of the school day is covered?_____

_____

_____

(C) 6. Describe the following community characteristics

a. Languages spoken (approximate percentage speaking each language)

_____

_____

b. Ethnicity (approximate percentage of each)_____

_____

c. Socioeconomic status (general description based on type of
   employment)

_____

_____

_____

d. Size of community_____

_____

_____

(C) 7. Describe the local education agency as follows:

a. Size_____

b. Financial status of district_____

_____

c. Facilities available for project_____

_____

_____

(C) 8. Describe the following school characteristics

a. Number of bilinguals in school by language group___._____

_____

_____

b. Number of students in bilingual program_____

_____

_____

c. Bilingual mix in the classrooms_____

_____

_____

(P) 9. Describe the project staff and its organization. List each member of the staff, the percentage of time committed to the project and their qualifications

Title	Name	Percentage time	Qualifications

_____

_____

_____

b. Describe the organizational structure of the project_____

_____

_____

_____

c. What selection procedures are used in selecting staff members?

_____

_____

_____

(P)   10. Describe the project director's role with respect to the following items:

a. Funds and budgets_____

_____

_____

b. Public relations_____

_____

_____

c. Administration_____

_____

_____

d. Overseeing instruction_____

_____

_____

e. Staff training_____

_____

_____

f. Developing and ordering materials and equipment_____

_____

_____

g. Staff recruiting and hiring_____

_____

_____

]

PROJECT DIRECTOR INTERVIEW SCHEDULE

(G)   1. The goals of the program as stated in the proposal are as follows:
(Present the goals orally or in writing as obtained from the pro-
posal.)

_____

_____

_____

What evidence will show that these goals have been met?_____

_____

_____

_____

Which goals have the highest priority?_____

_____

_____

_____

(G)   2. How would you define the project as to the extent which it is a
maintenance, transitional or partial bilingual program?

_____

_____

_____

(C)   3. Describe the mobility of the community including any specific data
available

_____

_____

_____

C = refers to program content
G = refers to program goals
S = refers to students
P = refers to instructional program

(P) 4. How are students assigned to classrooms?_____

_____

_____

_____

S; 5. Describe the student entry and exit criteria and procedures. Do the
       actual procedures conform to the planned procedures?

_____

_____

_____

(P) 6. Describe the scheduling of instruction including daily schedules and
       grouping and regrouping across and within classes

_____

_____

_____

(P) 7. Describe the staff and its organization in terms of the following
       dimensions

       a. Staff members' time commitments_____

_____

_____

       b. Staff organizational structure_____

_____

_____

       c. Staff qualifications_____

_____

_____

       d. Staff selection procedures_____

_____

_____

(P)   8. What is your general leadership style as program director?

_____

_____

_____

_____

_____

(P)   9. What is your role as program director with respect to each of the
         following areas?

a. Funds and budgets_____

_____

_____

b. Public relations_____

_____

_____

c. Administration_____

_____

_____

d. Overseeing instruction_____

_____

_____

e. Staff training_____

_____

_____

e. Developing and ordering materials and equipment_____

_____

_____

f. Staff recruiting and hiring_____

_____

_____

(P) 10. What is the teacher's role in the following areas?

a. Planning instruction_____

_____

_____

b. Implementing instruction_____

_____

_____

c. Noninstructional responsibilities_____

_____

_____

(P) 11. What is the role of the aides in the program?_____

_____

_____

_____

(P) 12. What is the role of other staff members such as the following?

a. Instructional coordinator_____

_____

_____

b. Community coordinator_____

_____

_____

c. Evaluator_____

_____

_____

174

d. Other (please specify)_____

_____

_____

e. Other (please specify)_____

_____

_____

(P) 13. Describe the program's staff development activities related to the
following aspects.

a. Needs assessment_____

_____

_____

b. Structure of training (pre-service and in-service)_____

_____

_____

c. Characteristics of training

(1) Appropriateness for staff of differing levels of knowledge
and experience.

_____

_____

_____

(2) Practicality_____

_____

_____

(3) Coordination with degree programs_____

_____

_____

(4) Integration with other training_____

_____

_____

c. Audiences trained (program and/or nonprogram staff)_____

_____

_____

_____

(P) 14. Describe the involvement of the community and parents with respect
to the following items.

a. Parent involvement in school affairs_____

_____

_____

b. Community input in program planning_____

_____

_____

c. Evidences of community support for the program_____

_____

_____

d. Parent education_____

_____

_____

e. Parent conferences/counseling_____

_____

_____

(P)　15. Describe the means of communication of the following groups.

　　　a. Among program staff_____

　　　_____

　　　_____

　　　b. Program staff with the following nonproject staff:

　　　　(1) Principals_____

　　　_____

　　　_____

　　　　(2) Other district administrators_____

　　　_____

　　　_____

　　　　(3) Nonprogram teachers_____

　　　_____

　　　_____

　　　　(4) School board_____

　　　_____

　　　_____

(P)　16. What means are used to disseminate project information to school
　　　personnel, parents and community?

　　　_____

　　　_____

　　　_____

PROGRAM STAFF INTERVIEW SCHEDULE

(Check one) _____Project staff   _____Bilingual teacher

(G)   1. What is the intended content of instruction (i.e. the <u>theoretical curriculum</u>) with respect to the following matters?

a. Content areas covered_____

_____

_____

b. Relationship of content to project goals_____

_____

_____

c. Who determines the content?_____

_____

_____

d. What articulation is there between project content and the extant district curriculum?

_____

_____

_____

(P)   2. Describe the presentation of content with respect to the following items.

a. Type of instructional model or theory (e.g. concurrent, alternate week/day, preview-review, half day, resource room, and/or bilingual aide)

_____

_____

_____

_____

C = refers to program content
S = refers to students
G = refers to program goals
P = refers to instructional program

b. Organizational practices (e.g. individualized, large group, learning centers, peer tutoring, small group instruction, and/or team teaching)

_____

_____

_____

_____

(P) .3. Describe the methodologies employed for bilingual education with respect to the following items.

a. Language of instruction

(1) General language use plan of teacher and student over length of program

_____

_____

(2) Daily instructional time in each language_____

_____

_____

_____

(3) Variations for different student groups_____

_____

_____

_____

(4) Criteria for establishing language of instruction_____

_____

_____

_____

b. Approach to nonstandard forms

(1) Acceptance_____

_____

_____

(2) Form of corrections_____

_____

_____

c. Approach to second language instruction

(1) Formal instruction_____

_____

_____

(2) Functional use of second language for content instruction
and other activities

_____

_____

d. Approach to reading instruction

(1) Language in which students learn to read_____

_____

_____

(2) Criteria for beginning reading in second language_____

_____

_____

(P) 4. Describe the specific instructional methodologies used in each
subject area

_____

_____

_____

_____

_____

_____

(P)   5. Describe those aspects of the program that are intended to motivate
         students and improve their self-concept with respect to the follow-
         ing matters:

         a. Appropriate content and language of instruction

            (1) Using L, for instruction_____

            _____

            _____

            (2) Accepting language of the student_____

            _____

            _____

            (3) Content that relates to experiences of students_____

            _____

            _____

            _____

            (4) Culturally relevant material_____

            _____

         b. Improved affective climate

            (1) Placing equal value on both languages and cultures_____

            _____

            _____

            _____

            (2) Insuring student success_____

            _____

            _____

(3) Involving parents_____

_____

_____

(4) Teacher as a role model_____

_____

_____

c. Discipline approach

   (1) Philosopny_____

_____

_____

   (2) Guidelines/approach tc control_____

_____

_____

   (3) Special reward systems (e.g. prizes andorivileges)_____

_____

_____

(P) 6. What materials are used within each of the following categories?

   a. Core materials in use

   (1) Commercial_____

_____

_____

   (2) Locally developed_____

_____

_____

11-76

132

b. Appropriateness

(1) Linguistic_____

_____

_____

2) Cultural_____

_____

(P) -7. Describe the role of each of the following personnel in the classroom

a. Teachers_____

_____

_____

b. Aides_____

_____

_____

c. Parents_____

_____

_____

d. Peers_____

_____

_____

e. Resource staff_____

_____

_____

(P)   3. Describe the program director's work with respect to the following:

a. Leadership style_____

_____

_____

b. Role or responsibilities in connection with each of the following

(1) Funds and budgets_____

_____

_____

(2) Public relations_____

_____

_____

(3) Administration_____

_____

_____

(4) Overseeing instruction_____

_____

_____

(5) Staff training_____

_____

_____

(6) Developing and ordering materials and equipment_____

_____

_____

_____

(7) Staff recruiting and hiring_____

_____

_____

LUCAL AND DISTRICT ADMINISTRATORS INTERVIEW SCHEDULE

(G)   1. Describe the school district's general goals_____

_____

_____

(C)   2. What is the school district's philosophy toward language and cul-
         tural diversity?

_____

_____

(P)   3. To what extent is there articulation of project content with the
         existing district curriculum?

_____

_____

(P)   4. What is the relationship between the project staff and each of the
         following categories of district personnel? Comment specifically
         on program acceptance.

      a. Principals_____

_____

_____

      b. Central office administrators_____

_____

_____

      c. Nonproject teachers_____

_____

_____

      d. The school board_____

_____

_____

      C = refers to program context
      S = refers to students
      G = refers to program goals
      P = refers to instructional program

11-79   195

P\   5. Describe the dissemination of program information to the following
two groups.

a. School personnel _____

_____

_____

b. Parents and the community_____

_____

_____

CHAPTER III

CONDUCTING THE EVALUATION OF PROGRAM OPERATIONS

The successful completion of the planning activities and the establishment of the baseline data for the evaluation enable the program director to initiate the actual evaluation of the bilingual program. As described before, the actual evaluation of the bilingual program takes two thrusts: the evaluation of program operations and the evaluation of student outcomes. These may be viewed as totally separate activities. However, the outcomes or outputs of both activities are used during the analysis function to interpret the overall evaluation results and formulate recommendations for changes in the program. This chapter presents guidelines and procedures for conducting one part of the evaluation, the evaluation of program operations.

The evaluation of program operations employs the discrepancy evaluation design described earlier. Therefore, in simple terms, the evaluation of program operations is performed by first establishing the baseline data on the program. This activity was hopefully accomplished in accordance with the recommended procedures in Chapter II. Most importantly, this activity should have been completed during the first or, at least, by the end of the second month of the program year. The second activity required to perform this facet of the evaluation is to collect another set of data similar to the baseline data. Decisions on what data to collect, how and when to collect the

data, and who will collect the data have already been made during the planning phase of the evaluation activity (See Chapter I). Most of these data are collected by monitoring classroom activities and interviewing various persons associated with the bilingual program. This set of data, decribing actual program operation (e.g., the instructional method being used; the amount of instruction in English; the number of teacher aides assigned to a class, etc.) is compared to the baseline data collected at the beginning of the school year, which describes the program design. The comparison provides the basis for determining if the program was operated as planned. If this is the case, there should be few or minor discrepancies in the two sets of data which describe the program. If the comparison reveals significant discrepancies or deviations, the evaluation must document why this occurred.

Discrepancies in program operations should not necessarily be viewed as a negative finding, since there are many reasons why a program may deviate from its original design. This information, however, is very important in determining if this deviation impacted the instructional program in such a way that it affected student performance. For example, the program may have been designed to provide one hour of instruction in social studies using the student's native language. However, due to scheduling modifications, teacher shortage or other factors, a change was made during the fourth month of the program and instruction in the home language did not occur. The evaluation, nevertheless, was designed to assess student performance in social studies. The resulting student outcomes could show that progress was

188

minimal. An immediate conclusion would be that the program failed.
However, knowing that instruction in the students' native language did
not occur, the program director and evaluator can explain the
resulting student outcomes. The question(s) to be addressed, then, is
why was the program design changed? Should the original design be
reinstated? Does performance data from students who received
instruction in their native language show achievement? Answers to
these and other questions begin to formulate a set of findings and
recommendations for the improvement of the overall program. This
interpretation activity also begins to merge and integrated the two
types of data from the evaluation.

While the example above ties the evaluation of program operations to
the evaluation of student outcomes, it should be clearly understood
that the primary purpose of this part of the evaluation is to examine
and monitor the manner in which the program is being implemented.
Additionally, the discrepancy evaluation design makes no attempt to
infer or determine program impact.

This chapter, therefore, describes procedures for evaluating the
instruction, staff development, and parent involvement components of
the bilingual program. While there are other facets of the program
operations that merit attention, these components are the most
significant to the overall operation of the program. The level of
effort allocated to the evaluation of each of these components depends
upon its emphasis and/or importance to the overall program, as
established during the priority setting activities of the planning

process. These issues should be addressed and resolved by the program director and evaluator in planning and designing the evaluation (see Chapter I).

Most of the activities required to evaluate the program operations are conducted throughout the program year rather than at one time. They can therefore be properly planned and scheduled, taking the other responsibilities of program personnel into consideration. Program personnel should be aware, however, that as various activities of the evaluation process begin to feed data into the analysis activity, analysis may become taxing if not planned properly. The program director, with assistance from the evaluator, must therefore schedule the analysis, interpretation, and reporting activities with this in mind.

The guidelines and procedures recommended in this chapter, in conjunction with those in Chapter II, may appear to be overwhelming in light of the program's resources. In reality, the prescribed procedures should be able to be conducted well within the resources of the program. The Handbook recognizes the fact that most bilingual programs, in addition to their personnel, only have an average budget of approximately $2,000 - $5,000 per year to secure the services of independent evaluators. The baseline data gathering activity may require extensive effort the first time that it is performed, however, updating the data for use in subsequent years should not require a great amount of time. The majority of the evaluation activities, if properly planned and scheduled, should be able to be performed by the program personnel.

GUIDELINES FOR CONDUCTING THE EVALUATION OF PROGRAM OPERATIONS

1. Evaluate the Program Instruction Component

The evaluation of the instructional program is intended to answer the following two questions:

.1.   Are planned instructional methods actually being used?

2.   Are changes needed in the instructional methods?

Data needed to answer these questions are obtained by observing classroom activities and interviewing program teachers and administrative staff.   This core of information is then compared to baseline information, obtained through activities described in Chapter II (Worksheets No. 7 through 10), in order to determine if the program is operating as intended.   The program director is assumed to have the primary responsibility for conducting activities that monitor program instruction.   Therefore, the program director will need to fine-tune the recommended procedures and worksheets to ensure the unique needs and intents of the bilingual education program are met.   The instructional program is the core of the bilingual program.   The program director must ensure that the level of effort allocated to evaluate this activity is appropriate.

Information on operating instructional programs is obtained by (a) conducting classroom observations, (b) interviewing the teachers whose

classrooms are observed, and (c) conducting supplemental interviews with a sample of program teachers and administrative staff. Each of these activities is discussed below.

Conducting Classroom Observations -- Prior to observing the classroom, the program director should review the program description so that program features which satisfy the goals and objectives can be observed. The features to be observed should be identified during the planning process. The program description is part of the baseline data identified in Chapter II. Classroom observations should become a planned activity of the program director. Following each informal observation, the program director should write a summary of the classroom instruction as it was observed. These brief summaries should be synthesized into brief reports at least three times during the year -- fall, winter, and spring. Later, these brief reports should be used during the comparison activity and incorporated into the final evaluation report. Thus, over time, the program director develops a complete picture of how the classroom instruction is actually being performed. Quality information can only be acquired from frequent, informal classroom visits, not from a few structured observations.

Topical areas that should be observed by the program director will, of course, depend on how the particular program is designed. Some general categories or features to observe include:

o   Language use;

o   Content of the lessons;

o   Teaching methods;

o   Diagnosis and grouping of students;

o   Recordkeeping;

o   Staff roles in the classrooms (teachers and aides);

o   level of participation by students; and

o   Attitudes and general morale of the students.

Worksheet No. 11, which follows, will assist the program director to develop a precise picture of classroom instructional activities. This worksheet should also be used by the program evaluator in conducting observations. The evaluator, who will have less time to spend in the classroom, should conduct several observations to see classrooms in operation. These observations can be informal or more structured depending on the need of the evaluation. These informal visits will provide a relatively unbiased outsider's perspective that is an insightful supplement to the program director's observations. In addition, this information will be beneficial to the evaluator in preparing the final evaluation report.

How to Use Worksheet No. 11 -- The Classroom Observation Schedule (Worksheet No. 11) is designed to collect information about: instructional methodologies employed; amount of time instruction is conducted in each language; variations for different student groups; rate of presentation; indicators of self-concept development and motivation, and the role of the various classroom personnel. This worksheet, when completed by the program director or evaluator conducting the observation, will provide information about program

instruction which when compared to intended program data will form the
basis to determine what changes have occurred in the program, as well
as provide information with which to make decisions for program
improvements.

CLASSROOM OBSERVATION SCHEDULE

Date:_____   Instructor:_____

Class Hour:_____   Observer:_____

I. List the content areas covered during the class period as they occur.

1.   time started:_____
     time ended:_____
2.   time started:_____
     time ended:_____
3.   time started:_____
     time ended:_____
4.   time started:_____
     time ended:_____
5.   time started:_____
     time ended:_____
6.   time started:_____
     time ended:_____
7.   time started:_____
     time ended:_____
8.   time started:_____
     time ended:_____
9.   time started:_____
     time ended:_____

II. List the instructional methodologies employed as they occur during the period:

Summary statement (enter at end of period):

III. The beginning and ending time for each of the instructional components of the close period can be indicated in item I above. In addition the observer can indicate here estimates of how much time fell within each of three categories during each three minute segment of the class period.

Three Minute Period	Off-task Time	On-task Students Active*	On-task Students Passive	Three Minute Period	Off-task Time	On-task Students Active*	On-task Students Passive
1				11			
2				12			
3				13			
4				14			
5				15			
6				16			
7				17			
8				18			
9				19			
10				20			

* One or more students engaged in behavior for which they get feedback from the teacher.

11-89

195

196

IV. Describe any variations in
teaching approach used for
different student groups
(include any variations in
pace of instruction for in-
dividuals or groups)

V. Describe any evidence of self-
concept development and mo-
tivation including indicators
of (a) accepting the language
of the student and (b) con-
tent that relates to the
experience of the students

Summary statement (enter at end of
period)

Summary statement (enter at end of
period)

VI. Describe the role of all of the following personnel who were present
in the classroom.

(1) Teachers:

(2) Aides:

(3) Parents:

(4) Parents:

(5) Resource staff:

Conducting Teache: Interviews -- Interviews with the teachers whose classes were observed may answer questions of whether instructional methods have changed from the original planned instruction, the reasons for the changes, and what changes in instructional methods may be needed. Worksheet No. 12 provides a sample interview schedule to use in conducting these interviews with teachers.

How to Use Worksheet No. 12 -- This worksheet contains a series of questions which help to direct the teacher interviews and may be used by the program director or evaluator to interview the teachers whose classrooms were observed. Within a week after each of the classroom observations, interviews should be completed by the individual who conducted the classroom observation to ensure that the interview is focused on the particular methodologies the teacher employed and the manner in which these methods were utilized.

PROGRAM OPERATIONS INTERVIEW SCHEDULE FOR TEACHERS

1. What are the major instructional methods that you employ?

   _____

   _____

   _____

2. Why do you use these particular methods, i.e. are these particular
   methods directed to particular instructional objectives?

   _____

   _____

   _____

3. Are there other instructional methods that you would prefer to employ if
   it were not for various circumstantial constraints that you face?

   _____

   _____

   _____

4. If so, what are these constraints?_____

   _____

   _____

   _____

5. What program changes would you recommend that would facilitate your
   efforts to provide the best instruction possible?

   _____

   _____

   _____

6. How typical would you say the class period that we observed was in terms of
   the instructional approach used and the nature and amount of interaction
   with students? How was it atypical?

   _____

   _____

   _____

7. How do the entry and exit criteria and procedures actually used differ from
   those planned for the project? (Interviewer: Be prepared to describe the
   planned procedure.

   _____

   _____

   _____

Supplemental Data Collection -- In establishing the baseline data
(Chapter II), interviews were conducted with program personnel,
parents, and district personnel. Worksheets Nos. 8, 9, and 10 were
used to guide the interviews and record the responses from the
individuals about their understanding of the intended goals,
audiences, and activities of the bilingual program. A similar set of
activities need to be undertaken to identify information about the
actual operation of the program. Thus, the final step in evaluating
the instructional program is to interview a sample of program
personnel, parents, and local and district administrators.
Information obtained from these interviews becomes a direct link to
the interview data used in establishing the baseline data. Comparing
these to sets of data is crucial in identifying discrepancies which
guide program improvement.

The program director should plan to re-interview a sample of program
personnel as well as local and district administrators to elicit
information about actual instructional operations. Worksheets Nos. 9
and 10, when slightly modified, can be used as a tool to guide the
interview and record responses. The program evaluator can modify and
use Worksheet No. 8 as a tool to re-interview the program director.

Once the interviews have been completed, the information should be
synthesized by the program director and evaluator. This information
is then compared to the baseline data so that discrepancies between
planned and actual program operations can be noted.

Analysis of Program Instruction Data -- A determination of whether or not the instructional component of the program is operating as intended is made by comparing baseline information about the design and plan of the instructional program (see Chapter II) to the information acquired from the evaluation of program instructional activities. This comparison leads to the identification of discrepancies between intended and actual program operations. Noted discrepancies identify areas or issues which may require decisions to correct the discrepancies. Later, these discrepancies may also be taken into account in the interpretation of student outcome data if the changes in the instructional program are determined to have influenced student performance. The triad of intended operations/instruction data, actual operations/instruction data, and student outcome data forms the basis for identifying final recommendations for the evaluation report.

Interpretation and Use of Results -- The results of these analyses is presented to those persons responsible for decisionmaking. The program director reviews and analyzes the data to determine if either immediate or future changes should be sought in the program operations and instructional methods employed. Frequent and immediate reports to the program staff should be provided by the program director. Such reports enable staff to review the intended changes, identify means of implementing the changes, and, consequently, be a part of the program improvement process.

$2()_{i1}$

Additional interpretation is performed by the evaluator. Using data
from the various sources, the evaluator can examine the triad of
intended instruction, actual instruction, and student outcomes to
recommend changes which should be sought and ways to implement these
changes.

2. Evaluate the Staff Development Component

The evaluation of the staff development activities of the instructional program compares the actual training provided to teachers to that which was planned. The comparison provides decisionmakers with information about what training actually took place and how this training is related to the intended goals of the program, as well as whether the training met the needs of the program. Specifically, the evaluation of the staff development activities answers the following questions.

1. Were the staff development activities conducted as planned?

2. Did staff training activities meet the needs identified at the onset of the program?

3. Did staff participants acquire the intended knowledge and skills?

4. Were staff satisfied with the training provided?

5. Were skills acquired through training implemented in the classroom?

Answers to these questions when compared to the baseline information (Chapter II--Worksheets No. 7 and 8), will identify discrepancies between actual staff development activities and intended staff training, as well as provide information on the actual training. A variety of data collection methods can be employed to obtain the data needed to answer the above questions. Methods such as questionnaires, knowledge tests, and observations of instructional techniques can be used to provide the necessary information.

202

Questionnaires -- Information regarding satisfaction with and outcomes of staff training activities can be obtained by questionnaires completed by the program director and staff. Worksheet No. 13 provides a sample questionnaire which can be used to collect information on the actual staff training activities.

How to Use Worksheet No. 13 -- The Staff Development Questionnaire, Worksheet No. 13, should be administered to the staff being trained by the person(s) responsible for conducting the evaluation of the staff training activities: This questionnaire provides information about the type and duration of training; numbers of program staff involved in the training; and planned and unmet expectations and objectives for the training. This data should be collected within one week following the completion of all training activities which occur throughout the program year, or at the very least, near the end of the program year.

Appropriate analytic methods for analysis of questionnaire data are determined by the form of the data. The evaluator or appropriate member(s) of the program staff should review the questionnaire responses and systematically categorize the information according to the evaluation questions posed.

STAFF DEVELOPMENT QUESTIONNAIRE

Name of training activity_____

Date of training_____

Name of person completing questionnaire (optional)_____

1. In general, what expectations did you have for the staff training pro-
   vided as part of this project?

   _____

   _____

   _____

2. To what extent were these expectations met?_____

   _____

   _____

   _____

3. Based on your knowledge of the objectives for this staff training, which
   objectives do you think have been met?

   _____

   _____

   _____

4. Which objectives do you think have not been met?_____

   _____

   _____

   _____

Knowledge Tests -- A more immediate source of information on the impact of staff training is information derived from administering knowledge tests to trainees during or at the end of the training. These tests, devised by the instructors, should focus directly upon the instructional content of the training. Because of the specificity of such tests, no sample instruments are included in this manual. The results of the knowledge tests can be examined from one or more perspectives. The tests could be administered prior to and subsequent to training, thus allowing comparisons to be made between pre- and post-test scores. An alternative approach would be to use a control group not involved in the training program as a basis for comparison. An additional comparison could be made between the test results and the stated objectives of the training program.

Observation of Instructional Techniques -- The classroom observation process, should yield information on the instructional approaches that are actually being used by teachers. To the extent that staff training is expected to affect instructional approaches used by teachers, the data acquired from the classroom observations are also pertinent to determine whether or not the training accomplished its purposes and is being implemented as planned. For example, it may be possible to determine if staff development activities intended to provide teachers with skills that are to be used in the classroom (such as how to use new materials, or administer tests) were successful by observing the teachers in the classroom.

Classroom observation data should be analyzed according to procedures described earlier in this chapter in order to identify discrepancies between intended and actual staff development activities. Specifically, the major goals of the staff training which pertain to teachers' instructional approaches (Worksheets No. 7 and No. 8) should be compared with actual classroom practices as evidenced by classroom observation data (Worksheet No. 11).

Interpretation and Use of Results -- The program director should examine the results of the analyses described above and determine if the goals of the staff training were met, as well as determine if findings related to staff training can be issued periodically throughout the program year, possibly in conjunction with recommended changes in program instructional operations. Program personnel then will be able to provide reactions to the recommended changes and identify possible approaches for implementation.

· 3. Evaluating the Parent Involvement Component

The evaluation of the parent involvement component should address four questions. These questions are:

1. To what extent did the level of parent involvement match the planned level?

2. Were parents satisfied with their level of involvement?

3. Was the program staff satisfied with the level of parent involvement?

4. To what extent and in what ways has parent involvement changed over the life of the program?

Data collected and used to answer these questions when compared to information about the planned level of parental involvement, identified in Chapter II (Worksheets No. 7 through No. 10), should identify if discrepancies exist. Data needed to answer these questions can be gathered by conducting a variety of interviews.

Parental Involvement Interviews -- An comprehensive interview should be conducted with the individual most knowledgeable about parent activities. This person could be the program director, parent activities coordinator, principal, or some other staff member. The Interview Schedule for Leader of Parent Activities (Worksheet No. 14) can be used to conduct this interview. This worksheet can also be used to elicit information from a sample of program staff and administrators about actual parental involvement activities. In addition, several interviews should be conducted to obtain information from a sample of parents whose children are involved in the program. Worksheet No. 15 provides a sample interview schedule for conducting these interviews.

The Parent Interview Schedule (Worksheet No. 15) provides a sample interview schedule for conducting parent interviews either in person or by telephone. The evaluator should select a representative sample of parents to be included in this evaluation activity. Parent involvement interviews should be conducted during the last few months of the program year.

How to Use Worksheets No. 14 and 15 -- These worksheets provide
guidelines for interviewing a sample of program staff and the
individual most knowledgeable about parent activities. Depending upon
the program's information needs, certain questions can be pursued with
more or less detail and others can be omitted. It may be desirable to
add additional questions which assess the degree of involvement and
satisfaction of parents with the program.

Analysis of Collected Data -- The program director or evaluator should
analyze the data through a simple process of categorizing responses to
open-ended questions, and recording simple averages and tallies of the
frequency of various activities. These data can then be used in
subsequent interpretations.

Interpretation and Use of Results -- Data interpretation should
provide a thorough description of current activities and compare
actual parental activities to previously determined goals. A
consistent and compatible set of recommended changes and future goals
can then be established. The evaluator, program director, Parent
Advisory Council chairperson, and key staff should review the data and
identify changes to be made for the upcoming program year.

The program director and evaluator should report findings to the staff
periodically throughout the year, along with any recommended changes
in program operations. The staff's reactions and suggestions should
then be solicited so that the desired changes can be made through
mutual endeavor.

INTERVIEW SCHEDULE FOR LEADER OF PARENT ACTIVITIES

1. What is the general scope of parent involvement which was planned for the project this year?

_____

_____

_____

2. To what extent have these goals changed since the beginning of the project year?

_____

_____

_____

3. To what extent have these goals been met?_____

_____

_____

_____

4. Are you satisfied with the level of parent involvement? Is the staff as a whole satisfied?

_____

_____

_____

5. To what extent and in what ways has parent involvement changed over the life of the project?

_____

_____

_____

6. What are the most positive aspects of parent activities?_____

_____

_____

_____

7. What aspects of the parent involvement have the most potential for improvement?

_____

_____

_____

8. What changes are you recommending be made in parent activities in the future?

_____

_____

_____

PARENT INTERVIEW SCHEDULE

(C) 1. To what extent have you been involved in school affairs?

_____

_____

_____

(P) 2. To what extent are you aware that the school has gotten suggestions and reactions from the community in planning its bilingual education program?_____

_____

_____

(C) 3. How much community support do you believe there is for the bilingual education project?_____

_____

_____

(P) 4. How much education has the school district provided for you as a parent as part of the bilingual education project?

_____

_____

_____

(P) 5. To what extent are you aware that the school has provided parent counseling or conferences?_____

_____

_____

(P) 6. What information have you received about the bilingual education project from the school district?

_____

_____

_____

(P) 7. The bilingual program has as one of its goals (fill in the goals related to parent involvement). To what extent do you think this goal has been met? What evidence do you know of that indicates this goal has been met?_____

_____

_____

4. Analyze and Interpret Program Operations Data

The analysis and interpreting of program operations data is a straightforward comparison activity. The evaluator simply examines and compares the information collected on the actual operation of the program to the baseline information describing how the program was meant to operate. For example, if the goal of the program was to provide instruction in all academic subjects using the native language of the students, the analysis function, using the second set of information, simply ascertains if this indeed occurred. If the goal was met, the analysis activity documents this. If the instruction did not occur, the analysis activity also documents this and should attempt to ascertain what caused the change in the program design. Both types of findings are recorded and reported in the overall evaluation report. This type of comparison analysis is all that is needed by this component of the evaluation.

Interpreting the findings or attempting to find an association between the findings of this facet of the evaluation to the results obtained from the student outcomes component should be performed very cautiously. The two sets of information are not meant to be "scientifically merged" in accordance with sound methodological evaluation practices. However, an alert and perceptive evaluator may be able to develop some "intelligent perceptions" about the program based on the two sets of information. For example, knowing that history was taught using the home language in the fourth grade, but not in the fifth, the evaluator may want to closely examine the

student outcome data for these two grades. If the data from the fourth grade students shows significant higher achievement than that of the fifth graders, the evaluator can highlight this fact and then present a "professional opinion" suggesting that the instruction in the native language fostered this difference in achievement.

5. Report the Evaluation Results

The information resulting from the evaluation of program operations should be summarized, written, and presented in the format in which it will appear in the Final Evaluation Report. The format for reporting the results will most likely be the same used to establish the baseline data. However, the reporting should contain a section on the evaluation findings and the recommendations being made to improve the program.

CHAPTER IV

CONDUCTING THE EVALUATION OF STUDENT OUTCOMES

The most important goal of any educational program is to improve the performance of the students enrolled in the program. Therefore, determining student outcomes is perhaps the most important part of a program evaluation. The purpose of this chapter is to describe procedures for evaluating student outcomes. The student outcomes to be evaluated can be divided into the following four areas: English (L2) language skills; non-English or first (L1) language skills; academic achievement (e.g., in science, social science, and mathematics); and affective areas of student performance.

Conducting an evaluation of student outcomes is neither very technical nor complicated if the evaluation is designed to simply describe student performance. A student performance evaluation is interested only in determining how the students in the program performed, rather than determining what caused the observed level of performance. An attempt to measure the latter requires a more comprehensive evaluation design than the former. These two different approaches to the evaluation of student outcomes are commonly referred to as evaluations of student performance and program impact or effectiveness. The terms program impact and program effectiveness are used interchangeably in this Handbook.

These two types of evaluations are widely confused when conducting evaluations of most educational (bilingual and other) programs. In particular, many evaluation reports make statements about program impact or effectiveness when actually they have only measured student performance. That is, they have observed that students have done better (or worse) than some standard or comparison group and then have taken the unwarranted step of concluding that the program was responsible. This distinction is so important for those who plan to use evaluation results that a discussion of these two types of evaluation is presented below.

Evaluating Student Performance

Evaluations of student performance and evaluations of program impact are both based on the same kinds of measures such as tests scores or other quantitative measures, such as attendance rates. In both types of evaluation, student scores are compared to some scale or standard to give them meaning. Evaluations of student performance usually group student standards of performance into two categories. Those are:

o   Absolute standards of performance which compare performance such as:

-    Comprehension level (of textbooks, newspapers, job application forms, etc.);

-    Mastery of specific skills such as language, math, or science; or

-    Proportion of days present in school.

These standards of performance are measureable in absolute terms. That is, they provide information on what a student can or cannot do and are not compared to any other external criteria.

o   Relative standards of performance (typically reported as percentile ranks or standard scores) may compare student performance against:

   -    Norm groups (National, State, and local);

   -    Other bilingual students (National, State, and local);

   -    Groups of non-bilingual students in the same school or district; or

   -    Bilingual program students in previous years.

These, of course, are only examples. There are many other comparisons that can be made. However, the more comparisons made the more technical the evaluation becomes, often resulting in inappropriate comparisons and misinterpretation of results.

Measuring absolute performance is often suggested as a solution to the many problems of evaluation, since absolute performance levels are supposed to indicate whether the students learned what was expected of them. Measuring absolute performance, however, is difficult because reliable tests are difficult to develop and criteria for success in academic areas are largely arbitrary. Nevertheless, absolute measures have an important role in evaluating bilingual programs, especially when testing bilingual students in their first language since appropriate comparison groups may be difficult to find.

Relative performance measures are probably the most common measures currently used to evaluate bilingual and other education programs. Standardized tests are the most widely used for this purpose because they enable comparisons, in the forms of percentiles and standard scores, to be made of local student performance to that of a nationally representative norm group. However, locally made tests, attendance records, and virtually any other measures can also be used to compare bilingual students performance to other students in the same district or school.

Relative measurement, like absolute measurement, also requires adequate tests. However, relative measurement can be thought of as going a step beyond absolute measurement because it uses performance data from comparison groups, which provide criteria for success that are not completely arbitrary. Therefore, relative performance measures can be used to measure performance in English skills and academic subjects taught in English. However, using these measures in bilingual program evaluations is not without problems. There is a real danger of making unreasonable comparisons between the comparison group and the students in a bilingual program, resulting in unreasonable conclusions. For example, it may be useful to compare the English reading skills of a group of low-income bilingual program students to those of a group of affluent native speakers of English from the same district. Assuming that the bilingual students scored lower in reading, it would not necessarily mean that the bilingual program had failed or that the bilingual students could not learn, since low-income groups tend to score lower than affluent groups even

where no language difficulties exist. Some evaluator, however, may arrive at the opposite conclusion.

Measures of relative performance should be the backbone of student outcome evaluation measuring English-language skills and academic subjects tested in English. Performance in other languages, generally must be measured in absolute terms because meaningful comparison groups will be difficult to find.

Evaluating Program Impact

Although determining the level of student performance should be the primary goal of most program evaluations, many evaluations attempt to go beyond this to demonstrate that <u>the program is effective and responsible for the observed level of student performance</u>. Explicitly or implicitly, this question of program impact underlies most evalation designs. This Handbook recommends that bilingual programs do not attempt to conduct an impact evaluation.

Demonstrating program impact requires documenting evidence that the program and nothing else was responsible for the student outcomes. This is more difficult than it appears. To do this, the impact evaluation design must immediately address and be able to answer two questions. The first question is what constitutes the "program" and the other is how the students would perform without the program. Most evaluations, however, never define exactly what the "program" includes. Implicitly, the program may be treated as the sum total of

all the methods, materials, teachings, community factors, and other things that affect the students. .. the evaluation is trying to determine whether the specific features (methods, curriculum, use of two languages, etc.) of the program are effective, then the definition of the program becomes very important. For example, some research (and the intuition of many educators) suggests that the teachers are the most important part of a successful program and that specific materials, methods and so on make much less difference. Therefore, an impact evaluation design must be able to differentiate results emanating from the methods and materials on the one hand, and the personnel on the other.

A practical consequence of this distinction might be that even if the evaluation shows new methods to be effective when performed by the "best" teachers, it does not necessarily follow that the same methods should be adopted by all teachers in the program. In order to make such a decision, the program would have to be defined as being only the methods and materials; and the evaluation would have to demonstrate their effectiveness with a variety of teachers in a variety of settings.

Determining how the students would perform without the program is a very troublesome question for a program impact evaluation. The data may show that students are meeting program objectives and that they score very well in comparison to National and/or local norms. This, however, does not prove that the program is effective. Someone might argue that the same students would do just as well or even better in a regular non-bilingual classroom or in an ESL program.

The laboratory approach to answering this question would be to divide the students _randomly_ into groups--one or more groups for each type of program--and then to compare the effects of the different programs after some reasonable amount of time. In practice, however, because of the diversity of services and the characteristics of bilingual students, this is almost never possible. The result is that the effect of a program cannot be separated from effects of other factors in a conclusive manner. An evaluation using data from a single academic year probably should not even try to prove impact. However, data collected over several years can probably be used to develop an argument that, while not completely definitive, will be reasonably convincing as to the impact of the program. Bilingual programs should attempt to collect multi-year data on student performance.

Problems Associated With Accurate Measurement

In addition to the issues described above, impact evaluations, as well as evaluations of student performance, are themselves impacted by the measurement techniques available to measure performance. The predominant factor is the ability of the evaluation design and the evaluator to control the "noise" or more commonly, the error of measurement. The characteristics of a bilingual program further complicate the problem.

An Analogy: The Signal-to-Noise Ratio

It is generally accepted that test scores include some measurement

error, and that student performance is affected by many things outside of the program. To use the popular term from the stereo recording industry, these various kinds of errors can be thought of as the "noise" in any test score. To pursue the analogy, think of the true changes in student performance (which may or may not represent impacts of the bilingual program) as the "signal" in the test score, just as the music is the signal on a stereo tape or record. If there is a lot of noise in the stereo system, very soft passages of music will be lost in the hiss and static, although very loud passages may be quite clear. In the same way, if there is a lot of noise in an evaluation, small changes in student performance will be obscured, even though dramatic changes would show up quite clearly.

The important issues for anyone involved in evaluation are (1) how much noise is there in a carefully done evaluation? and (2) can changes be expected in students (or impacts due to the program) that are big enough to stand out from the background of noise?

To oversimplify a bit, the answer depends on both how well the evaluation is done and on the evaluation questions that are asked. It is probably safe to say that in the vast majority of program impact evaluations (for all kinds of programs, not just bilingual programs), we are dealing mainly with noise. On the other hand, questions that ask only about student performance can usually be answered quite well. This issue of error in measurement is explained more fully in the section on data analysis.

The characteristics of bilingual programs which compound the measurement of error problems are:

Programs Cover Several Grades -- Most bilingual programs cover several grades, are often started at the lower grade levels and expanded upward, one grade per year. Therefore, a K-6 program cannot be evaluated by simply observing one or two of the lower grades, but require multi-year evaluation designs. Multi-year evaluations present many methodological problems. In fact, student turnover makes most program evaluations longitudinal in theory only.

Programs Change From One Year to the Next -- Bilingual education is characterized by new and constantly evolving instructional approaches and the programs are under great pressure to provide immediate evidence of positive results. However, there is simply no way to do a meaningful outcome evaluation cf a program that is only partially in place or is in a state of flux.

Different Students Get Different Instructions -- Meaningful evaluation requires a clear understanding of what happens to each student. Instruction in bilingual programs often varies widely among students, even within a single classroom. When the instructional program is described clearly it becomes obvious that only a few students received any one treatment. This creates difficulty, since the different groups may be too dissimilar to aggregate, but too small to analyze separately.

Young Children are Difficult to Test -- The testing of young children, especially those below the third grade, is notoriously difficult. Many bilingual programs, however, focus heavily on the lowest grades. There is no obvious answer to this problem, but it should be acknowledged prior to conducting an evaluation.

Popular "Solutions" That Do Not Work

The frustrations generated by the kinds of problems described above have led to many misguided attempts to find solutions. Some fail to answer the impact question, but do answer other questions of possible interest. Others are of no use at all. Approaches that should never

be used are:

o   Raw score posttest minus raw score pretest for
    English language subjects.   In lieu of any better
    ideas, many evaluators simply subtract raw score
    pretest scores from posttest scores and compute
    the difference.   Since almost all groups of
    children make some gains in English language
    subjects, even when they are falling rapidly
    behind their peers, this approach is of no value
    at all for these subjects. A popular variation,
    selecting a gain of some arbitrary number of
    raw-score points as the program target, is no
    improvement.

o   Grade-equivalent scores (the month-for-month gain
    myth).   Analyses based on grade-equivalent scores
    still, unfortunately, appear all too frequently.
    They are based on the mistaken belief that a gain
    in test scores of one or more months for each
    month of instruction represents good progress.
    This is not true. Grade-equivalent scores provide
    an illusion of simplicity but, in fact, they are
    virtally impossible to interpret, even for
    specialists in test construction.
    Grade-equivalent scores should never be used for
    any purpose whatsoever.

o   IQ-based formulas.   From time to time, an attempt
    to use IQ scores appears as the basis for
    evaluating reading or math performance. The idea
    that IQ tests provide an absolute standard against
    which to compare a specific skill is simply a
    misunderstanding.   IQ-based formulas are not
    appropriate for use in bilingual program
    evaluations.

o   Subjective data.   As a last resort, evaluators
    sometimes fall back on subjective data, usually
    teacher reports. Such reports are always useful in
    interpreting results and supplementing
    standardized scores.   However, they can never be
    assumed to represent reliable, valid measures of
    student performance when used alone.

In an effort to find appropriate solutions to these problems,

evaluators have turned to practices which appear to solve these

problems.   However, some of these practices are often misused.

Approaches that are widely misused are:

o Criterion-referenced testing. Some evaluators
   suggest that criterion-referenced tests can solve
   the major problems faced by evaluators. Actually,
   what the criterion-referenced test advocates have
   done is to change the question that is being
   asked. Criterion-referenced tests can provide
   information as to whether program objectives have
   been met. However, measuring student performance
   or program impact still requires reliable, valid
   tests with an adequate range (no floor or ceiling
   effects). In principle, criterion-referenced
   tests could meet these requirements but, in
   practice, most do not.

o Gap-reduction models. "Gap-reduction" is a term
   that appears in the bilingual program evaluation
   literature. It usually means either (a) students
   get closer to the national norms, or (b) students
   get closer to some dissimilar comparison group.
   The former is simply an application of the
   norm-referenced model, which is useful for
   student-performance evaluation but generally not
   for program-impact evaluation. The latter is an
   example of non-random comparison groups (see
   below). The important point is that
   "gap-reduction" is simply a new name for familiar
   designs. The new name does not change their
   strengths or weaknesses.

o Non-random comparison groups. Many bilingual
   program evaluations make use of non-random
   comparison groups, that is, different kinds of
   students who are receiving different instructional
   treatments. As part of any evaluation of student
   performance, such comparisons may be of great
   interest to local decision makers and program
   staff. In general, however, such comparisons do
   not by themselves provide program impact
   information because student differences are
   confounded with program differences.

By this time, the program director may ask if there is really any use
in conducting the evaluation. The answer is yes, provided that the
program director and evaluator fully understand the problems.
Secondly, for these reasons, the Handbook strongly recommends that
evaluations of bilingual programs concentrate their efforts in

conducting evaluations of student performance, rather than impact, when evaluating student outcomes. This, together with the evaluation (description) of program operations meets the Federal requirements, as well as provides the program with sufficient information with which to make informed decisions on how to improve the program.

1. Developing the Evaluation Design

The first steps in performing the evaluation of student outcomes is to determine the type of evaluation that will be conducted and what questions the evaluation is designed to answer. The type of evaluation conducted, however, must address the minimum Title VII requirements.

Title VII requires that bilingual program evaluation include provisions for measuring the accomplishments of the instructional objectives, the progress of the students in improving their English language skills and a procedure for using the information to improve the operation of the program. Meeting these requirements is relatively simple and can be accomplished by following the procedures recommended in the Handbook. In order to meet these requirements, the Handbook recommends conducting an evaluation of student performance, rather than attempting to determine program impact. This can be accomplished by using the basic evaluation design provided in this Handbook.

The Basic Evaluation Design

Because of the difficulty in conducting program impact evaluations, the recommended approach to evaluate student outcomes is simply to evaluate student performance. This approach is referred to in this Handbook as the basic evaluation on the basic evaluation design. This basic evaluation design, therefore only answers the relative

performance question, "to what extent are the bilingual students achieving?"

The basic design has minimal requirements. These are:

o   Testing only the students enrolled in the
     bilingual program;

o   using adequate norm-referenced tests (NRTs)
     capable of measuring English language skills,
     first (L1) language skills, if applicable, and
     academic subjects (e.g., math, science, etc.);
     and

o   measuring performance for only one academic year.

Applying these minimal design requirements to the first student outcome component, English language performance, is all that is reqired to meet the Federal evaluation requirements. However, most bilingual programs should at least evaluate performance in two other outcome areas, first (L1) language and academic subjects. Additionally, although the basic design does not require a multi-year evaluation design, the Handbook does recommend that bilingual programs attempt to collect multi-year performance data. At a minimum, programs should strive to collect data over the duration of their grant period. It is conceivable that data showing progress over the life of the program, can be used to argue that the bilingual program was responsible for the outcome.

Expanding the Evaluation

Programs wishing to extend the evaluation beyond a description of

student performance to measure program effectiveness and/or impact will need to enhance the requirements of the basic design. At a minimum, these evaluation designs may require three modifications. They will have to obtain test scores for comparison purposes from students enrolled in other bilingual or non-bilingual programs. In practice, this option may only be realistic for programs located in school districts that employ district-wide testing programs, where scores for all district students are readily available through computer services or some other easy-to-use form, or if a comparable group of students can be identified and tested.

Single-year evaluations only serve the purpose of the basic evaluation design and can only document if the program is effective compared to baseline data, but cannot show year-to-year changes. Therefore, evaluations attempting to measure effectiveness will most likely require multi-year evaluation designs capable of tracking students throughout their participation in the program. Multi-year evaluations require the use of the same measurement instruments throughout the evaluation period and strict recordkeeping.

Evaluations attempting to measure effectiveness will most likely also need to expand their measurement instruments beyond norm-referenced tests. These may include criterion-referenced tests (CRTs), mastery tests, and other types of measures. Some programs administer these tests as part of their instructional program. The costs to include results from these tests in the evaluation could be minimal and very productive.

How to Select Among the Options -- Ideally, if you want the most complete picture of your program, you should include all of the options. This Handbook certainly recommends that you incorporate any options that can be added at little cost in money and effort. Beyond that, you must decide on the basis of tradeoffs between the amount of effort involved and the importance of the additional evaluation questions that can be answered by adding the different options. The levels of effort and additional resources required for adding the options depend very much on local factors such as the ones described below.

o   Use of local comparison groups. Identifying and testing local comparison groups can easily double the level of effort of your evaluation. On the other hand, if your district has a district-wide testing program with computerized results, comparison group data may be available to you at little or no cost and minimal effort.

o   Use of CRTs (and other tests). Many programs administer diagnostic or mastery tests as part of the instructional program. It may be easy to include results from these tests in the evaluation. At the opposite extreme, some programs make elaborate attempts to develop tests to measure local objectives. Such an effort may be useful for monitoring instruction, but is probably not justified for purposes of evaluating student outcomes.

o   Using longitudinal evaluation. The main requirements for longitudinal evaluation are continuity of personnel, proper planning, and careful recordkeeping. This option is essential if you are really interested in monitoring the progress of your students. Single-year evaluations serve little purpose beyond meeting funding-agency requirements.

o   Using and developing baseline data. Baseline data are obtained by testing bilingual students before the program starts. If these data don't already exist (e.g., from a district-wide evaluation

228

program) it cannot be reconstructed. Before
considering baseline input, make sure your
district maintains all of the required information
in a form that a new bilingual program can use.

Additional Evaluation Questions Can Be Answered -- In the introduction
to this chapter, we discussed three kinds of student outcome
evaluation questions: (a) absolute student performance, (b) relative
student performance, and (c) program impact. The basic evaluation
consists of administering a norm-referenced test to the program
students. This design lets you answer the relative performance
question "How do students compare to a National norm group?" The
options, described above, that you add to the basic design will
determine which additional questions you can answer. These may
include:

o   Absolute student performance questions. In
     general, these questions require the addition of
     appropriate tests such as CRTs, mastery, etc.

o   Other relative student performance questions. The
     different options enable you to compare your
     student to various other groups such as (a) other
     (dissimilar) students in the district (from
     comparison groups), (b) previous program students
     (from longitudinal designs), and (c) pre-program
     students (from baseline data).

o   Program-impact questions. Each piece of
     student-performance information will provide some
     clue for possible program impacts. However,
     strong evidence would have to include both (a)
     evidence that students had improved as compared to
     baseline data, and (b) that other students in the
     district had not made a similar improvement (local
     comparison groups). You will also need evidence
     that the characteristics of program students
     (entering language skills, SES, etc.) have not
     changed. Longitudinal data can show impact if the
     program is improving each year. However, a
     program could be very effective as compared to
     baseline data, but show no changes from year to
     year.

Preparing for the Evaluation

Because the evaluation resources are limited, the evaluation may not be able to answer all questions. Priorities must be determined with respect to the most useful information to be obtained from an evaluation. The evaluation does not have to provide data on each student's learning outcome. The evaluation may provide data only on the students as a group. For example, measurements may be made of changes in reading achievement of third graders but not on reading achievement of a specific student in that grade. The evaluation does not have to provide data on sub-skills such as phonetic analysis but rather on general skill levels such as reading achievement.

Certain decisions must be made before any data is collected to ensure that the analyses can be conducted as desired. Program goals need to be organized according to several key student or program features such as:

o   Subject area (e.g., reading, writing, speaking);

o   Language used in instruction (e.g., English, Spanish);

o   Student language proficiency category (e.g., English: limited or proficient, Spanish: limited or proficient);

o   Grade level of students; and

o   Year of the program.

Worksheet No. 16 allows the evaluator to organize students according to these categories in preparation for measuring student performance.

How to Use Worksheet No. 16 -- The Evaluation Design Worksheet is to be used as a planning worksheet for developing the evaluation of each of the four areas of the student outcome component of the evaluation. The worksheet provides space for listing the different languages and subject areas to be evaluated and the tests to be used. Identification of comparison data and evaluation questions to be answered for the four areas: English language skill, first language skills, student achievement, and affective areas may also be listed.

This worksheet will aid the program director and/or evaluator in keeping track of the decisions to be made for each outcome area. The program director and/or evaluator will need a separate worksheet for each area. Thus, multiple copies of this worksheet will have to be made. In filling out Section I of the worksheet, list the subject areas to be evaluated, the test to be used (name); and the language in which the test is to be administered (e.g., reading, CTBS in Spanish). In the case of norm-referenced tests, list the form, level, and date of the testing. For other tests, such as criterion-referenced or teacher-made tests, provide a brief description of the skill(s) they are designed to assess.

In Section .II, program and student description, list the grade levels in which the subject areas are to be evaluated, the student's language skills, and any other descriptions such as students enrolled in a

special language laboratory. This information at the time of analysis will enable the evaluator to break out the students into separate groups by grade level, language skill, and possibly by any relevant program feature (e.g., students attending a special language laboratory).

For Section III of the worksheet, identify the student groups that the bilingual program students will be compared to, the test to be used in these comparisons, and whether these comparisons involve current or past year test scores. If norm-referenced tests are not to be used, there will be no norm-referenced comparisons to be made. However, scores from district developed mastery tests or criterion-referenced tests for similar or past students can be used to estimate the progress of students currently enrolled in the bilingual program.

Section IV requires a description of the actual comparisons to be made in addressing each evaluation question. In the section on Student Performances, indicate the relative comparisons to be made. An example would be comparing scores of students in the bilingual program with student groups identified in Section III. The absolute standards of performance require identification of past or current similar student progress and which mastery or criterion-referenced test were used.

Date: _____

EVALUATION DESIGN WORKSHEET

I. Subject Area and Language:

   Tests: NRT: _____

            Other: _____

II. Program Student Description:

   Grade Level(s): _____

   Language Skills: English: _____ Other: _____

   Other Descriptors: _____

III. Comparison Data (Groups and Years)

Student Groups	Test Code	Current Year	Earlier Years
A.			
B.		--	
C.		-	

IV. Evaluation Questions

   Student Performance

   1. Relative Standards
       of Performance: _____

       _____

       _____

       _____

   2. Absolute Standards
       of Performance: _____

       _____

       _____

       _____

233

2. Evaluating the English Language Component

The English language skills to be evaluated are the fundamental components to language use. These include knowledge of the sound system for oral language and comprehension of the orthographical system for written language. While each of the four language skill areas -- listening, speaking, reading, and writing -- can be considered individually, one component of language cannot easily be isolated from another. It simply cannot be assumed that mastery of one skill area necessarily indicates mastery of a related skill area; nor can it be assumed that lack of skill in one area indicates lack of skill in another. For this reason, the model recommends that proficiency in all four language skill areas be assessed.

The identification of appropriate norm or criterion referenced instruments is essential to conduct this facet of the evaluation. Although numerous instruments exist, many are not comprehensive or organic in design. This means that the evaluator must carefully select instruments or components of instruments to meet the evaluation objectives.

Measurement of oral language and listening comprehension can be performed by using informal measures. Informal reading inventories or cloze tests may be used to determine the basic reading level of the student. Informal written criterion-references measures may be useful for assessing basic writing skills. The evaluation of the language component may be overlapped with the academic achievement component if norm-referenced measures are used to assess the literacy skill areas.

Three Basic Design Decisions

For practical purposes, most programs must make three basic evaluation decisions: (a) which students to include, (b) what tests to use, and (c) what period of time to include. For each decision, the Handbook recommends a choice for a basic or minimal evaluation and then offer options that will let you answer additional questions if you have the necessary evaluation resources.

o   Which students to include? The basic evaluation requires only testing the students enrolled in the bilingual program. An option could be to obtain data from other students in the district for comparison purposes. Theoretically, the bilingual program staff could pick out comparison groups and test them. In practice, though, this option is realistic only where there is a district-wide testing program, and the scores for all district students are readily available on computers or in some other easy-to-use form.

o   What tests to use? The basic evaluation requires a reliable, standarized, norm-referenced test (NRT) of reading and other language skills. Usually, the test used for district-wide testing may be used. Options include criterion-referenced tests, teacher-made tests, mastery-tests, and tests included as part of commercial instructional packages. We will refer to these kinds of tests generically as "CRTs, etc."

o   What period of time to cover? The basic evaluation requires covering only one academic year and testing only once in the Spring. Two options are highly desirable: (a) multi-year designs following program students from one year to the next, and (b) baseline data on program-type students obtained before the program begins. A sub-issue is whether to test once or twice a year. The first choice should be to test only once a year in the Spring. Options are (a) once a year in the Fall or (b) twice a year, Fall and Spring.

235

~ These basic choices can be summarized as follows:

		Basic Evaluation	Optional Additions
1.	Students	Program only	Comparison groups from the district
2.	Tests	NRTs	CRTs (etc.)
3.	Term of Evaluation	Single year	Multi-year Baseline data
	(Time of Testing)	Spring only	Fall only Fall and Spring

Applying the Basic Design to the English Language Component

The basic evaluation design through the use of a norm-referenced
approach provides for comparing bilingual program students to a
national sample of students who scored at the same pretest percentile
on a nationally-normed test. For example, if the students in the
bilingual program scored at the 25th percentile on the pretest, their
growth can be compared to the growth of the students in the norm group
who scored at the same 25th percentile on the pretest.

The norm-referenced approach makes the equipercentile assumption that
a group of similar students who are not enrolled in the bilingual
instructional program will maintain the same percentile rank
throughout the year. This does not mean that the group without
bilingual instruction is not learning. It simply means that their
learning rate keeps them at a similar position relative to other
students in their grade. In contrast, the students in the bilingual
program will hopefully learn faster than they would in the program.

The question therefore being addressed is, "Do the students in the bilingual program increase their percentile ranking as compared to a national norm group who began at the same percentile?"

Key Comparisons to Be Made

There are many comparisons of performance that can be made. However, the five comparisons which follow are the ones that the evaluator may find useful and can be performed without using complex statistical procedures.

1. Are the students in the bilingual program making gains?

2. Is this year's student performance an improvement over past years?

3. Are students meeting the objectives of the program?

4. Are students doing better in the bilingual program than in another program?

5. Are students doing better than they would be expected to do without the program?

The answers to the first two comparisons can be easily answered by applying the basic design and using a norm-referenced test. The other comparisons require adding one or more of the options described earlier, such as a comparison group of students from another program.

The first question, "Are the students in the bilingual program making gains?" can be answered by administering a norm-referenced test (NRT) of English language skills and comparing the bilingual student

posttest scores with those of the norm group provided by the NRT. Answering this question will provide sufficient information to meet the Federal requirements.

The second question, "Is this year's student performance an improvement over past years?" can be answered by comparing the gains of the students in the program each year, taking into account the error of measurement. When making this comparison, it is very important to realize that it may not be easy to determine why the change from one year to the next occurred. However, other data from the evaluation (the program description and monitoring of program operations) could provide some clues for the observed change.

The third question is "Are students meeting the objectives of the program design?" This is both the most difficult and easiest question to answer. The difficulty comes in deciding what the goal level should be. To establish a realistic goal, the program staff and others need to carefully review the present skill level of the students; the amount and type of instruction required to make a certain change in student achievement; the motivation of students, staff, and parents to implement the change; the accuracy of the assessment instrument; and other similar conditions. Based on this information, the desired performance level on a test or other assessment device can be established.

The fourth question "Are students doing better in the bilingual program than in another program?" must be answered by first

identifying the other program to be compared to the bilingual program.
For example, there may be an alternate program in the school designed
to teach skills similar to those being taught in the bilingual
program, but using a different teaching method. Or, some schools in a
district may be using one method of instruction, and other schools a
second method. A comparison of these programs may be of interest. In
order to make such a comparison, the groups must be comparable, or a
plan to statistically adjust the results must be developed. It is
recommended that comparability of the two groups be established prior
to any comparison, because statistical adjustments of dissimiliar
groups require complicated and sophisticated analytic procedures,
which are not generally available.

The final question, "Are students doing better than they would be
expected to do without the program?" can be answered by the
information from the first question. It is assumed that students are
enrolled in bilingual programs because they need instruction in both
languages. Therefore, if they did not have access to these services,
they would probably not learn as well. If the data show that they are
achieving, then they are doing better.

Many other questions that involve comparisons by race, past
achievement level, social economic status, etc., are not addressed
here because they would either be very costly or very difficult to
measure. Programs attempting to make other comparisons should
approach the exercise with caution.

Selecting Appropriate Tests to Measure English Language Skills

The criteria for selecting an achievement test to measure English language skills in a bilingual program are the same as those used in selecting a test for any evaluation. However, some criteria are more difficult to meet because few tests have been developed with the needs and characteristics of bilingual students in mind. Note also that a major assumption is made about the measurement of the English language component -- that the students learning English language skills have enough English language facility so that testing can occur in English. If this is not true, the students are likely being instructed in their native language and they are acquiring language skills in that language.

The basic evaluation design recommends the use of a standardized, norm-referenced test (NRT) of reading and other language skills to evaluate the English language component. Most school districts now routinely administer one of these tests to all students. If the district does not use a norm-referenced test (NRT) and NRT scores are not readily available, the evaluator may choose to select one of the tests described in the Technical Appendix. These tests are reasonable, reliable, and valid. The main concern should be that the test content matches the program curriculum, at least on a general level. If this basic check is not made, it may later be discovered that the second-grade test covers third-grade curriculum, and vice versa.

240

In an evaluation using a norm-referenced test, the norm group is used as the comparison group for the bilingual program students. Therefore, it is preferable for the norm group to be as similar as possible to the program students. Most available norms for a given grade, however, are designed to be representative of the U.S. population as a whole. Some tests may have norms for different regions of the country or for special educational programs, such as ESEA Title I programs. Norms established for students in Title I programs may be similar to the norms of students in the bilingual program, since their students may reflect similar socio-economic backgrounds.

Finding a test with norms that are comparable to the bilingual students is unlikely, but having an idea of the nature of the differences will help in interpreting the final results. In addition, the test should have norms that are as current as possible. If norms are over 5 or 10 years old, the students were probably experiencing a significantly different curriculum or instructional method than the bilingual students currently being tested.

There are two major problems to consider in selecting NRTs. These are:

o   Test level (floor and ceiling effects). In some
     bilingual programs, the at-grade-level test is too
     difficult for program students at pretest. The
     next lower level may be too easy at posttest time.
     If the mean score on a test is less than 25% of
     the items correct or more than 75% of the items
     correct, floor or ceiling effects probably exist,
     and the test cannot give an accurate picture of
     either student performance or program impact (See

Out-of-Level or Functional Level Testing in the
Technical Appendix).

Multi-year and multigrade-level requirements. Most
bilingual programs cover several grade levels.
Therefore, it is desirable to have achievement tests
that can be used to compare progress across grades and
that can be used to follow groups of students as they
progress through the grades. In practice, this means
using any one of the recognized achievement tests.

Guidelines for Using Norm-Referenced Tests -- The following guidelines

for using norm-referenced tests (NRTs) should be adhered to in order

to produce a valid evaluation.

1.   Do not use the same test score to select students
     for the bilingual program as the pretest score.
     Doing so tends to over estimate the impact of the
     program.   The pretest and selection test scores
     can be separated by:

     o   Administering separate tests;

     o   Using last year's posttest scores as this
         year's selection scores;

     o   Using different subtests of the same test
         battery -- one to select students and one as
         the pretest (both subtests, of course, need
         to be related to the objectives of the
         project); and

     o   Readministering the same test used for
         selection as the pretest.

2.   Tests should be commensurate with the development
     and skill level of the students.

3.   Use the same test form for pretesting and
     posttesting. (Test forms have the same difficulty,
     but contain different, although comparable
     items).

4.   If a norm-referenced test is used, testing should
     occur within two weeks before or after the
     publisher actually administered the test to a
     national sample for norming purposes. These
     empirical norm dates differ from projected norms
     -- norms which are merely estimates of
     performance. Testing done at the same time as

that for the norm group provides more accurate comparisons. Deviations from the norm dates should be in the same direction and magnitude for both pretest and posttest. That is, if pretesting occurred a week before the norm date, the same should be true for the posttest.

Using CRTs (etc.) for Evaluating the English Language Component -- The choice of CRTs (etc.) is more of a curriculum decision than an evaluation decision in most districts. That is, when developing objectives and curriculum materials for a bilingual program, many districts either develop or buy tests matched to their curriculum and the instructional materials. These tests are the best candidates to use in your evaluation. If you have important objectives for student performance that are not covered by any other tests, you may wish to develop or buy special tests just for evaluating student outcomes.

Cautions for CRT Users -- If teachers keep good records of the number of students passing each test and the dates on which they pass, these records will provide a form of absolute student performance measure, as well as a progress record over the course of the year. The records are interesting in their own right, and can also be compared from year to year. Often, however, such tests are weak in the characteristics required for outcome evaluation (high reliability and validity, plus adequate floors and ceilings) so they should be viewed as ballpark measures that include a lot of noise (error), and they should be interpreted with great caution. In short, our recommendation is to look at the results from CRTs (etc.), but be careful.

3. Evaluating the Non-English Language Component

Bilingual programs, for evaluation purposes, can be divided based on their non-English language component into three types. These are:

o   Spanish only programs;

o   Single languages other than Spanish programs; and

o   Multiple language programs.

The major differences among these three types of programs, from the evaluator's perspective, are: (a) only Spanish-English programs will find commercial tests readily available, and (b) multiple-language programs often include small groups that cannot be combined for evaluation purposes.

Three Basic Design Decisions

The three basic decisions made for the English language component also apply to the non-English language component: (a) which students? (b) what tests? and (c) what time period? However, the decisions are even simpler for the non-English language component, because there are fewer alternatives available to the evaluator. The basic options can be summarized as follows:

o   <u>Which students?</u> In general, only the bilingual program students will speak the languages in question and therefore the <u>only</u> students that can be included in the evaluation. In a few districts, there may be comparison groups of interest from other programs or other districts who use the same tests. However, in most cases,

only your program students will be tested in the non-English language, making comparison groups unavailable.

o **Which tests?** A limited number of standardized tests are available in Spanish (although their norm groups are not analogous to those from English-language tests, and you should not use the norms as a simple standard of comparison). For other languages, you are limited to, at best, a few commercial, criterion-referenced tests, plus locally-made tests (CRTs, etc.).

o **What period of time?** Here, the evaluator has the option of single-year or multi-year designs since baseline data before the start of a new bilingual program could be collected. However, in practice, few districts will do this. In general, if the English language evaluation is multi-year, the non-English language evaluation should also be multi-year. Otherwise, both should be single-year evaluations.

The decision on once-a-year (Spring) verus twice-a-year (Fall, Spring) testing will probably also be the same for non-English testing as for the English language testing.

The basic choices are summarized below.

		Basic Evaluation	Optional Additions
1.	Students	Program only	None from the district
2.	Tests	CRTs, etc. (NRTs for Spanish)	None
3.	Term of Evaluation	Single year	Multi-year
	(Time of Testing)	Spring only	Fall only Fall and Spring

How to Select Among the Options As you can see, the only real option is whether to include the non-English language component in the

215

evaluation at all. If you want to know how your students are doing in this area, you will almost certainly be able to produce teacher-made tests that will serve your purposes, but you need to consider exactly which questions you can answer with such tests.

Type of Performance That Can Be Measured

At first glance, it might appear that the evaluation is only able to answer absolute student performance questions for the non-English language component. However, there is one key difference between English and non-English language performance that lets the evaluator consider program impact questions as well. It is a fact that most students improve their English whether or not they are in a bilingual progam. Therefore, the burden of proof in program impact evaluations falls on the evaluator to show that the students do better in the bilingual program than they would have done without it. In evaluating the non-English language component, however, the evaluator is probably safer in assuming students would learn little, or no reading or writing should occur, without the bilingual program. Therefore, the evaluator may be able to argue that the program is largely responsible for _any_ level of performance they achieve. With this in mind, the options and the questions that can be answered for the non-English language component are depicted below.

216

Type of Test		Single-Year Evaluation	Multi-year Evaluation
Test Program Students Only	CRT (etc.)	Absolute Performance Mastery of lesson content	
		Relative Performance Compared to no program	Relative Performance Improvement over time

Key Comparisons to be Made

The key comparisons that can be made relative to non-English or first (L1) language skill development/performance can be the same as those made for the English language component. Performance measurement agains norms will only be possible for Spanish language performance. Therefore, answering the first comparison question for other languages will have to be made by using locally developed mastery tests.

Answering the other questions may be done by following the same procedures as before. Answering the fourth question, which requires a comparison group, should not even be attempted.

Selecting Tests for the Non-English Language Component

Selecting tests for this component is difficult because there are very few tests available. Spanish versions are available for the Inter-American Tests, the CTBS, and the ETS Circus test. However, conventional non-English language norms do not exist. The Inter-American Tests (Spanish) provide user-norms based on students in bilingual programs using that test. The norms provided with the

247

Spanish CTBS do not represent the population of Spanish/English
bilingual students. Norms for both tests can only provide comparison
standards for student performance evaluation, and these comparisons
are difficult to interpret. So far as the review of literature
indicated, no large-scale norm groups have been tested in any other
languages.

Commercial Tests for Languages Other Than English -- The first two
Spanish language tests mentioned above, while not conventional
norm-referenced tests, are similar in terms of reliability and
validity to other standardized tests. The El Circo (Spanish) test for
primary students also represents a high degree of development. All
three can be used for measuring student performance and comparing
program students from year to year. Standardized tests to measure
language achievement, particularly in the first language, may be
difficult to find. In this situation, it would be appropriate to
utilize criterion-referenced measures or teacher-made tests.

Must English and Non-English Language Tests Come From the Same
Publisher? -- This question applies mainly to Spanish-English
programs, since few tests are available in other languages. While
there are some advantages to dealing with a single test publisher, it
is more important to get the most appropriate tests in each language.
Limiting choices to tests that are published in two languages is an
unnecessary restriction.

<u>Teacher-Made Tests</u> -- For languages other than Spanish, many projects
will have to depend on teacher-made tests. These tests should be
quite adequate for demonstrating that students are gaining skills in
their non-English languages. In general, they will not be adequate
for measuring small, year-to-year changes in program effectiveness.

249

4. Evaluating Student Performance in Academic Areas

Evaluation of performance in academic areas requires the specification of the skills to be assessed, selection of the language in which skills are to be measured, and the identification of appropriate tests in English and/or the first language of the student. The evaluator will need to determine which skill areas are to be included in the evaluation. Measurement of achievement in literacy as well as in major academic subject areas may be appropriate. This determination will have to be made on a program-by-program basis. If a student is not literate in L1 or L2, then achievement testing will not be appropriate. If the students are literate, the language in which to test the students will depend upon the language in which instruction in the particular subject has been given, as.well as the fluency of the student in that language.

The Basic Design

Many bilingual programs include non-language, academic subjects, such as math, social studies, and science. The same principles that apply to the English language component apply to this component if testing is done in English. A minimal evaluation would consist of (a) testing program students only, (b) using standardized, norm-referenced tests, and (c) a single-year design. Options include local comparison groups, longitudinal designs, and baseline data.

Language of Testing

The major issue in evaluating performance in academic subject areas is whether to test in English or in the first (L1) language. The evaluation will be easier to implement and the results easier to interpret if the testing is done in English. However, as a matter of common sense, if the students are weak in English and much stronger in their native language (e.g., new arrivals or young children from non-speaking homes), then testing in the native language may be required. In such cases, the evaluation design principles for non-English language components apply (see above).

Selecting NRTs

By and large, the discussion of tests for English language also applies to tests for academic subjects tested in English. The discussion of non-English language tests applies to tests of math, science, etc. in non-English languages. The basic rule here, as it was for English language, is to utilize the test that is used throughout your district. The Technical Appendix contains a discussion on the selection of achievement tests, as well as a listing of these tests for testing language, mathematics, science, etc.

Using CRTs (etc.)

As in language testing, if you have test data available from your instructional program on math, science, or other subjects, you may

want to include these data in your bilingual-program evaluation. For
subjects tested in languages other than English or Spanish, you may
have to depend on teacher-made tests, and the normal cautions apply.

252

5. Evaluating Affective Areas of Student Performance

Affective goals, like improving student attitudes or behaviors, are mentioned in connection with many bilingual programs. If your program has specific objectives in these areas and if the program includes specific components that are intended to change student attitudes or behaviors, then you should consider evaluating the effects of these components. However, you should be aware of two problems, which are discussed below.

Affective goals must be clearly defined. In many bilingual programs, the non-acadmic goals are defined in very general terms, such as "improving self-concept." The test chosen to evaluate changes in self-concept may be some readily available commercial attitude test that bears very little relationship to the self-concept of the program students. The esults are almost certain to be meaningless.

If you wish to evaluate affective components of your program, then you must define the goals clearly, describe the components of the program that are intended to address the goals, and then identify appropriate measures, such as tests, attendance records, and so on, that match your goals. Then you can begin to consider an evaluation design to evaluate absolute student performance, relative student performance, and program impact in the areas that you have designed.

Affective goals are very difficult to evaluate. While the general evaluation design principles apply theoretically, in practice it is

very difficult and frustrating too evaluate changes in attitudes, self-concept, and so on. This is because (a) there is a great deal of noise in the measurement, (b) most measures are insensitive to change in attitudes, (c) attitudes change greatly from month to month and even from hour to hour, (d) there are few good absolute criteria available, and (e) there are seldom any very good comparison groups available.

The net result is that few evaluations can provide convincing evidence of changes in attitudes or related characteristics of the students. For this reason, we would not advise bilingual programs to invest much of their effort in evaluating these goals unless they are a major focus of the program.

Programs wishing to measure affective areas may consult the <u>Technical</u> <u>Appendix</u>. This volume contains a discussion of self-concept and a listing of different tests available.

6. Conducting the Data Collection Activity

Data collection for the first component of the evaluation, program operations, consists of obtaining student background information, interviewing teachers, program administrators, and parents, as well as observing classroom operations. Data collected for evaluating student outcomes consist of test administration, scoring, and the recording of test scores. The latter activity probably requires a higher level of effort than the former. However, data collection for the student outcome component requires strict discipline and very precise procedures.

Testing the Students

Testing in the academic program areas -- language, math, science, and so on all require the same basic procedures. The main distinction that the evaluator should make is between formal testing for evaluating student outcomes and informal testing for diagnostic or other instructional purposes, and out-of-level or functional level testing. Each type of testing is discussed separately below.

Formal Testing for Outcome Evaluation -- Standardized, norm-referenced tests should always be administered and scored under carefully controlled conditions. If you are serious about using CRTs, teacher-made tests, or any other kinds of tests for purposes of outcome evaluation, the same rules apply. Most of these rules are familiar to all teachers. Two points deserve special mention. For

experienced testers using a familiar test, it is sufficient to bring the group together briefly within a few days of the beginning of testing to review the tests and testing procedures. For new tests or inexperienced testers, each tester should practice administering the entire test under the supervision of the evaluator.

Testing should be done within a few days of the same date each year. For norm-referenced evaluations, the testing should be within a week or two of the time that normative data were collected by the test publisher (or local district).

Informal Testing for Instructional Purposes -- In previous sections, the Handbook suggests including the results from CRTs, teacher-made tests, and so on in the outcome evaluation. In some areas, such as non-English subjects, these may be the only test results that you have. The problem is that many of these tests are given under informal classroom conditions. For example, progress checks or mastery tests are often taken by individual students while the teacher works with other students in the same classroom.

The simple fact is that when you give tests under informal conditions, you can expect a lot more noise (error) in the scores than if the same tests were given under carefully controlled conditions. In general, you will have to choose, at least to some extent, between instructional and evaluation uses for your tests. Tests that are given informally in the classroom will have provide only very rough measures of student outcomes.

Out-of-Level on Functional Level Testing -- Achievement tests provide
useful information for evaluating student performance. The value of
such information is obviously related to its accuracy. Achievement
tests are designed to accurately measure the achievement level of
average students in a certain grade level. However, they may not
accurately assess the achievement level of all students at that grade
level.

A student's functional level, at test time, may be below a test
publisher's recommended test level. This is often suggested by a very
low test score on a recommended test level and may indicate that
guessing (chance) by the student played an important role in the
result. Therefore, students whose scores are primarily a result of
guessing on a test that is too difficult may need to be tested out of
level. That is, they need to be tested with an easier, lower level of
the test.

Functional-level testing, therefore, involves testing students with
test levels most appropriate to their achievement levels.
Functional-level testing can involve testing students with the
recommended test level (in-level testing), or it can mean testing
students with a test level below or above the recommended level
(out-of-testing). Whatever the case, the goal is to test at a level
affording the students the most opportunity to demonstrate their
abilities. The Technical Appendix contains a more detailed
explanation of when to use out-of-level testing, as well as how to
properly conduct the testing.   257

Testing Procedures

Testing procedures simply require following the exact instructions of the test and making sure that pre- and posttesting conditions and procedures are identical. Scoring and recording test data are subject to clerical errors. These errors, however, can be easily held to an acceptable level through adequate care and accuracy checks. Scoring procedures which require the scorer to make qualitative judgments about the adequacy of a response are more difficult to control. These qualitative judgments may involve more than simply deciding whether a response is correct or incorrect.

The following guidelines should be followed during test administration.

253

Guidelines for Administering the Testing

1. Assembling the Students

   o Utilize similar testing conditions for all
      treatment and comparison groups. Consider the
      time, place, and date of test administration.
      Follow technical manuals testing administration
      often contain testing procedure recommendations
      (e.g., avoid afternoon testing, or testing on
      Monday and Friday).

   o Distractions should be minimized. Avoid testing
      in the hall or in the cafeteria when lunch is
      being prepared.

   o Coordinate testing efforts with district testing
      or assessment policies and procedures.

   o Consider teaching test-taking skills to students.
      This includes acquainting students with test
      formats, etc. (NOT teaching to the actual test).

   o Plan for make-up testing.

2. Administering the Test

   o Identify testers. If teachers do not speak the
      appropriate language, identify alternative
      testers.

   o . Conduct inservice training for all test
      administrators. If aides and parent will be used
      for testing, more intensive training will be
      required for them. The items on the list below
      should be addressed:

      - Familiarity with materials

      - Clarity of presentation

      - Adherence to guidelines and time limits

      - Control in the classroom

      - Attention to physical conditions (e.g., seat
         spacing)

      - Practice for individual testing

11-153.

259

- Correct choice of testing dates (e.g., norming dates)

- The need for the inevitable "fill-in" of absentees

o Clearly define roles and responsiblities of testers. Inservice training and determination of roles and responsibilities should be assertively coordinated by the program director.

3. Scoring the Test

o Train test scorers.

o Scored tests should be spot-checked by someone other than the person scoring the test.

o Check interrater reliability.

4. Scheduling

o Testing should be spread out over one or more days so that the burden on the students does not so great as to lower scores. Pre- and posttesting must follow similar schedules.

Scoring of Test Data -- One of the issues in scoring tests and recording the scores is whether to use computers. If the program is very large, the answer should probably be "yes," at least for norm-referenced tests. Many programs have access to district, university or state computer centers that can perform the scoring of the tests. If these services are not available locally, the test publishers or other scoring services can provide them. Hand scoring and recording may still have to be performed for very small programs. In addition, if non-standardized tests are used, it may be necessary to score the tests by hand before entering the scores into a computer for analysis.

A second issue involves the way that scores are organized for recordkeeping purposes. Since a student may stay in a program for several years, be tested several times and have several teachers, it will usually be necessary to keep individual student record files. However, for analysis purposes, it is desirable to group students by classroom. This will require keeping a second set of forms. This should not be a problem if the data are stored in a computer, since the computer can do the work of regrouping the records of the students. Commercial scoring services may be able to do this type of processing. Some commercial scoring services can provide complete analyses of the data, including comparisons across years upon request.

The type of score utilized is very important. Never use grade equivalent scores for any purpose. Use normalized standard scores (preferably NCEs) for all computations and calculations of impacts. Report pre- and posttest performance in percentiles. The use of NCEs is explained in the section entitled "Analyzing Student Outcome Data", which follows this section on testing.

Recording Test Data -- Recording the scores is the final step in the data collection process. To ensure that the scores will be usable, the details of recording should be planned well before pretest time. Where a commercial scoring service is used, the evaluator may have little control over the recording process, but if the program elects to do its own scoring or wishes to transfer scores from computer printouts to a more convenient form, the evaluator must consider two

important issues: (a) the accuracy of the data, and (b) the details of the data recording forms.

Copying scores accurately onto data forms is not a complicated problem. However, even the most conscientious recorders make errors. Therefore, all data forms should be carefully proofread, preferably with one person reading aloud while a second person checks the scores. Attention must also be given to data recording forms. Data forms might appear to be of little importance, but the way in which data have been recorded in many school districts virtually precludes any reasonable analyses. It is not possible to prescribe a standard data format because school requirements vary so widely, but it is possible to state two general principles which must be observed. First, data forms scores must be able to completely identify all scores, and second, data forms must be able to arrange data in a way that facilitates analysis.

Recording Data for Multi-year Evaluations -- A data recording form that works well for a single fall-to-spring evaluation may not be suitable for following student progress over several years. Thus, data recording forms that allow for attrition, regrouping of classes each year and the total number of scores must be developed and used for a multi-year evaluation.

The following guidelines should be used for recording test data:

262

Guidelines for Recording Test Data

1. Most sets of scores will require more than one page. The page should have a number identifying each page and the "number of pages" to ensure that no pages used for collecting data will be missed.

2. Every form containing important information should have a name and date to indicate who filled in the numbers in case any questions arise in the future about the accuracy of the information.

3. Each group for which data are recorded should be clearly identified at the top of the data form to simplify the retrieval of that group's data from a large data base.

4. Each page of forms containing student data should be arranged so that it can be photocopied without the students' names. This permits wide use of the data for research purposes without compromising student privacy.

5. The analysis of data is simplified if only one set of test scores (pre and post) are recorded on each sheet. The rules for listing students (see points 6-11 below) should be followed. The complete name of the pretest and posttest (taken exactly from the test booklets and including publication date) must be listed.

6. Identifying and organizing student names efficiently are the most difficult recording problems. Single year evaluations collecting data through fall and spring testing should have minimal problems. However, multi-year evaluations that follow students over several years are a more difficult task since students come and go from projects, and groups are reorganized every year. The simplest rule is to make sure that the posttest scores are all entered on the same form as the corresponding pretest scores. This at least eliminates the problem of the evaluator trying to find each student's name on two forms.

7. A second rule for listing student names is to establish a standard for listing of names, use it for the life of the evaluation, and for all tests that are used. If a student moves or fails to take some of the tests, then the appropriate entries are left blank, but the student's name should not be eliminated from the list. If new

students enter the program, their names should be added to the end of the list for all tests, even those for which no data will be entered. If there is a compelling reason to change the order of student names in the middle of a project, then either all forms should be changed, or a double set of forms (old and new order) should be maintained.

8.   A rule should be established for recording names. The simplest procedure is to allow plenty of space and to spell out first names and middle initials (e.g. Caldwell, Daniel E.).

9.   Each student should have an individual ID number that identifies the student. For example use a one-digit number to identify an experimental condition, a two-digit number to identify a group or class, a one-digit sex code, and a two-digit student number. In some evaluations, other codes (including letters) can be used, but careful consideration of the situation is necessary in order to permit any desired grouping simply by ID number.

10.   A page on any form should have some reasonable number of entries, probably 20 or 25. The same number of entries per page will facilitate the analysis of the data.

11.   Test dates are critical, especially in norm-referenced evaluations. If all students listed on a form have their pretests in one day and all are later posttested in a single day, then test date information is not really necessary. However, this is usually impossible to predict at the time the form is made up, so e columns should be made to provide space to indicate the dates of make-up tests and late entries into the program.

12.   Pre- and posttest scores should, in general, be in adjacent columns, rather than pairing each pretest raw score with its standard score, percentile score, etc., followed by each posttest score and its transformations.

7. Analyzing Student Outcome Data

The analysis of the student outcome data should be performed or at
least supervised by a trained evaluator. The analysis of student
performance data should simply answer the questions which the
evaluation was designed to answer and make the necessary comparisons
that were established during the evaluation design phase. There are
three steps in this approach:

o   Examine scores for serious mistakes or unusual
     results. The scores can be examined simply by
     drawing the frequency distributions of test
     scores. If two sets of scores are being compared
     for the same students (for example, second-grade
     and third-grade scores) then scatter diagrams of
     one test against the other should be used.

o   Compute the mean scores and standard deviations
     for program (and comparison) students. If the
     scores do not appear to reflect any serious
     problems or unusual program effects, then simply
     compute the mean score for each group of program
     students (and for each group of comparison
     students, if any). The standard deviation (a
     measure of how spread out the scores are) must
     also be calculated and reported for each group.
     The mean scores are used to draw comparisons or
     look for progress of the students.

o   Estimate the possible effect of error on your
     results. What may appear to be changes in student
     performance may only be random changes in the
     scores due to noise (error). Errors, in mea
     scores of 5 to 10 NCEs are not uncommon,
     especially with small groups of students.

In examining the data from the evaluation the evaluator should check
to see if the distribution scores resemble a normal curve (bell
shaped). If the distribution of scores is a different shape, this
could indicate possible problems with the tests, testing procedures,

the scoring procedures or the data computer programs. An abnormal distribution in the data may also be attributable to the effects of the program on specific students. For example, in one bilingual program, the mean scores could show second grade students making a moderate percentile or normal curve equivalent (NCE) gain in reading. However, when individual students scores are analyzed, it may be found that only a few students in that grade have made very large gains while the rest of the students have made little or no change in their percentile standings. This information is useful to the evaluator in concluding that the program is working for some students but not for others. Using this finding, the program director may be able adjust the program for those students not showing improvement in reading.

Another problem in analyzing the data from the evaluation is the kinds of noise (error) that remain in even the best evaluation data. Consideration should be taken to ensure that change in students test scores are not due to noise but too the effects of the programs. Error in mean scores of 5-10 NCEs are not uncommon, especiall' for programs with small numbers of students. Tests of statistical significance provide the best way of estimating the likelihood that the results are simply examples of random error. However, tests of statistical significance do not provide information about the educational importance of results, since small gains can be statistically significant for large groups of students, while what appear to be l~ge gains can be due toq random error with small groups of students. Tests of statistical significance also will not indicate flaws in your evaluation procedures. Thus, individuals responsible

206

for conducting the evaluation should look for possible problems in the
evaluation procedures.   In order to better understand this issue the
following information is presented.

An Analogy: The Signal-to-Noise Ratio

It is generally accepted that test scores include some measurement
error, and that student performance is affected by many things outside
of the program.   To use the popular term from the stereo recording
industry, these various kinds of errors can be thought of as the
"noise" in any test score.   To pursue the analogy, think of the true
changes in student performance (which may or may not represent impacts
of the bilingual program) as the "signal" in the test score, just as
the music is the signal on a stereo tape or record.   If there is a lot
of noise in the stereo system, very soft passages of music will be
lost in the hiss and static, although very loud passages may be quite
clear. In the same way, if there is a lot of noise in an evaluation,
small changes in student performance will be obscured, even though
dramatic changes would show up quite clearly.

Can the Signal Be Separated From the Noise in an Evaluation?

The important issues for anyone involved in evaluation are (1) how
much noise is there in a carefully done evaluation? and (2) can
changes be expected in students (or impacts due to the program) that
are big enough to stand out from the background of noise?

To oversimplify the answer depends on both how well the evaluation is done and on the evaluation questions being addressed. It is probably safe to say that in the vast majority of program impact evaluations (for all kinds of programs, not just bilingual programs), the amount of noise will be significant. On the other hand, questions that ask only about student performance can usually be answered quite well. Finally, even program impact questions can be answered in some districts where conditions and resources permit. Before getting more specific, however, we must pick a type of test score or "unit of measurement" that we can use to discuss the size of effects and the amounts of noise in program evaluation must be selected.

Selecting a Unit of Measurement: The Normal Curve Equivalent (NCE)

The type of test score that we will use is called the Normal Curve Equivalent, or NCE. Like any type of score that we might pick, the NCE has both good and bad features. Perhaps the worst is that it is unfamiliar to many educators. On the positive side, however, NCEs have many technical properties that make them useful in evaluations. They have been adopted by many evaluators in the last few years, and many standardized test manuals now include tables for converting to NCEs.

Basically, NCEs are one of the many varieties of normalized standard scores (others include stanines and T-scaled scores). Like all standard scores, they are generated by the test publishers from norm

group data, so they relate student performance to a nationally representative group of students.

Comparing NCEs and Percentiles -- The NCE scale runs from 1 to 99 like the percentile scale (see Figure 1). In fact, an NCE of 1 is equivalent to the 1st percentile of the national norms, and an NCE of 99 is equivalent to the 99th percentile. Similarly, an NCE of 50 represents the mean of the national norm group, just as a percentile of 50 does. However, there are important differences. According to a popular model of student skills, each percentile unit at the end of the scale represents a large increment of skill, while a percentile in the middle of the scale represents a small increment of skill. For example, a student who wants to raise his or her score from the first to the second percentile (or from the 98th to the 99th percentile) must learn about 15 times as much as the student who goes from the 49th to 50th percentile.

This means that the number of percentile points that a student or a class improves does not tell us much unless we also know the starting point. NCEs, on the other hand, cover the same range but divide the range into 99 equal units in terms of skills. Thus, if we say that a student gains one NCE, we can assume that it always means the same thing regardless of where the student started on the scale.

Measuring Gains in NCEs -- One last point about NCEs is important here. This is the difference between raw score gains (i.e., improvements in the number of items answered correctly) and NCE gains

269

PERCENT OF SCORES
UNDER THE NORMAL CURVE

NORMAL CURVE EQUIVALENTS (NCES)

PERCENTILES

STANINES

STANDARD DEVIATION UNITS
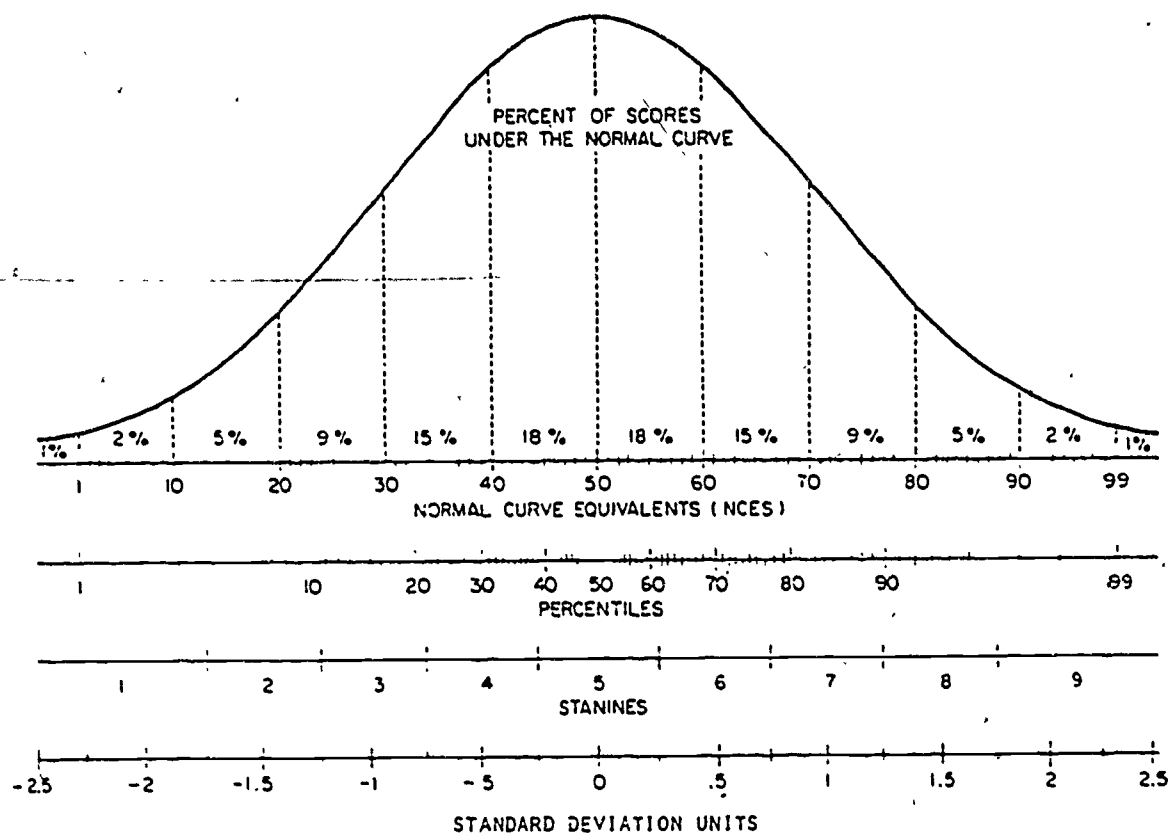
Figure 1.   Area under the normal curve divided into NCEs,
percentiles, stanines, and standard deviation units.

(improvement in relation to the national norm group). If we had a test that covered several grade levels and we gave it to program students each spring, we would expect their raw scores to go up each year. However, we would not necessarily expect their NCE scores to go up. For example, let's say a student is exactly at the mean of the national norm group for his or her grade level (NCE = 50). The next year, our student's raw score will almost certainly go up, but so will the scores of all the other students of the same age. All things being equal, we would expect our student to stay at the mean of the norm group, so the NCE score would still be 50. Discounting any error in the score, any change from an NCE of 50 would indicate that our student was learning faster (or slower) than the average student in the norm group. This could be due to an unusually effective school program or to ways in which our student (or community) differs from those in the national norm group.

In Practice, How Big is an NCE? -- The NCE is, therefore, a useful measure for evaluators, but what does it mean in terms of, let's say, reading skill? A few examples may give you some ideas. Suppose you compared two second graders - one who reads at the average level for second graders and the other (a very good reader) who reads at the average third grade level. The one who reads at the second grade level would get an NCE score of about 50. The better reader would get an NCE score of about 70 or 80 (it is possible to figure this out by studying the norms tables from standardized reading tests). In other words, a difference of roughly 20 to 30 NCEs represents the difference in skill between an average second grade reader and an average third grade reader.

271

By the time students reach junior high school, the average student has developed his or her basic reading skills considerably, and the difference from year to year is not so great as it was at the second grade. At the junior high school level, this difference works cut to roughly 10 NCEs.

As another example, think of comparing good and poor readers at a single grade level. Poor readers in special programs, such as Title I, often average around the 20th percentile. This corresponds to a NCE score of 32. An 80th percentile reader (NCE = 68) would be a fairly good one. In round numbers, then, a rather poor reader must improve about 20 NCEs to become an average reader. A gain of 40 NCEs would take a reader from "rather poor" to "quite good." Similarly, an 80 NCE gain (from NCE = 10 to NCE = 90) would take a student from very poor to very good.

One final example may add to your sense of how big an NCE really is. Suppose that you taught two classes of students in reading, each with about 20 to 30 students. Suppose further that each was a fairly normal class with a normal range of reading abilities. Now suppose that your evaluator told you that one class, on the average, was slightly better than the other. How small an average difference (measured in NCEs) could you expect to detect just by working with the students?

The answer appears to be "somewhere around seven NCEs." That is, if the average scores of the two classes are within seven NCEs of each

other, you probably would notice little if any difference between the classes. With differences greater than seven NCEs you would begin to be aware that one class was noticeably better than the other.

To summarize:

o   Less than 7 NCEs is scarcely noticeable to an observer.

o   The difference between second and third grades is about 20 to 30 NCEs.

o   By junior high school, one grade level is down to about 10 NCEs.

o   A difference of 20 NCEs is quite noticeable. It is the difference between "average" and "rather poor" or between "average" and "quite good."

o   A difference of 80 to 90 NCEs is the difference between the very poorest readers and the very best readers in the typical district.

How Much Noise is There in Measures of Student Performance?

Error of Measurement in a Single Student's Score -- The answer to the noise question is "It depends on whether we are talking about an individual student's score or about an average (mean) score for a group of students." There is almost always a certain amount of random error in a single student's test score. For standardized reading tests, this error will fall somewhere within the range of about ± 10 NCEs for the majority of students, but for some it will be even greater. For about five percent of the students (one out of 20), the error may be greater than ± 16 NCEs. Young students (e.g., second grade) tend to have somewhat more errors in their scores than do older

students, but we are speaking very roughly here and the figures given above are close enough for our purposes.

Error of Measurement in the Mean Score for a Group of Students -- You can see that the amount of error in an individual score can sometimes be very large and that you must, therefore, be very cautious about assigning a student to a special program or to special materials on the basis of a single test score. Fortunately for the evaluator, however, the error of measurement in the _mean_ score for a group of students tends to be much lower than the error in individual scores. This is because the positive and negative errors from the different students tend to cancel out. In fact, for very large groups of students, the random error cancels out almost entirely and the mean score for the group is certain to be very accurate.

Of course, the amount of error in any particular single score or group mean cannot be calculated in most evaluations. However, a simple calculation gives us a good idea of how much error is _likely_ to be present. If we know for a given test that about five percent of the individual student scores will have errors of ± 16 NCEs or greater, we simply divide by the square root of the number of students in our group to get the range of likely errors in the mean score for a group of this size. For example, suppose we have 25 students in the group. The square root of 25 is 5. Sixteen NCEs (the range that covers the errors in most of the individual student scores) divided by 5 equals about 3 NCEs:

27 1

$$\frac{16 \text{ NCEs}}{25} = 3.2 \text{ NCEs.}$$

Thus, when looking at groups of 25 students, about five percent of the group mean scores will be in error by more than ± 3 NCEs.

Similarly, from groups of nine students:

$$\frac{16 \text{ NCEs}}{9} = 5.3 \text{ NCEs or about 5 NCEs.}$$

So, about five percent of group means for groups of nine students will be in error by ± 5 NCEs or more. For the other 95 percent, the errors will be smaller. With 4 students, the range goes up to ± 8 NCEs:

$$\frac{16 \text{ NCEs}}{4} = 8 \text{ NCEs.}$$

Error of Measurement When Comparing Two Groups -- There is one further complication to be aware of. When one compares the mean scores of two groups (or the same group at two different times), each will include some error, and the error in the difference score may be greater than in either score by itself. For example, suppose you test a group of nine students at the end of second grade, and again at the end of third grade. Suppose further that the group mean is 20 NCEs in the second grade and 30 NCEs in the third grade.

Our first reaction is to say that they have improved by 10 NCEs (a small, but probably noticeable improvement). But we also know that each of the scores could be in error by ± 5 NCEs. Could the 10 NCE gain be in error by <u>double</u> this amount? In the worst case, could there be a combined error of -10 NCEs, or in other words, no gain at all?

Statisticians can show that an error this large is not likely. To find the error in the difference between two scores, we should not multiply the error for a single score by two. The correct multiplier is the square root of two (which is 1.4).* In our example, 1.4 times ± 5 NCEs is about ± 7 NCEs. Thus, our apparent gain of 10 NCEs could actually be a true gain of 3 NCEs (i.e., 10 - 7 NCEs). Of course, it could also be a true gain of 17 NCEs (10 + 7 NCEs). In fact, about five percent of such groups (nine students) with real gains of 10 NCEs will appear to have gains greater than 17 NCEs or less than 3 NCEs. For groups of 25 students, the range is about 1.4 times ± 3 NCEs or ± 4.2 NCEs.

Analyzing the Data for Program Impact Evaluations

Once information has been analyzed for student performance, the next step is to analyze data for determining program effectiveness. Analyzing the data for program impact requires a demonstration that the program has had an impact on student performance, it must be shown that student performance is better than expected, and that the program and nothing else is responsible. This does not require any special

276

analysis of the data. It requires the use of data from the program operations evaluation component and student outcomes to build a convincing argument. In addition to the three analytic steps described above, proving program impact will require three basic elements to build a convincing argument. These are:

o **Evidence that students have improved their performance.** This type of information documents that similar students in the same schools had lower scores in the past. This requires compiling data from several different years.

o **Evidence that non-program students have not made a similar improvement.** This type of information examines the possibility that something outside of the bilingual program, such as a new principal or a new district-wide curriculum, is responsible for the improvement in bilingual student performance. This information can only be generated by having local comparison groups -- preferably from district-wide test data.

o **Evidence that the characteristics of the bilingual students have not changed since entry into the program.** In some districts, the student population can change drastically over a period of a year or two (as when large numbers of new arrivals enroll). Some evidence that changes in student population are not responsible for the changes in student test scores must be demonstrated.

Analyzing evaluation data, especially program impact evaluation, is careful, systematic detective work. It consists of looking for clues and followup of any leads that may help to explain the effects (or lack of effects) that are observed in data. A clever and thoughtful evaluator can often build a convincing case by assembling a variety of evidence. Unless it is specifically required that the impact of program be assessed, it is better to spend the effort in developing the instructional program.

2.7

Other Issues in Data Analysis

<u>Single year vs. multi-year analysis</u> -- Many bilingual program evaluations are only concerned with measuring the effects of the program for a single year. These evaluations are not convincing to show program effectiveness. It is, therefore, necessary to demonstrate that there is continuing year-to-year progress toward program goals.

<u>Effects of attrition on multi-year evaluation</u> .-- The effect of student attritions on multi-year evaluations are a problem that all evaluations must be concerned with. Multi-year evaluation means following the same students over a period of years. However, as students transfer out of the program, the number of students in the program gets smaller and smaller until the groups may not be large enough for drawing any comparisons. Another problem is that the ones who transfer will probably be different in many ways from the ones who stay. While multi-year evaluation can give you very useful information, it may be impossible to interpret these results since the progrm may experience constantly changing students.

<u>Floor and ceiling effects</u> -- Floor and ceiling effects are pervasive problems in bilingual program evaluation. A minimal check for these effects on multiple-choice tests is performed by making sure that classroom means or school raw scores are no lower than 25 percent of the correct itmes on four-choice tests, 33 percent for three-choice, and so on. Mean raw scores should not exceed 75 percent of the total

2~3

possible raw score on any test. Outside of these values, the likelihood of floor or ceiling effects, respectively, should be noted in the report.

Other Analysis Techniques

Additional analytic techniques such as Analysis of variance (ANOVA), Analysis of covariance (ANCOVA), and regression techniques can also be used in analyzing the data. These sophisticated statistical analyses can be found in most textbooks on evaluation and are not even mentioned in this volume. This is because these approaches require many special conditions (like random assignment of student to different treatments, and large numbers of students in each group) that simply cannot be met in most bilingual programs. The following guideline should be used when conducting the data analysis activity.

Guidelines for Data Analysis

I. General principles

A. Analyze data both by individual years for short-term goals and cumulatively for long term goals.

B. Separate data according to language proficiency groups.

C. Separate data further according to instructional treatment.

II. Preparation (applies to most evaluation designs)

A. Convert raw scores to standard scores (preferably normalized standard scores such as NCEs). Use these scores for all analyses.

B. Separate out those students with both pre- and posttests.

1. Compute means and standard deviations.
2. Plot the distributions of pretest scores.

3. Plot the distributions of posttest scores.
4. Plot the joint distribution of pretest and posttest scores.

C. For students with pretest scores only:

1. Compute the mean and standard deviation.
2. Plot the distribution of scores.

D. For students with posttest scores only.

Save the scores for student files and for use as next years' pretest scores.

III. Check for irregularities in the data:

A. Floor or ceiling effects

B. Large changes in standard deviation from pretest to posttest.

C. Low correlations between pre- and posttest scores, or irregular joint distributions.

D. Differences between students who took the posttest, and those who dropped out.

E. Look for any other features of the data that strike you as strange, and be sure that you can explain them. Ideally, item data should be examined.

IV. Apply the statistical or other procedures relevant to the particular evaluation design in use.

Be sure that your analyses are relevant to the questions you are trying to answer.

8. Interpreting the Results of the Evaluation

The analysis of student outcome data described above, provides the program director and evaluator with the quantitative information on student performance. If a norm-referenced test was used, the data will show how the bilingual students compared in achievement to a national norm group. Hopefully, the results will show that bilingual students achieved as well or better. These results, however, do not provide answers as to why the students achieved. The answer to this question may possibly be found by carefully examining the results emanating from the evaluation of program operations.

The evaluator should understand that the two components of the evaluation model, the discrepancy evaluation of program operations and the evaluation of student performance, are not methodologically linked together. As a matter of fact, each component may stand alone. The baseline data developed for the evaluation of program operation, however, does play a role in designing the evaluation of student performance. That is, the baseline data provides information to determine what outcome areas should be evaluated.

In addition, the results of the program operations evaluation can provide the evaluator with valuable information on how the program was operated, the instructional approach used, and the amount of instruction provided in the first language for each academic subject area, etc. This information can be used to "understand" the results of the student outcomes component of the evaluation. This information

is valuable to a perceptive evaluator wishing to find answers to explain student performance. For example, if the discrepancy evaluation shows that history was taught using the first language to fourth grade students, but not to students in the fifth grade, the evaluator may want to closely examine the test scores in history for those two grades. Depending on what the test scores show, the evaluator may be able to make some assumptions on what caused either the same or different level of performance. The evaluator may then want to more closely examine "how" the instruction was provided. For example, the evaluator may want to ascertain the level of language proficiency of the teacher teaching in the first language or compare the language assessment scores, if available, of the students in the two grades. All this information, when processed together, could provide clues for understanding what caused the level of performance.

Because the two components of the evaluations are not methodologically linked, there are no specific procedures that can be described for merging the two sets of data. Nevertheless, the recommended approach provides the evaluator with a significant amount of information to use in arriving at conclusions about the program. The analysis techniques required for the evaluation, as described earlier, are relatively simple and can usually be performed by following the instructions in the test manuals as well as the discrepancy procedures described in this Handbook. The other ingredient is the creativity of the evaluator and project director in their ability to use the information to better understand the program and how it might have impacted student performance.

Basically, two general categories of information will be gathered by
the recommended evaluation activities outlined in the Handbook. These
include facts such as: the number of students, the instructional
methods used, the test scores of students, etc. The other category
includes opinions generated by this information such as: whether
there should be more or fewer students in the program, whether the
instructional methods used are appropriate, and whether the test
scores are as high as they should be. It is essential to keep this
distinction in mind when reporting information about the program
evaluation.

The general approach in reporting evaluation results should be first,
to present the facts and second, to present opinions about these facts
clearly identifying the source of the opinions. For example, when
discussing test scores, the fact may be that, as a group, the
bilingual students gained ten normal curve equivalents (NCEs) from
pretest to posttest time. If presented with this information,
different people may interpret this fact in different ways.
Differences in interpretations may result from differences in
understanding of how much gain is typical in a bilingual program, the
nature of the students involved, the instructional methods used, etc.
Therefore, the report must include careful interpretation of the
data.

The procedures and results of the evaluation should be clearly
described. For example, the goals of the English language component,
may be:

1. Students will gain seven or more NCEs in reading, compared to similar students not in the program, as determined by comparing their average NCE gain from pretest to posttest on the New Improved Rural Achievement Test with students in the tests' norm group.

2. Students will gain seven or more NCEs in language art skills compared to similar students not in the program, as determined by comparing their average NCE gain from pretest to posttest on the New Improved Rural Achievement Test with students in the tests' norm group.

Following a statement of the goals, a description of the evaluation procedures used to evaluate this goal should be presented. These descriptions should include the measurement instruments used, the data collected and the analysis procedures. In addition, any information about the evaluation process that would effect interpretation should also be discussed. For example, a description of the evaluation procedures related to the above goals may be stated as follows:

> Attainment of the goal was measured by administering the New Improved Rural Achievement Test to all students in the program during the first week of October and again in the last week of April (the same times when the norming population was tested). Teachers were trained to administer the tests and did so within their classrooms. The analyses performed were a comparison of the pretest-posttest average NCEs to determine the amount of gain as compared to that of the norm group. Separate analyses were conducted for the two content areas (reading and language arts), for each grade level (2-6), and for students at two different levels of English language proficiency. (Students were categorized by these levels of language during the selection process for entry into the program.)

This description should be followed by a presentation of outcomes related to specified goals. The presentation of the outcomes of the evaluation should include two parts. First, the results of the evaluation measurement (i.e., test scores) should be reported. Then a

judgment or well reasoned discussion about the meaning of the results should be offered. These discussions should explain why the program is considered to be responsible for the observed outcomes, or conversely, why the results should not be attributed to the program.

This information should be used to make interpretative comments about the results. Since these comments will inevitably be somewhat subjective, it is important to clearly note whose interpretation is presented. Interpretations may be made by the evaluator based on opinions gathered from program personnel, parents, and administrators. In some cases, an interpretative panel may be established officially to review and interpret the data. Recommendations which logically stem from the results and interpretations are presented in the final section of the report, since the recommendations generally are derived from several sets of results or interpretations (e.g., looking jointly at student outcomes and parent involvement).

The recommendations made for program change should stem from a careful review of all the descriptive information and evaluation results and interpretations presented thus far. The recommendations may best be generated by a team consisting of program staff and the evaluator. However generated, the recommendations should be reviewed by the program director and selected staff to ensure that no major factors which influence the results have been overlooked. Recommendations should then be organized according to the aspect of the program they relate to -- program operations, parent involvement, staff development, or student effects.

CHAPTER V

PREPARING THE EVALUATION REPORT

Preparation of the final evaluation report is an important activity of
the evaluation. The evaluation report is the final and most visible
product of the evaluation. Steps should be taken to assure that the
report addresses the purposes and specific questions of the
decisionmakers for whom the evaluation was planned. In addition, the
evaluation results should be reported in a timely manner, taking care
to ensure that the technical aspects of the evaluation effort are
clearly presented. Together, these steps increase the usefulness of
the evaluation results.

Preparation of the final evaluation report can be a time-consuming and
burdensome process if not properly planned. However, reporting should
be a continual process occurring throughout the evaluation cycle. As
recommended in earlier chapters, brief summaries or reports on
specific activities of the evaluation (e.g., classroom observations)
should have beer prepared and shared with program staff as well as
with key decisionmakers. For example, Chapter III recommended that
following each classroom observation, a brief report should be
prepared. These brief reports were in turn to be summarized at least
three times during the program year--fall, winter, and spring -- and
were to be shared with program personnel so that they could become
part of the program improvement process. Thus, these brief reports
and summaries prepared throughout the evaluation cycle can all feed
into the final evaluation report thus simplifying the reporting

process. The preparation and sharing of evaluation information throughout the evaluation cycle also serves to strengthen communication between the evaluation audiences and those conducting the evaluation, thereby increasing the use of evaluation results.

The focus of this chapter is the preparation of the final evaluation report. The suggestions and guidelines in many cases also apply to the reporting mechanisms recommended throughout the evaluation cycle. The information in this chapter will prove useful to program personnel involved in the evaluation effort as well as to the person(s) responsible for preparing the final evaluation report.

There are a number of basic principles which pertain to the reporting process and serve to simplify preparation of the final evaluation report. This discussion assumes that completion of the report is the primary responsibility of the program evaluator(s) contracted to undertake major segments of the bilingual program evaluation. Basically, the evaluator has three important tasks: develop an understanding of the audiences who will use the information, select a proper reporting format(s), and assist the audiences in using the results. Proper planning of the reporting requirements will make this final activity easy to complete.

2 7

1. Develop an Understanding of the Audiences

The evaluator must understand that clear communication requires knowledge and understanding of the evaluation audiences. The identification of the audiences should have been completed during the planning stages. However, it is helpful to review who the audiences are at the time of reporting. The evaluator communicates with the audiences to identify their information needs and their understanding of evaluation issues, such as testing. This will help the evaluator to tailor the report specifically to the level of understanding of the audiences and to determine the best form in which to report the results. Contact with the audiences also increases the probability that evaluation results will in fact be used.

Understanding the role played by the various audiences in using the evaluation results is also crucial. Some may be involved in clarifying the results of the evaluation, while others will be involved in interpreting these results. Still others are involved in making decisions, and thus are considered to be the key audiences. The roles of the audiences determines the time at which information is reported to them. For example, those involved in clarifying the results enter the reporting process somewhat earlier than those who aid in interpreting the results and making recommendations. Understanding the roles of the audiences assists the evaluator to directing the evaluation report to the proper decisionmakers.

288

2. Select a Reporting Format(s)

Evaluation reports can take different forms, but whatever the form, the report should be designed for a specific audience and be presented in a manner that allows for response and interaction. Although the most common format is a written report, which describes the entire evaluation, consideration should be given to alternative versions for various groups.

A news release is a type of written report. Because news reporters do not have the time to read full evaluation reports, there is a risk that they may write an inadequate or inaccurate news article. To avoid this, preparation of a news release is recommended. The newspaper will probably adapt the news release to its own style and size limitations. In some cases, a press conference may be held for reporting the results to television, radio, and newspaper reporters. Interviews with representatives of the media are even more common. These may be taped for broadcast on television or radio, or they may be the basis for an article by a print journalist.

Oral presentations are also a major vehicle for reporting to professional audiences such as teachers' and program staff. Oral presentations are particularly important for highlighting the major findings, conclusions, and recommendations, and for establishing two-way communication that will clarify, interpret, and influence decisionmaking. Such presentations can be enhanced by a panel discussion and/or small group discussions of the reported results.

Whatever reporting formats are used, the evaluator must focus on the audience(s) and their specific needs. The amount of attention given to the form of reporting may make the difference between a report that is simply received and one that influences practice.

Several standard elements should be included in the report. These include:

o   Statement of purpose;

o   Program overview and background;

o   The goals and objectives of the bilingual program;

o   Description of the program and students;

o   Discussion of the methodology used; including design, sampling strategy, instrumentation, and data analysis procedures; and

o   Presentation of the findings, conclusions, and recommendations for program change.

The report should be concise and should include easily interpreted tables, graphs, and other figures limiting the amount of narrative material presented. Important issues should be identified and highlighted in the report if the results of the evaluation effort are to be maximized. Techniques such as boxing in recommendations or using a different type face are useful to highlight the most important points of the report. Examples of actual data collection instruments should be included in an appendix.

290

3. Assist the Audience Using the Results

Once the written report is completed, copies must be submitted to the funding agency. Plans should also be initiated to present the results of the evaluation to specific audiences. Consideration must be given to identify the appropriate person responsible for presenting the results. It is recommended that this be the program director and the evaluator. A decision as to which of the two will report to which audiences is dictated by the situation and deserves careful consideration.

Arrangements should be made to present the results of the evaluation to the staff, parent groups, school boards, and school administrators. Presentations should include a verbal discussion of the evaluation procedures and findings as well as a discussion of the implications of the findings. Ample time should be available for questions and answers.

Even though most of the information presented at such a meeting is contained in the evaluation report, it cannot be assumed that the audience has either read or understands the complete report. Oral presentations of evaluation findings frequently enhance the credibility of the evaluation and provide the evaluator with important feedback on the comprehensibility of his/her written work. This can be very helpful in improving subsequent evaluation products. Finally, a personal explanation of the evaluation provides evaluation users with an opportunity to ask questions and receive answers and

explanations; something that simply reading the report cannot accomplish. Worksheet No. 17, which follows the instructions for its use, provides a detailed outline for the report.

How to Use Worksheet No. 17 -- This worksheet serves as an outline or checklist which can be used to ensure that all necessary information is included in the report. Generally, Worksheet No. 19 follows the format of this Handbook. The report outline provides a format for the presentation of facts and opinions about the bilingual program. Four major categories of information are presented: evaluation summary, program overview and description, program and student effects, and recommendations. Each of these is discussed below.

Evaluation Summary Information -- This summary information provides a concise overview of the evaluation findings, conclusions, and recommendations. This section of the evaluation report, commonly referred to as the Executive Summary, is a three-to-five page section which should provide the reader (who may be totally unfamiliar with the program) with a brief overview of the program's purpose and structure, as well as a concise description of how well the program is operating and accomplishing its goals. Specific data indicating student and program outcomes should be presented. Recommendations for program changes based on the data should also be included. The Executive Summary is often the only section of the report read by the most influential audiences. The Executive Summary can be provided to persons who most likely do not want or need the more technical information contained in the complete report. The full report,

however, should be made available to interested parties requesting the report.

Program Overview and Descriptive Information -- This information reports on several of the evaluation activities. Overall, factual information is presented about the type of students in the program (e.g., language proficiency, achievement level, biographic data, etc.), their needs, program goals, methods of operation, student selection criteria, instructional approach, etc. In addition, this information also presents factual information on the purpose of the evaluation, its design and the audience(s) who the evaluation is intended to serve.

Program and Student Effect Information -- This information reports on the more technical aspects of the evaluation which includes opinions or evaluative information on the success or failure of the bilingual program. Included is information on each program goal or operation that was evaluated as well as a description of the evaluation procedures used to evaluate each goal. This description should be followed by a presentation of the outcomes related to the specific goals. Included in this description is a discussion of the related results as well as an interpretation of the results.

Recommendations -- The recommendations made for program change stem from a careful review of all the descriptive information and evaluation results and interpretations presented thus far. The recommendations may best be generated by a team consisting of program

staff and the evaluator. However generated, the recommendations should be reviewed by the program director and selected staff to ensure that no major factors which influence the results have been overlooked. Recommendations should then be organized according to the aspect of the program they relate to -- program operations, parent involvement, staff development, or student effects. The recommendations may relate to changes in goals or changes in the way tasks are carried out.

294

BILINGUAL EVALUATION PROJECT REPORT OUTLINE

	Section	Check When Done
I. Executive Summary (3-5 pages)	_____	_____
A. Overview of project goals, numbers and types of students served, instructional approach and evaluation design	_____	_____
B. Summary of findings	_____	_____
1. Instructional methods	_____	_____
2. Parent involvement component	_____	_____
3. Staff development	_____	_____
4. Student outcomes	_____	_____
a. English language	_____	_____
b. NonEnglish language	_____	_____
c. Nonlanguage academic	_____	_____
d. Nonacademic student effects	_____	_____
C. Recommendations	_____	_____
II. Program Overview and Background (2 pages)	_____	_____
A. Context of program including community characteristics, LEA, and school description	_____	_____
B. Student description and needs	_____	_____
C. Program's major goals	_____	_____
D. Program methods	_____	_____
E. Size, scope, and definition of the program	_____	_____

	Section	Check When Done
III. Description of Evaluation (3 pages)	_____	_____
A. Purposes and audiences	_____	_____
B. Evaluation staff and roles	_____	_____
C. Design	_____	_____
1. Questions addressed (includes standards for comparison)	_____	_____
2. Constraints and questions not addressed	_____	_____
3. Relationship to past and future years' evaluations	_____	_____
IV. Program and Student Description	_____	_____
A. Target students	_____	_____
1. Definition of project student	_____	_____
2. Student selection criteria and method	_____	_____
a. Tests and cut-off scores used	_____	_____
b. Role of teacher judgment	_____	_____
c. Role of parent wishes	_____	_____
d. Method of combining criteria	_____	_____
3. Exit criteria and follow-up	_____	_____
4. Student turnover	_____	_____
5. Student characteristics at beginning of year	_____	_____
a. Language proficiency	_____	_____
(1) English	_____	_____
(2) NonEnglish language	_____	_____
b. Achievement level	_____	_____
c. Biographic data	_____	_____

296

	Section	Check When Done
B. Instructional Approach	___	___
1. Self-concept and cultural emphasis	___	___
2. Content of instruction	___	___
3. Presentation of content	___	___
a. instructional model or theory	___	___
b. methodologies for bilingual education	___	___x___
c. specific methodologies for each subject area	___	___
d. role of presentation	___	___
e. self concept development and motivation	___	___
f. materials	___	___
g. personal role in classrooms	___	___
4. Scheduling	___	___
C. Program Management		
1. staff organization	___	___
2. staff roles	___	___
a. Project Director	___	___
b. Teachers	___	___
c. Aides	___	___
d. Other staff	___	___
3. staff development	___	___
a. needs assessment	___	___
b. structure of training	___	___
c. characteristics of training	___	___
d. audiences trained	___	___
4. parents and community	___	___
5. communication	___	___
6. dissemination of project information	___	___

297

	Section	Check When Done
VII. Parent Involvement Component	___	___
A. Goals and objectives	___	___
B. Description of activities to be evaluated	___	___
C. Evaluation procedures	___	___
1. Measures used	___	___
2. Data collection procedure	___	___
3. Analysis procedures	___	___
D. Evaluation Outcomes	___	___
1. Results (including unanticipated outcomes)	___	___
2. Interpretations	___	___
VIII. Staff Development	___	___
A. Goals and objectives	___	___
B. Description of activities to be evaluated	___	___
C. Evaluation Procedures	___	___
1. Measures used	___	___
2. Data collection procedures	___	___
3. Analysis procedures	___	___
D. Evaluation	___	___
1. Results (including unanticipated outcomes)	___	___
2. Interpretation	___	___
V. Student Effects	___	___
A. English language component	___	___
1. Goals and objectives	___	___

298

	Section	Check When Done

V. Continued

    2. Evaluation procedures ___ ___

       a. Measures used ___ ___

       b. Data collection procedures ___ ___

       c. Analysis procedures ___ ___

    3. Evaluation Outcomes ___ ___

       a. Results (including unanticipated outcomes) ___ ___

       b. Interpretation ___ ___

       c. Recommendations ___ ___

  B. NonEnglish language component ___ ___

    1. Goals and objectives ___ ___

    2. Evaluation procedures ___ ___

       a. Measures used ___ ___

       b. Data collection procedures ___ ___

       c. Analysis procedures ___ ___

    3. Evaluation Outcomes ___ ___

       a. Results (including unanticipated outcomes) ___ ___

       b. Interpretation ___ ___

       c. Recommendations ___ ___

  C. Nonlanguage academic component ___ ___

    1. Goals and objectives ___ ___

    2. Evaluation Procedures ___ ___

       a. Measures used ___ ___

       b. Data collection procedures ___ ___

       c. Analysis procedures ___ ___

    3. Evaluation Outcomes ___ ___

       a. Results (including unanticipated outcomes) ___ ___

       b. Interpretation ___ ___

       c. Recommendations ___ ___

299

	Section	Check When Done
VI. Continued		
C. Nonacademic component	___	___
1. Goals and objectives	___	___
2. Evaluation procedures	___	___
a. Measures used	___	___
b. Data collection procedures	___	___
c. Analysis procedures	___	___
3. Evaluation Outcomes	___	___
a. Results (including unanticipated outcomes)	___	___
b. Interpretation	___	___
Recommendations	___	___
A. Program Operations	___	___
1. Instructional approach	___	___
2. Program management	___	___
B. Parent involvement	___	___
C. Staff Development	___	___
D. Student Effects	___	___
VII. Program Operations Evaluation	___	___
A. Instructional Approach	___	___
1. Goals and objectives	___	___
2. Description of activities to be evaluated	___	___
3. Evaluation procedures	___	___
a. Measures used	___	___
b. Data collection procedures	___	___
c. Analysis	___	___
4. Evaluation outcomes	___	___
a. Results	___	___
b. Interpretations	___	___

300

	Section	Check When Done
3. Program Management	_____	_____
1. Goals and objectives	_____	_____
2. Description of activities to be evaluated	_____	_____
3. Evaluation procedures	_____	_____
a. Measures used	_____	_____
b. Data collection procedures	_____	_____
c. Analysis	_____	_____
4. Evaluation Outcomes	_____	_____
a. Results	_____	_____
b. Interpretations	_____	_____

301

VOLUME III

A

TECHNICAL APPENDIX

FOR

EVALUATING ESEA TITLE VII

BILINGUAL EDUCATION PROGRAMS

Prepared By:

INTERAMERICA RESEARCH ASSOCIATES, INC.

1555 Wilson Boulevard

Rosslyn, Virginia 22209

MARCH 1982

302

.3'3

TABLE OF CONTENTS

Volume III

304

This document represents the third and final volume of the <u>Handbook</u> <u>for Evaluating ESEA Title VII Bilingual Education Programs</u>. The Handbook provides practical guidelines and recommended approaches for bilingual education program directors and evaluators to use in evaluating bilingual programs.

In the development of the Handbook, it was readily recognized that a single document could not be equally suitable to all bilingual education programs. Obviously, bilingual education programs cover a range of languages and grade levels in a variety of settings. In addition, some programs have large evaluation budgets and access to teams of highly sophisticated evaluators, while others have limited budgets and no evaluation specialists at all. Thus, Volume III, entitled <u>Technical Appendix,</u> contains a collection of reference material addressing various evaluation issues, as well as lists of tests available. These are intended to assist program directors and program evaluators in building upon or expanding the evaluation activities identified and discussed in Volumes I and II. The appendix also contains full-size reproducible copies of all the worksheets contained in Volume II.

The volume is divided into three sections. Section One includes a fairly comprehensive list of references relevant to the evaluation of ESEA Title VII bilingual education programs. Section Two includes

reference documents addressing issues related to evaluation and testing. This section also contains lists of tests which may be used in the evaluation. The section also includes a reference paper on ethnographic methods for describing bilingual programs. Section three of this volume includes a set of worksheets for use with Volume II of the Handbook. The inclusion of the worksheets in this volume is intended to facilitate the reproduction, dissemination, and use of the worksheets.

Volume I, entitled The User's Guide on Evaluation Basics, summarizes evaluation procedures providing a summary describing the five components of a bilingual education program evaluation. These include: planning, managing, and staffing the evaluation; establishing baseline data required for evaluation; monitoring program operations; evaluating student outcomes; and analyzing and reporting evaluation results.

Volume II, entitled The Designer's Manual for Conducting an Evaluation, describes how to implement each of the components. The Designer's Manual contains recommended approaches, forms, and worksheets--all designed to assist the program director and/or program evaluator in completing the specific tasks associated with the overall program evaluation.

SECTION I

REFERENCES

The following is a fairly comprehensive list of references pertinent
to the evaluation of ESEA Title VII bilingual education programs.
Many of the more technical issues discussed in the Handbook can be
found in these publications.   Program directors and evaluators are
encouraged to familiarize themselves with these publications.   .

LISTING OF REFERENCES

Agarwala-Rogers, Rehka. "Why Is Evaluation Research Not Utilized?" In: Evaluation Studies Review Annual, Vol. 2, edited by M. Guttentag. Beverly Hills: Sage, 1977, 327-333.

Allen, Virginia French. "Toward a Thumb-Nail Test of English Competence." In: Papers on Language Testing (1967-1974). Washington, D.C.: Georgetown University, 1975, 2-5.

Anderson, J.K., Johnson, R.T., Fishbein, R., Stonehill, R.M., and Burnes, J.C. The U.S. Office of Education Models to Evaluate ESEA, Title I: Experiences After One Year. Washington, D.C.: Office of Education, 1978.

Anderson, S.B. and Ball, S. The Profession and Practice of Program Evaluation. San Francisco, California: Jossey-Bass, Inc., 1978.

Angoff, William H., and Sharon, Auriel T. "A Comparison of Scores Earned in TOEFL by Native American College Students and Foreign Applicants to U.S. Colleges." In: Papers on Language Testing (1967-1974). Washington, D.C.: Georgetown University, 1975, 154-162.

Aronson, Sidney H. and Sherwood, Clarence. "Researcher Versus Practitioner: Problems in Social Action Research." Social Work, 12 (4), 1967, 89-96.

Bay Area Bilingual Education League. Bilingual Testing and Assessment, Proceedings of Bay Area Bilingual Education League (BABEL) Workshop and Preliminary Findings. Multilingual Assessment Program. Berkeley, California, 1969 (ED 065-225).

Bennett, C.A. and Lunsdaine, A.A. (Eds.). Evaluation and Experiment. New York, New York: Academic Press, 1975.

Bennett, Carl A. and Lumsdaine, Arthur A. "Social Program Evaluation: Definitions and Issues." In: Evaluation and Experiment: Some Critical Issues In Assessing Social Programs, edited by C.A. Bennett and A.A. Lumsdaine. New York, New York: Academic Press, Inc., 1975, 1-38.

Bernal, Ernest. "Adapting Assessment Procedures to Specific Population Characteristics: The Chicano Child." Paper presented at the Annual Convention of the American Psychological Association. San Francisco, California, 1977 (ED 145-943).

Boruch, R.F. and Wortman, P.M. "Implications of Educational Evaluation for Evaluation Policy." In: Review of Research in Education, Vol. 7, edited by D.D. Berliner. Washington, D.C.: American Educational Research Association, 1979.

Boruch, R.F. and Cordray, D.S. An Appraisal of Education: Federal, State, and Local Agencies. Washington, D.C.: Department of Education, 1980 (Contract Number 300-79-C467).

Biere, Eugene J. and Brown, Richard H. "Norming Tests of ESL Among Amerindian Children." In: Papers on Language Testing (1967-1974). Washington, D.C.: Georgetown University, 1975, 121-127.

Burt, Marina and Dulay, Heidi. "Some Guidelines for the Assessment of Oral Language Proficiency and Dominance." Tesol Quarterly, 12, No. 2, June 1978, 177-192.

Burt, M. and Dulay H. "Relative Proficiency of Limited English Proficient Students." In: 31st Annual Georgetown University Round Table on Languages and Linguistics, edited by J.E. Alatis. Washington, D.C.: Georgetown University Press, 1980.

Bye, T.T. Tests That Measure Language Ability: A Descriptive Compilation. Berkeley, California: BABEL/LAU Center, 1977.

Campbell, D.T. Assessing the Impact of Planned Social Change. Kalamazoo, Michigan: Evaluation Center, Western Michigan University, Occassional Paper, No. 8, 1976.

Canale, M. "From Communicative Competence to Communicative Language Pedagogy." In: Language and Communication, edited by J. Richards and R. Schmidt. New York, New York: Longman, 1981.

Carroll, John B. "Fundamental Considerations in Testing English Language Proficiency of Foreign Students." In: Testing. Arlington, Virginia: Center for Applied Linguistics, 1961.

Carter, Reginald K. "Clients' Resistance to Negative Findings." In: School Evaluation: The Politics and Process, edited by E. House. Berkeley, California: McCutchan, 1973.

Cartier, Francis A. "Criterion Referenced Testing of Language Skills." In: Papers on Language Testing (1967-1974). Washington, D.C.: Georgetown University, 1975, 19-25.

Clark, John L.D. Foreign Language Testing: Theory and Practice. Philadelphia, Pennsylvania: Center for Curriculum Development, Inc., 1972.

Cohen, Andrew D., Bruck, Margaret, and Rodriguez-Brown, Flora V. Evaluating Evaluation: Bilingual Education, Series 6. Arlington, Virginia: Center for Applied Linguistics, 1979.

Cook, T.A. and Pollard, W.E. "Guidelines: How To Recognize and Avoid Some Common Problems of Mis-Utilization of Evaluation Research Findings." Evaluation, 4, 1977, 161-164.

Cooley, W.W. and Lohnes, P.R. "How Can Evaluations Influence Decisions and Policies?" In: Evaluation Research in Education: Theory, Principles, and Practice. New York, New York: Wiley and Sons, 1976, 339-350.

Cox, Gary B. "Managerial Style--Implication For the Utilization of Program Evaluation Information." Evaluation Quarterly, 1 (3), 1977, 499-500.

Cronbach, L.J., and Associates. Toward Reform of Program Evaluation. San Francisco, California: Jossey-Bass, Inc., 1980.

Cronbach, L.J. "Course Improvement Through Evaluation." Teachers College Record, 1963, 64, 672-683.

Cronbach, L.J. and Furby, L. "How We Should Measure 'Change'--Or Should We?" Psychological Bulletin, 74, 1970, 68-80.

Cross, C.T. "Title I Evaluation--A Case Study In Congressional Frustration." Educational Evaluation and Policy Analysis, 1, 1979, 15-21.

Cummins, J. "Linguistic Interdependence and the Educational Development of Bilingual Children." Review of Education Research, 49, 1979, 222-251.

Cummins, J. "The Entry and Exit Fallacy in Bilingual Education." NABE Journal, 4, No. 3, Spring 1980(a).

Cummins, J. "The Cross-Lingual Dimensions of Language Proficiency: Implications for Bilingual Education and the Optimal Age Question." TESOL Quarterly, 14, June 1980(b).,

Cummins, J. "Four Misconceptions About Language Proficiency in Bilingual Education." NABE Journal, 5, No. 3, Spring 1981(a).

Cummins, J. "The Role of Primary Language Development in Promoting Educational Success for Language Minority Students." In: California State Department of Education Compendium on Bilingual-Bicultural Education, 1981(b).

David, J.L. "Local Uses of Title I Evaluation." Final Report on HEW Contract HEW-1G0-77-0095. Menlo Park, California: SRI International, July 1978.

Davis, Howard R. and Salasin, Susan E. "Applied Social Research: In Combat With Waste and Suffering." Evaluation Studies Review Annual, Vol. 2, edited by M. Gattentag. Beverly Hills: Sage, 1977, 320-326.

Dieterich, Thomas G., Freeman, Cecila, and Crandall, Jo Ann. "A Linguistic Analysis of Some English Proficiency Tests." TESOL Quarterly, 13, No. 4, December 1979, 535-550.

Dissemination and Assessment Center for Bilingual Education. Evaluation Instruments for Bilingual Education: An Annotated Bibliography. Austin, Texas: DACBE, 1976.

310

Dyer, H.S. "Toward Objective Criteria of Professional Accountability in the Schools of New York City." Phi Delta Kappan, 52, 1970, 206-211.

Echternacht, G. "The Use of Different Models In the ESEA Title I Evaluation System." Paper presented at the Annual Meeting of the American Educational Research Association. Toronto, Canada, 1978.

Farhady, Hossein. "The Disjunctive Fallacy Between Discrete-Point and Integrative Tests." TESOL Quarterly, 13, No. 3, September 1979, 347-357.

Federal Register, Vol. 45, No. 67, April 14, 1980, Paragraph 123A.24.

Fetterman, D. "Ethnographic Techniques In Educational Evaluation: An Illustration." Journal of Thought, Vol. 15, No. 3, Fall 1980, 31-48.

Fletcher, B.P., Locks, N.A., Reynolds, D.F., and Sisson, B.G. A Guide to Assessment Instruments for Limited English Speaking Students. New York, New York: Santillana Publishing Company, Inc., 1978.

Gamel, N.N., Tallmadege, G.K., Wood, C.T., and Binkley, J.L. "State ESEA Title I Reports: Review and Analyses of Post Reports, and Development of a Model Reporting System and Format." Mountain View, California: RMC Research Corporation, 1975.

Gilmore, G. and Dickerson, A. The Relationship Between Instruments Used for Identifying Children of Limited English Speaking Ability in Texas. Houston, Texas: Region IV Education Service Center, 1979.

Glass, G.V., Willson, V.L., and Gottman, J.M. Design and Analysis of Time--Series Experiments. Boulder, Colorado: Laboratory of Educational Research, University of Colorado, 1972.

Green, J.L. and Wallet, J. Ethnography and Language in Educational Settings. Norwood, New Jersey: Ablex Publishing Company, 1981.

Green, J.L. and Wallat, C. "Mapping Instructional Conversation--A Sociolinguistic Ethnography." Ethnography and Languages in Educational Settings, edited by J.L. Green and J. Wallat. Norwood, New Jersey: Ablex Publishing Company, 1981.

Guba, E.G. "Toward a Methodology of Naturalistic Inquiry in Educational Evaluation." Los Angeles, California: Center for the Study of Evaluation, UCLA, 1979.

Guba, G. and Lincol, Y.S. Effective Evaluation. San Francisco, California: Jossey-Bass, Inc., 1981.

Gunnarsson, Bjarni. "A Look At the Content Similarities Between Intelligence, Achievement, Personality, and Language Tests." Language in Education: Testing the Tests. Rowley, Massachusetts: Newbury House Publishers, Inc., 1978, 17-35.

31

Gurel, Lee. "The Human Side of Evaluating Human Services Programs: Problems and Prospects." Handbook of Evaluation Research, Vol. 2, edited by M. Guttentag and E.L. Struening. Beverly Hills, California: Sage, 1975, 11-28.

*Hanson, J.B. "Measurement Problems in Title I Evaluation." Paper presented at the Annual Meeting of the American Educational Research Association. Toronto, Canada, 1978.

Hardy, Roy. "Should Jose Be Tested In Spanish?" Paper presented at the Annual Meeting of the National Council on Measurement in Education. New York, New York: 1977 (ED 163-043).

Harris, C.W. (Ed.) Problems in Measuring Change. Madison, Wisconsin: University of Wisconsin Press, 1963.

Harris, David P. "Report on an Experimental Group-Administered Memory Span Test." Papers on Language Testing (1967-1974). Washington, D.C.: Georgetown University, 1975, 7-17.

Harris, David P. Testing English as a Second Language. New York, New York: McGraw-Hill Book Company, 1969.

Hawkins, J. David, Roffman, R.A., and Osborne, P. "Decision-Makers' Judgments--The Influence of Role Evaluative Criteria and Information Access." Evaluation Quarterly. 1978, 2·(3), 435-454.

Helmer, Susan. "Demonstration of Assessment of Language Dominance of Spanish-Speaking Bilingual Children." Occasional Papers in Linguistics, No. 1. Carbondale, Illinois: Southern Illinois University, Carbondale Department of Linguistics, 1977 (ED 144-417).

Henry, J. "The Cross Cultural Outline of Education." Current Anthroplogy, 4, 1960, 269-305.

Hernandez-Chavez, E., Burt, M., and Dulay, H. "Language Dominance and Proficiency Testing: Some General Considerations." NABE Journal, 3, 41-54.

Hernandez-Chavez, E. "Critique of a Critique: Issues in Language Assessment." Journal of the National Association for Bilingual Education, Vol. II, No. 2, 1978.

Hill, P., et al. "Compensatory Education Study." Washington, D.C.: National Institute of Education, 1978.

Hoepfner, R. "Achievement Test Selection for Program Evaluation." In: Achievement Testing of Disadvantaged and Minority Students for Educational Evaluation, edited by M. J. Wargo and D. R. Green. New York, New York: Hart Publishing Company, Inc., 1974.

Horst, D.P., et al. An Evaluation of Project Information Packages (PIPs) as Used for the Diffusion of Bilingual Projects. Mountain View, California: RMC Research Corporation, 1980 (RMC Report UR 460).

Horst, D.P. Title I Evaluation and Reporting System: Examination of the Models at the Project Level. Mountain View, California: RMC Research Corporation, in press.

House, E. Evaluating With Validity. Beverly Hills, California: Sage Publications, 1980.

House, E.R., Glass, G.V., McLean, L.D., and Walker, D.F. "No Simple Answer: Critique of the Follow Through Evaluation." Evaluation Studies Review Annual, Vol. 3, edited by T.D. Cook, et al. Beverly Hills, California: Sage Publications, 1978.

Houts, P.L. The Myth of Measureability. New York, New York: Hart Publishing Company, Inc., 1974.

Hubert, J. "An Investigation of the Language Assessment Battery (English, Level I) for Title VII Students in Hartford." Unpublished manuscript, 1978.

Ilyin, Donna. "Structure Placement Tests for Adults in English-Second-Language Programs in California." Papers on Language Testing (1967-1974). Washington, D.C.: Georgetown University, 1975, 128-136.

Jaeger, R.M. "The Effect of Test Selection on Title I Project Impact." Educational Evaluation and Policy Analysis, 1979, 1, 33-40.

Joint Committee on Standards for Educational Evaluation. Standards for Evaluations of Educational Programs, Projects, and Materials. New York, New York: McGraw-Hill, 1981.

Jones, Randall L. and Spolskey, Bernard (Eds.). Testing Language Proficiency. Arlington, Virginia: Center for Applied Linguistics, 1975.

Kuhlmann, Natalie A. "Language Dominance Testing: Some Questions." Paper presented at the 7th Annual Meeting of the California Association of Teachers of English to Speakers of Other Languages. San Francisco, California: 1978 (ED 154-614).

Lado, Robert. Language Testing. New York, New York: McGraw-Hill Book Company, 1961.

Lange, Robert R. and Speiss, M. An Empirical Study of Four Providing Variables Effecting Utilization of Evaluation. Paper presented at the Annual Meeting of the American Educational Research Association. Chicago, Illinois: 1974.

Law, A. Proceedings of the Bilingual Instrument Review Committee (AB 3470). Sacramento, California: Office of Program Evaluation and Research, California State Department of Education, September 28, 1978.

Linn, R.L. "Validity of Inferences Based on the Proposed Title I Evaluation Models." Educational Evaluation and Policy Analysis, 1979, 1, 23-32.

323

Linn, R.L. and Slinde, J.A. "The Determination of the Significance of Change Between Pre- and Posttesting Periods." Review of Educational Research, 47, 1977, 121-150.

Locatis, Craig N., Smith, J.K., Blake, Virgil. Effects of Evaluation Information on Decisions. Paper presented at the Annual Meeting of the American Educational Research Association. San Francisco, California: April 1979.

Longwood, Robert and Simmel, A. "Organizational Resistance to Innovation Suggested by Research." Evaluating Action Programs: Readings in Social Action and Education, edited by C.H. Weiss. Boston, Massachusetts: Allyn & Bacon, 1972, 311-317.

Lorec, P.G., et al. Anchor Test Study. Washington, D.C.: U.S. Government Accounting Office, 1974.

Marco, G.L. "A Comparison of Selected School Effectiveness Measures Based on Longitudinal Data." Journal of Educational Measurement, 1974, 11, 225-234.

Matluck, Joseph H. "The Multiple Role of Oral Language Assessment: Language Status, Research and Curriculum Development." Paper presented at the 23rd Annual Meeting of the International Reading Association. Houston, Texas: 1978 (ED 161-058).

Michigan State Department of Education. Guidelines for Selecting Test Instruments and Procedures for Assessing the Needs of Bilingual Children and Youth, Parts I and II, 1976 (ED 161-951).

Millman, J., Paisley, W., Rogers, W.T., Sanders, J.R., and Womer, F.B. Performance Review of USOE's ESEA Title I Evaluation Technical Assistance Program. Final Report on HEW Contract HEW-300-78-0355. Washington, D.C.: Hope Associates, Inc., April 1979.

Moller, Alan. Validity in Proficiency Testing. British Council, London, England: English Teaching Information Center, 1975 (ED 122-603).

Morris, L.L. and FitzGibbon, C.T. How to Measure Program Implementation and Evaluation Handbook. Beverly Hills, California: Sage Publications, 1978.

Morris, L.L. and FitzGibbon, C.T. "Determining How Well a Test Fits a Program." How to Measure Achievement. Beverly Hills, California: Sage Publications, 1978.

Natalicio, Diane S. and Williams, Frederick. "What Characteristics Can 'Experts' Reliably Evaluate in the Speech of Black and Mexican-American Children?" In: Papers on Language Testing (1967-1974). Washington, D.C.: Georgetown University, 1975, 196-171.

31

Nielsen, V.G. "Why Evaluation Does Not Improve Program Effectiveness." Policy Studies Journal, 3 (4), 1975, 383-390.

Northwest Regional Education Laboratory, Center for Bilingual Education. Assessment Instruments in Bilingual Education: A Descriptive Catalogue of 342 Oral and Written Tests. Los Angeles, California: National Dissemination and Assessment Center, 1978.

Northwest Regional Education Laboratory. Oral Language Tests for Bilingual Students: An Evaluation of Language Dominance and Proficiency Instruments. Portland, Oregon: NWRL, 1976.

Office of Bilingual Bicultural Education. Bilingual Education Program Quality Review Instrument. Sacramento, California: California State Department of Public Instruction, 1980.

Oller, John W. "How Important is Language Proficiency to IQ and Other Educational Tests?" In: Language in Education: Testing the Tests. Rowle, Massachusetts: Newbury House Publishers, Inc., 1978, 1-16.

Oller, John W. "Assessing Competence in ESL: Reading." Papers on Language Testing (1967-1974). Washington, D.C.: Georgetown University, 1975, 26-36.

Oller, John W. and Inal, Neven. "A Cloze Test of English Prepositions." Papers on Language Testing (1967-1974). Washington, D.C.: Georgetown University, 1975, 38-49.

Oller, John W. and Perkina, Kyle. "Language Proficiency as a Source of Variance in Self-Reported Affective Variables." Language in Education: Testing the Tests. Rowley, Massachusetts: Newbury House Publishers, Inc., 1978, 103-125.

Oller, John W. and Perkins, Kyle. Language in Education: Testing the Tests. Rowley, Massachusetts: Newbury House Publishers, Inc., 1978(a).

Oller, John W., Jr. "The Language Factor in the Evaluation of Bilingual Education." Georgetown University Round Table on Languages and Linguistics, edited by J.E. Alatis. Washington, D.C.: Georgetown University Press, 1978(b).

Oller, John W., Jr. Language Tests at School: A Pragmatic Approach. New York, New York: Longman, 1979.

Oller, John W. and Perkins, Kyle. Research in Language Testing. Rowley, Massachusetts: Newbury House Publishers, Inc., 1980.

Palmer, Leslie and Spolsky, Bernard (Eds.). Papers on Language Testing (1967-1974). Washington, D.C.: Georgetown University, 1975.

Patton, M.O., et al. "In Search of Impact: An Analysis of the Utilization of Federal Health Evaluation Research." Evaluation Studies Review Annual, Vol. 3, edited by T.D. Cook and Associates. Beverly Hills, California: Sage Publications, 1978, 59-81.

315

Patton, M.Q. Utilization-Focused Evaluation. Beverly Hills, California: Sage Publications, 1978.

Petersen, Calvin R. and Cartier, Francis A. "Some Theoretical Problems and Practical Solutions in Proficiency Test Validity." Testing Language Proficiency. edited by Randall L. Jones and Bernard Spolsky. Arlington, Virginia: Center for Applied Linguistics, 1975.

Pierce, Lorraine Valdez. "Invalidity of Language Proficiency Tests Used in the Language Assessment of Bilingual Students." Unpublished paper. Washington, D.C.: Georgetown University, 1980.

Pletcher, Barbara P., Locks, Nancy A., Reynolds, Dorothy F., and Sisson, Bonnie G. A Guide to Assessment Instruments for Limited English Speaking Students. New York, New York: Santillana Publishing Company, Inc., 1978.

Politzer, Robert L., Hoover, Mary Rhodes, and Brown, Dwight. "A Test of Proficiency in Black Standard and Nonstandard Speech." Papers on Language Testing (1967-1974). Washington, D.C.: Georgetown University, 1975, 90-100.

Popham, W.J. Evaluation in Education. Berkeley, California: McCutchan Publishing Corporation, 1974.

Provus, M. Discrepancy Evaluation. Berkeley, California: McCutchan Publishing Corporation, 1971.

Pulu, Tupou L. and Silverman, Robert. Guidelines for the Selection, Adaptation, and Development of Language Proficiency Tests. Alaska Department of Education, National Bilingual Materials Development Center, Northwest Regional Laboratory and Bilingual Education Services, 1979.

Rhodes-Hoover, M., Politzer, R.l., and Taylor, O. Bias in Achievement and Diagnostic Reading Test: A Linguistically Oriented View. Unpublished manuscript, Standford University, 1975.

Riecken, H.W. and Boruch, R.F. Social Experimentation: A Method for Planning and Evaluating Social Intervention. New York, New York: Academic Press, 1974.

Roberts, A.O.H. Out-of-Level Testing. Mountain View, California: RMC Research Corporation, 1978.

Rosansky, Ellen J. "A Review of the Bilingual Syntax Measure." Advances in Language Testing Series, edited by Bernard Spolsky. Arlington, Virginia: Center for Applied Linguistics, 1979.

Rosenshine, B. and Furst, N. "The Use of Direct Observation to Study Teaching." Second Handbook of Research on Teaching, edited by R.M. Travers. Chicago, Illinois: Rand McNally College Publishing Company, 1973.

Sanders, J.R. and Nafziger, D.H. A Basis for Determining the Adequacy of Evaluation Designs. Kalamazoo, Michigan: Evaluation Center, College of Education, Western Michigan University, 1976.

Sanders J.R. "Using Multiple Approaches in Field Research." Evaluation News, No. 16, Summer 1980.

Schmidt, R.E., Scanlon, J.W., and Bell, J.B. Evaluability Assessment: Making Public Programs Work Better. Washington, D.C.: The Urban Institute, 1978 (Also published in Human Services Monograph Series, Project Share, No. 14, November 1979).

Shavelson, R. "Self Concept: Validation of Construct Considerations." In: Review of Educational Research, 46; 1976, 407-441.

Skutnabb-Kangas, T. and Toukomaa, P. "Teaching Migrant Children's Mother Tongue and Learning the Language of the Host Country in the Context of the Sociocultural Situation of the Migrant Family." Helsinki: The Finnish National Commission for UNESCO, 1967.

Spolsky, Bernard. "Language Testing--The Problem of Validation." In: Papers on Language Testing, edited by Leslie Palmer and Bernard Spolsky. Washington, D.C.: Teachers of English to Speakers of Other Languages, 1975.

Spolsky, Bernard. "What Does It Mean to Know a Language, or How Do You Get a Person to Perform His Competence?" In: Focus on the Learner, edited by J.W. Oller and T. Richards. Rowley, Massachusetts: Newbury House Publishers, Inc., 1973.

Spolsky, Bernard, et al. "Three Functional Tests of Oral Proficiency." Papers on Language Testing (1967-1974). Washington, D.C.: Georgetown University, 1975.

Stake, R.E. Program Evaluation, Particularly Responsive Evaluation. Kalamazoo, Michigan: Evaluation Center, College of Education, Western Michigan University, 1975.

Stake, R.E. "Measuring What Learners Learn." Urbana, Illinois: Center for Instructional Research and Curriculum Evaluation, University of Illinois, 1971(a).

Stake, R.E. "Testing Hazards in Performance Contracting." Phi Delta Kappan, 52, 1971(b), 583-589.

Stolz, W. and Bruck, M. "A Project to Develop a Measure of English Language Proficiency." Final Report to the National Center for Education Statistics, Contract No. 300-75-0253. Arlington, Virginia: Center for Applied Linguistics, June 15, 1976.

Storlie, T., et al. "Revised RMC Materials Committee Recommendations." Evanston, Illinois: Educational Testing Service, August 1978.

317

Streiff, Virginia. "Relationships Among Oral and Written Cloze Scores and Achievement Test Scores in a Bilingual Setting." Language in Education: Testing the Tests. Rowley, Massachusetts: Newbury House Publishers, Inc., 1978, 65-102.

Stufflebeam, D.L., Foley, W.J., Gephart, W.J., Guba, E.G., Hammond, R.L., Merriman, H.O., and Provus, M.M. Educational Evaluation and Decision-Making. Itasca, Illinois: Peacock, 1971.

Stufflebean, D.L. Meta-Evaluation. Kalamazoo, Michigan: Evaluation Center, College of Education, Western Michigan University, 1974.

Stufflebeam, D.L. and Webster, W.J. "An Analysis of Alternative Approaches to Evaluation." Educational Evaluation and Policy Analysis, 2, 1980, 5-20.

Stufflebeam, D.L. "The Use of Experimental Design in Educational Evaluation." Journal of Educational Measurement, Vol. 8, No. 4, Winter 1971.

Stump, Thomas A. "Cloze and Dictation Tasks as Predictors of Intelligence and Achievement Scores." Language in Education: Testing the Tests. Rowley, Massachusetts: Newbury House Publishers, Inc., 1978, 36-64.

Symes, Dal S. "A Description and an Analysis of Tests for the Bilingual Child." New Mexico State Department of Education. Santa Fe, New Mexico: Bilingual Teacher Training Unit, 1975 (ED 128-359).

Tallmadge, G.K. "Cautions to Evaluators." Achievement Testing of Disadvantaged and Minority Students for Educational Program Evaluation, edited by M.J. Wargo and A.R. Green. New York, New York: CTB/McGraw Hill, 1977.

Tallmadge, G.K. and Wood, C.T. User's Guide. Mountain View, California: RMC Research Corporation, 1978.

Technical Assistance Centers (TACs). Program Improvement Evaluation (PIE) Modules. Palo Alto, California: ESEA Title I Clearinghouse, 1981.

Texas Education Agency. Report from the Committee for the Evaluation of Language Assessment Instruments. 1977.

Ulibarri, Daniel M., Spencer, Maria L., and Rivas, Guillermo A. "Language Proficiency and Academic Achievement: A Study of Language Proficiency Tests and Their Relationship to School Ratings as Predictors of Academic Achievement." NABE Journal, 5, No. 3, Spring 1981.

United States Commission on Civil Rights. A Better Chance to Learn: Bilingual Bicultural Education. Washington, D.C.: May 1975.

318

Upshur, John A. "Objective Evaluation of Oral Proficiency in the ESOL Classroom." Papers on Language Testing (1967-1974). Washington, D.C.: Georgetown University, 1975, 53-65.

Valette, Rebecca. Modern Language Testing: A Handbook. New York, New York: Harcourt, Brace and Jovanovich, 1967.

Walters, Joel. "Language Variation in the Assessment of the Communicative Competence of Bilingual Children: Evidence For the Linguistic Interdependence Hypothesis." University of Illinois, 1979 (ED 174-040).

Wargo, M.J. and Green, D.R. Achievement Testing of Disadvantaged and Minority Students for Program Evaluation. New York, New York: CTB/McGraw Hill, 1977.

Wargo, M.J., Tallmadge, G.K., Michaels, D.D., Lipe, D., and Morris, S.J. "ESEA Title I: A Re-Analysis and Synthesis of Evaluation Data From Fiscal Year 1965 through 1970." Palo Alto, California: American Institutes for Research, March 1972.

Weiss, Carol H. "The Politicization of Evaluation Research." Evaluating Action Programs: Readings in Social Action and Education, edited by C.H. Weiss. Boston, Massachusetts: Allyn and Bacon, 1972(c), 327-338.

Weiss, Carol H. "Utilization of Evaluation: Toward Comparative Study." Evaluating Action Programs: Reading in Social Action and Education, edited by C.H. Weiss. Boston, Massachusetts: Allyn and Bacon, 1972(d), 318-326.

Weiss, Carol H. "Evaluation Research in the Political Context." In: Handbook of Evaluation Research, Vol. 1, edited by M. Guttentag and E.L. Struening. Beverly Hills, California: Sage Publications, 1975, 13-26.

Wholey, J.S. Evaluation: Promise and Performance. Washington, D.C.: The Urban Institute, 1979 (URI 25100).

Wisler, C.E. and Anderson, J.K. "Designing a Title I Evaluation System to Meet Legislative Requirements." Educational Evaluation and Policy Analysis, 1, 1979, 47-55.

Worthen, B.R. and Sanders, J.R. Educational Evaluation: Theory and Practice. Belmont, California: Wadsworth Publishing Company, Inc., 1973.

Wylie, R.C. The Self Concept: A Review of Methodological Considerations and Measuring Instruments. Revised edition. Lincoln, Nebraska: University of Nebraska Press, 1974.

Yavorsky, D. Discrepancy Evaluation: A Practitioner's Guide. Charlottesville, Virginia: University of Virginia, 1977.

3.19

Yeh, Mary. "The Issues in the Measurement of Bilingual Language Dominance." Paper presented at the Annual Meeting of the American Educational Research Association. San Francisco, California: 1976 (ED 121-853).

Zirkel, Perry A. and Brown, Mark E. "Emerging Instrumentation for Assessing Language Dominance." Occasional Papers in Linguistics, No. 1. Carbondale, Illinois: Southern Illinois University, Department of Linguistics, 1977 (ED 144-410).

320

SECTION II

OBSERVATIONS ON TESTING

This section contains reference material to provide program directors
and evaluators with a theoretical and practical background on testing
issues as well as a series of descriptions of several
testing/evaluation instruments. The material provides information on
the selection of achievement tests, language proficiency tests, and
self-concept scales. Included are abstracts and/or test summaries of
tests and scales often used in the evaluation of bilingual programs.
The documents in this section are included in order to make this core
of information readily available to program directors and evaluators,
thereby facilitating their evaluation activities.

An additional document found at the end of this section is a
presentation and discussion of ethnographic methods to develop a
program description.

INDEX TO SECTION III

OUT-OF-LEVEL OR FUNCTIONAL-LEVEL TESTING

Purpose

This document is designed to give teachers, parents and administrators a simple overview of the concept called functional-level testing. It can be used separately for an awareness of the topic or with other available resources to promote skills for matching test levels to student achievement levels.

The information presented in this document address five questions about functional-level testing. Each of these questions are identified and discussed in detail in the following sections.

How Do I Know If I Need to Use Functional-Level Testing?

When bilingual teachers evaluate the effectiveness of their projects, one piece of evidence to consider is the students' improvement on an achievement test. Students' scores at the beginning of the project can be compared to their scores at the end. This comparison will provide a true picture of the students' improvement if the teacher has accurate measures from the test.

A test that is too difficult or too easy may provide very little information about students' actual achievement levels. Students who are frustrated by a test that is too difficult may give up early, or

they may simply guess their way through the test. If a test is too easy, students will find it unchallenging. In either case, test scores will not provide an accurate indication of their achievement level. Such results are a major concern of teachers, especially when they realize before testing that most items on a test are too difficult or too easy for some students.

Functional-level testing is an alternative that can be helpful in situations like these. Because functional-level testing results in improved information by matching a student's ability with the difficulty of the test he or she takes, it has been recommended in current evaluation guidelines. This paper will provide information about functional-level testing; what it requires, and how it can be implemented.

What Are Test Levels?

Many initial questions need answers when a commercial publishing house plans development of an achievement test battery. For example, the publisher must determine which basic topical areas will be measured, what span of grades the test should cover and the length of time required for test administration. Other major considerations include reading and vocabulary levels of the test items, specific content to be covered within the given topical areas, and the relative interest and difficulty of the material on which the test will be based. In weighing these considerations, the publisher understands that a single test covering all grades would be much too long and inefficient to

administer to any student (see Figure 1). One solution is to publish

a series of tests, each known as a _level_.* A test level may be

defined as one of a number of strata, the content and difficulty of

which are appropriate to a given grade or span of grades. Note that

in Figure 2 Level C covers a span of grades from the second through

eighth.
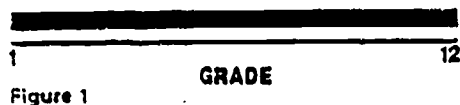
Figure 1
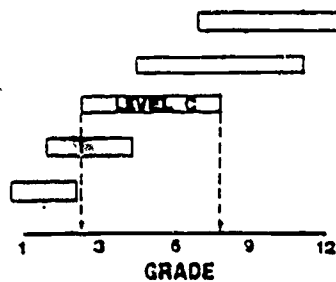
Figure 2

---

* The term _level_ must be clearly distinguished from the term _form_.
Form, more appropriately termed _alternate form_ or _equivalent form_, is
a second test at a given level designed to measure the same content
using a different, but equivalent, set of test items.

Next, the publisher must decide how many levels the planned series should incorporate. If only a few levels will be developed, each must cover the content for several grades. For example, the test levels shown in Figure 3 cover a broader range of grades than those in Figure 4. The broader the content covered by a given level, the more likely that item content and difficulty will be appropriate for the low or high achieving students -- perhaps both. Narrow content coverage within a level may be more relevant for a single grade (see Figure 4). However, focusing on such narrow content coverage can result in too many tests and be too costly.

TEST LEVELS

Figure 3

Figure 4
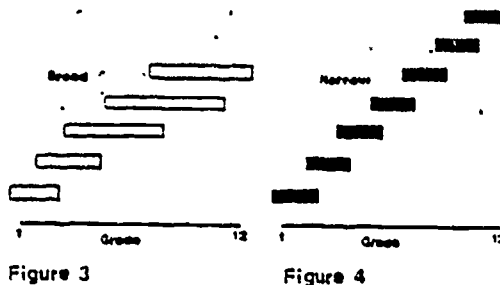
After considering which topics to cover, what content to cover within topics and the difficulty of content, test publishers select items for inclusion in a test series. Each level, designed for typical students in a given grade or span of grades, is known as the recommended level. There is not always one test level for each grade level. Sometimes a test level spans two or more grades.

326

What is Functional Testing?

Whenever a student is given a test level appropriate for his or her functioning educational level, it is considered functional-level testing. Most students' functional levels are best served when the test publisher's recommended level for their grade is administered. This practice is known as in-level testing. Functional-level testing, however, allows testing at, below or above the publisher's recommended level.

What is Out-of-Level Testing?

The recommended level of a test does not always contain the most appropriate content or difficulty for students with very low or very high performance levels. When testing such students, it may be desirable to administer a test level other than the specific level recommended by the test publisher for typical students in that grade. This practice, called out-of-level testing, is employed when the recommended test level is expected to be much too easy or too difficult for the students.

The use of tests at levels below those recommended by the publisher is an option if the content of the program can be measured better this way. Students in bilingual programs may be learning skills, such as English reading, at a later time than other students and therefore, should receive the same test at a later point. In order for any test to be suitable, the average score of the group tested should be

327

between 1/3 and 3/4 of the maximum (Roberts, 1976). Otherwise, ceiling or floor effects depress estimates of student gains. Some publishers provide norms for the administration of a single test in several grades. Other publishers provide expanded standard scores that link up all levels of a test on a common scale, and occasionally, locator tests, to facilitate out-of-level testing. Generally, a test should be used no more than one level below that recommended by the publisher. But care should be taken that in testing out-of-level, pretest floor effects are not being replaced by posttest ceiling effects.

Why Test Out-of-Level?

Achievement testing is used to obtain a reliable and valid measure of student achievement. Factors contributing to unreliable and invalid test scores may include test administration procedures (e.g., adhering to timing and directions), physical surroundings (e.g., spacing of chairs, temperature, lighting, etc.), student characteristics (e.g., motivation, physical well-being, etc.), and test characteristics (e.g., difficulty level, content, format, etc.).

Although functional-level testing does not address all of these concerns, it does consider test characteristics and has the potential to affect students' motivation. Test characteristics of content and difficulty level are very important. For example, consider test content. Different levels of a test series emphasize different skills and the content can be quite different even though the subject area

remains the same. A selected test level should match the content
material taught. If a test does not match what students are being
taught it will not be sensitive to learning and gains which actually
occur may not be shown.

Now consider test difficulty. When a test is too difficult for a
student, guessing is likely to occur, creating problems for both the
reliability and validity of the resulting test scores. In turn, the
assessment of student achievement and the evaluation of programs are
affected. Guessing increases most students' scores in multiple-choice
tests. Some students' entire scores can be a reflection of the luck
involved in random guessing. The laws governing these scores based
upon random guessing are the same as those governing who wins and who
loses at Las Vegas; consequently, they are known as chance scores.
For example, if a group of students were to take a 100-item test with
four options per item, and randomly guess at all items, the average
score for the group would be approximately 25. Obviously, chance
scores do not provide accurate information about a student's level of
skill development. Students whose scores are primarily a result of
guessing on a test that is too difficult may need to be tested with an
easier, lower level of the test. In Figure 5 we see that the students
scoring in the chance range (shaded area) of Test Level C may need to
be tested at a lower test level -- Level B in this case.

Chance Score Range **LEVEL C**

Test at Lower Level ▼
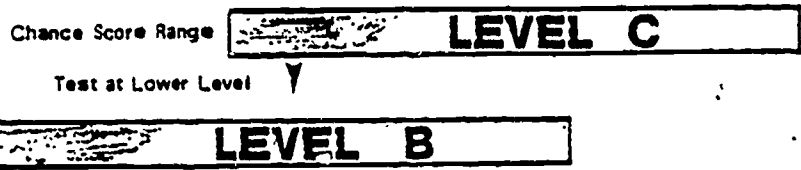
LEVEL B

Figure 5 For students scoring in the chance score range the test level was probably too difficult. They may need to be tested with a lower level of the test.

On the other end of the achievement spectrum there are students for whom the test level is too easy, limiting such students' ability to demonstrate their skill development. They too may need to be tested out-of-level, but with a more difficult test. The shaded area in Figure 6 depicts the high range for two test levels. Students scoring in the high range of Level C may heed to be tested with a more difficult level of tests--Level D in this case.

LEVEL D

▲ Test at Higher Level

LEVEL C   High Score Range
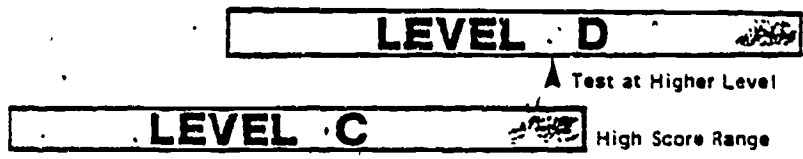
Figure 6 Students scoring in the high range of a test level may need to be tested with a higher level of the test.

33

In addition to being misleading about a student's true skill level, a test that is either too easy or difficult can misrepresent student gains in achievement. Consider the following diagram in which the line at the bottom represents all there is to know about a certain topic and the lines above indicate the portions of the topic covered by various test levels.

Zero Knowledge   C   A   B   100% Mastery

Apparent Gain (B-A)
Actual Gain (B-C)

331
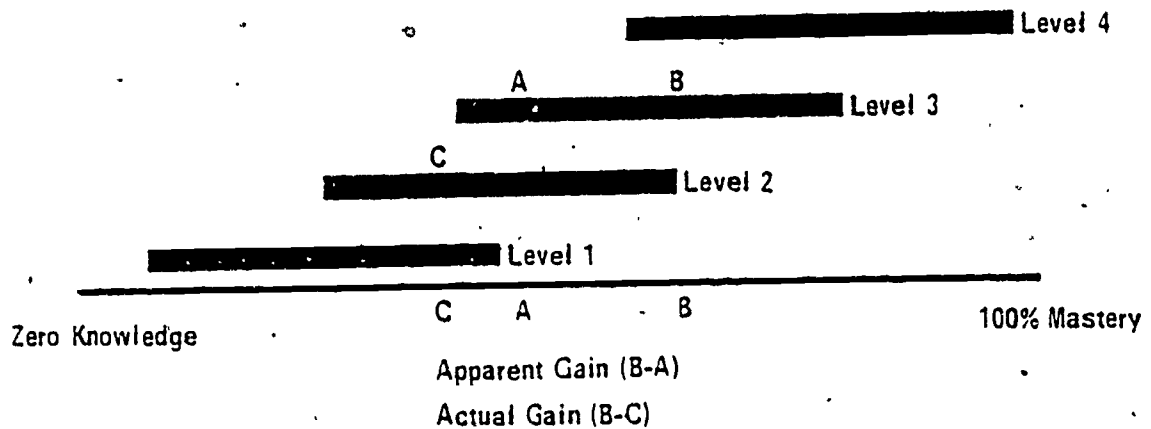
Suppose a group of students is given Level 3, but it is too difficult for them. They may guess on may items and score in the chance range of the test--at point A. Let's assume their posttest performance would show improvement, and they would score at point B. Their apparent gain is the distance between the pretest and posttest (B-A). (However, if students had been tested at their functional level, Level 2 probably would have been given at the pretest. Guessing would be less a factor since the test difficulty at this lower level is more closely matched to student achievement. Their score may have been something near point C. So their actual gain from pretest to posttest is B-C.

In summary, the recommended test level for the average student in a certain grade may not accurately measure the achievement level of every student in that grade. Some students will function at a higher achievement level, some at a lower level. In either case, out-of-level testing could provide a better measure of student achievement.

332

SUMMARY

Achievement tests provide useful information for evaluating the effectiveness of programs. The value of such information is obviously related to its accuracy. Achievement tests are designed to accurately measure the achievement level of average students in a certain grade level. However, they may not accurately assess the achievement level of _all_ students at that grade level.

A student's functional level may be below a test publisher's recommended test level. And a very low test score on a recommended test level may indicate that guessing (chance) played an important role in the result. Students whose scores are primarily a result of guessing on a test that is too difficult may need to be tested out of level; tested with an easier, lower level of the test.

Functional-level testing, therefore, involves testing students with test levels most appropriate to their achievement levels. Functional-level testing can involve testing students with the recommended test level (in-level testing), _or_ it can mean testing students with a test level below or above the recommended level (out-of-testing). Whatever the case, the goal is to test at a level affording students the most opportunity to demonstrate their abilities.

SELECTING AN ACHIEVEMENT TEST*

In selecting achievement tests for the evaluation of bilingual programs, evaluators must consider all the same criteria that are used in selecting any achievement test as well as additional criteria that relate to the nature of the program and the student population. This discussion will give most emphasis to issues in test selection that are especially important for bilingual education evaluations.

Test Bias

During the last ten years extensive attention has been given to the effects of test bias for culturally different populations (Wargo, 1977; Houts, 1974). As a result, test publishers have made concerted efforts in this area and many standardized achievement tests have been revised. The technical manual of a test will often include a discussion of what procedures were undertaken to minimize bias. The two most common procedures are: (1) review of the content of the items by a culturally sensitive panel and (2) statistical item analyses.

---

* Adapted with author's permission from: "An Evaluation of Project Information Packages (PIPs) As Used for the Diffusion of Bilingual Projects," RMC Research Corporation, U.S. Department of Health, Education and Welfare, Office of Education, 1980.

Review of Content -- Reading and examining the content of items may result in rewriting items so that they seem fairer to all groups involved. However a visual examination alone cannot determine if an item is biased, i.e., that it will function differently for different groups of students. What can be accomplished is the elimination of stereotypical wording or content. External review panels have the advantage of insuring a disinterested reading, although in-house groups may also be effective. This procedure may result in a more acceptable test, but will not necessarily eliminate biased items.

Item Analysis -- Item analysis is a statistical procedure that is performed routinely in test construction. The scores of students on each item are compared to their scores on the whole test in order to determine if each item is measuring what the whole test measures, and in fact should be part of that test. When this procedure is used to eliminate bias towards a specific group, the test is administered to both the general population and to the specific group. Then item analysis is performed in order to determine that the same items function similarly for both groups. For example, if an item is difficult for one group it should be difficult for the other regardless of the mean test scores for each group. If an item is easy for one group but difficult for another, then such an item exhibits bias, and should probably be eliminated.

Additional Selection Issues

Consideration of subtest content and weight in scoring is important for selecting the test that most closely matches the curriculum and for determining whether in-level testing is appropriate. Such issues are important for all students, but they may be even more critical for students of limited English proficiency. Although the curriculum of bilingual programs may contain the same final objectives, skills such as English reading may not be taught in the same grade levels as other programs.

The wording of the instructions to the test should be considered. The language of the instructions should not be more difficult than the language used in the items that actually appear in the test. Although directions containing needlessly complex sentence structures are a handicap for all students, they will cause an even greater difficulty for students of limited English proficiency. Examiners may want to consider systematically simplifying test directions, but if norms are to be used, this may affect their validity.

Additionally, the content of the test should be examined to determine the extent to which it tests the out-of-school experience of the children. The experience of the culturally different child and of the low SES child may differ significantly from that assumed by the authors of the test. Therefore, the more the test relies on out-of-school experience, the more it may discriminate against the

target population and the less valid it will be for evaluating program impact.

Finally, if bilingual tests are used, the nature of the translation should be considered. Some tests are direct translations except where such a translation would clearly be impossible. Other tests provide equivalent versions where the kinds of items and the difficulty level are roughly equivalent, but the content of the item may be completely different. Other tests are a combination of both methods. In a translated test, the difficulty level may not be the same for both versions. However, very few test publishers provide equivalent versions.

Language of Testing

In many bilingual education evaluations, the evaluator must decide what testing language is appropriate. Several questions have to be considered individually and in relation to each other. First, what is the language of instruction for the subject that will be tested? Because the language of instruction for math, for example, may be different for students in the same class or may be different at various times during the year, this question may not be answered simply. Second, what is the dominant language of the child as established by a systematic assessment procedure? Third, what are the project goals? Goals may require testing in a particular language. Ideally, of course, students should be tested in the language in which

337

they will peform the best. However, that language may not always be

the dominant one. For example, a student may be more fluent in

Spanish, but if almost all math instruction has been in English, the

student may perform better on an English test.

There are other issues involved in planning testing in more than one

language that have not yet been studied in sufficient detail. Some

evaluators double-test the project students, avoiding the choice of

test language by testing in both languages. The benefits of this

practice are clear: more information is obtained about the students'

proficiency in content and language and the dangers of testing only in

the weaker language are avoided. However, the additional expense, the

added burden on teachers and students, and the possibility of practice

effects represent significant disadvantages. In addition, the

language of some students may be neither standard English nor standard

Spanish.

Where tests exist in two languages, the non-English language may be

the most appropriate language for the pretest. However, after a year

of English instruction, English may be more appropriate for the

posttest. Longitudinal studies will almost certainly include scores

in both languages reported at different stages of a student's

progress. Evaluators will have to consider carefully the

interpretations of such scores.

Limits to the Usefulness of Norms

The use of national norms as a comparison standard in an evaluation relies on the validity of a principle known as "the equipercentile assumption." This assumption implies that in the absence of any special instructional treatment students in the project would have grown at a rate comparable to that of students in the norming sample who obtained the same mean pretest value. Such an assumption can only be valid if the project population is similar in educationally relevant ways to the population represented in the norming sample. This is not usually the case in bilingual education programs which are generally comprised of students of limited English proficiency, bilingual students, and a larger proportion of low SES students than is found in the general population. While the accuracy of the equipercentile assumption for such populations has not yet been systematically assessed, it is unlikely that norms for English achievement tests can provide precise no-treatment expectations for bilingual project students. There are no statistical techniques to adjust for differences in expected growth between the project students and the norming population (Tallmadge, 1976).

Recently, data have been gathered on Spanish language achievement tests. The most recent editions of the Comprehensive Test of Basic Skills (CTBS) and the Inter-American Series both furnish norms tables for English and Spanish versions of their tests, but the manner in which such norming data were compiled limits their usefulness for

evaluating the impact of bilingual projects. The CTBS Español norms were developed by administering the CTBS in both languages to a balanced bilingual, biliterate population as determined by scores on the SERVS test. The assumption was made that a student's standing in the norms would be the same in English and Spanish. Student's scores in Spanish were then equated with their rank in the English norms. Although the assumption that a perfectly bilingual person will possess the same knowledge of content in two languages is logical, the possibilities for error are so large that the Spanish norm conversions can provide only very rough estimates of student achievement. There are several other reasons why the CTBS norms cannot be used to provide a precise estimate of project impact. Because the scores in the norms table are extrapolated rather than derived empirically, they are subject to a certain amount of error inherent in any estimation procedure. In addition, the balanced bilingual population in the sample is not comparable to the population of most bilingual programs which include students with a range of language proficiences. Finally, because the students in the sample were in bilingual programs, they do not provide an estimate of how similar students would have performed without any special instruction.

The Inter-American norms were not constructed from a national probability sample. They are "user norms" derived only from those groups in the population to whom the Inter-American tests were administered in the course of local evaluations. For certain tests, the sample obtained in this way numbers over a thousand students, but

340

for others the N is less than 100, severely limiting the reliability of normative data, particularly in the extreme score ranges where estimates are based on relatively few cases. Because the norming group was not specifically constructed to represent the population of limited English and bilingual students, unknown biases may exist in the sample. Because students in the sample are also in bilingual programs, the norms do not provide an estimate of how similar students would have performed in the absence of a special program.

The question of how a group of students would have performed without a bilingual project cannot be answered by simply consulting currently available norms. But existing norms can be used to answer other evaluation questions. Well constructed norms based on national probability samples, such as those provided by the major achievement tests, can be used to show how the bilingual project students compare to national averages. Norms based on more specific populations, such as those constructed for the Spanish versions of the CTBS and the Inter-American, can be used to show how project students compare to the bilingual/biliterate CTBS sample or the bilingual project students in the Inter-American sample.

Out-of-Level Testing -- The use of tests at levels below those recommended by the publisher is an option if the content of the program can be measured better this way. Students in bilingual programs may be learning skills, such as English reading, at a later time than other students and therefore should receive the same test at

a later point. In order for any test to be suitable, the average score of the group tested should be between 1/3 and 3/4 of the maximum (Roberts, 1976). Otherwise, ceiling or floor effects depress estimates of student gains. Some publishers provide norms for the administration of a single test in several grades. Other publishers provide expanded standard scores that link up all levels of a test on a common scale, and occassionally, locator tests, to facilitate out-of-level testing. Generally, a test should be used no more than one level below that recommended by the publisher. But care should be taken that in testing out-of-level, pretest floor effects are not being replaed by posttest ceiling effects. (Note: This topic is discussed further in a preceding document entitled "Out-of-Level or Functional-Level Testing.")

Introduction to Test List and Summaries

An extraordinary number of tests could be used to evaluate basic subject areas for bilingual programs. Some of these tests are locally developed and have not been administered to large samples of the population. Therefore, they are less likely to have the technical qualities required by most evaluators. Other tests are limited to only one content area, and cannot be used by themselves to evaluate a bilingual project which includes several content areas. Finally, many evaluators will first consider the appropriateness of tests already in use in the district for the evaluation of the bilingual program. Certain tests may be mandated or choices may be constrained in other

312

ways. Selection of a test already being used for district-wide
assessment introduces the possibility of comparison with local
non-project students. This comparison alone cannot provide a precise
estimate of project impact, but may answer other evaluation questions,
such as how project students compare in achievement level and rate of
growth to other students in the district.

The following sections of this document are intended to provide
helpful information about tests that, for the reasons discussed above,
are already likely to be under consideration by project evaluators.
First, an annotated test list is presented which includes information
about major tests of achievement that include both math and reading or
language subtests which are available in two languages. Second, a set
of achievement test summaries, developed by the Region V Technial
Assistance Center (TAC), are provided as an additional information
resource for program directors and evaluators. Finally, a list of
publishers is provided for future reference for program directors and
evaluators seeking additional achievement test information.

Annotated Test List

The annotated test list contains only major tests of achievement that
include both math and reading or language subtests. All such tests
available in two languages were included. Tests only available in
English were limited to those included in the Anchor Test Study
(Loret, 1974). Finally, all of the tests were discussed only as they

$3 \cdot 3$

apply to evaluations of grades K-6.

The same categories of information are provided for each test to
facilitate comparison. All of the tests are available from major
publishers. Technical aspects of such tests are likely to be as good
as the state-of-the-art. All of the tests have technical manuals
describing the process of test construction and standardization.
Except for an occasional subtest, all of the tests are designed to be
administered in groups. Administration time for each test varies
according to the number of subtests used. Subtests are listed only
where they contribute to a total score in reading, language arts, or
mathematics, three major areas of interest to bilingual program
evaluation.

EL CIRCO
1979

1. Languages: Spanish and English

   Spanish tests allow the test administrator to select among alternatives the word most appropriate for the students' dialect of Spanish.

2. Publisher's recommended in-level use: Tests can be used at pre-school, kindergarten, and beginning of first grade.

3. Subtests:*

   Cuanto y Cuantos

   Para Que Sirven Las Palabras

   What Words are For

   Quanto y Quantos is a direct translation of Level A of How Much and How Many of CIRCUS. Para Que Sirven Las Palabras and What Words are For are equivalent, but one is not a translation of the other. For example, each test has items testing comprehension of the past tense but the items will have different content.

4. Norming: The El Circo measures were administered to a nationwide sample of children from the Spanish-speaking cultural groups. Empirical norms exist for fall only.

5. Out-of-level testing: Separate norms exist for preschool, kindergarten, and first grade.

6. Procedures for minimizing bias: Items were reviewed by a cultural advisory committee composed of speakers of Puerto Rican, Mexican, and Cuban Spanish.

---

*Several tests have been developed as part of El Circo, but only the ones listed are available as of spring 1980.

California Achievement Test, 1977-78
Forms C and D

1. Languages: English

2. Publisher's recommended in-level use:

Level	Grade
Level 10	K.0-K.9
Level 11	K.6-1.9
Level 12	1.6-2.9
Level 13	2.6-3.9
Level 14	3.5-4.9
Level 15	4.5-5.9
Level 16	5.5-6.9

3. Subtest Components:

	Level: 10	11	12	13	14	15	16
Pre-Reading							
Listening for Information	X						
Letter Forms	X						
Letter Names	X						
Letter Sounds	X						
Visual Discrimination	X						
Sound Matching	X						
Reading							
Vocabulary		X	X	X	X	X	X
Comprehension		X	X	X	X	X	X
Phonic Analysis		X	X	X			
Structural Analysis			X	X			
Language Total							
Language Mechanics			X	X	X	X	X
Language Expression			X	X	X	X	X
Mathematics Total							
Computation			X	X	X	X	X	X
Concepts and Applications			X	X	X	X	X	X

4. Norming: Weeks rather than midpoint dates are provided for empirical fall and spring norms. These are the week in which November 3rd falls, and the week in which May 4th falls. Tests can be administered two weeks on either side of these weeks without the use of interpolated norms.

5. Out-of-level testing: Provides an expanded standard score scale and a locator test.

6. Procedures for minimizing bias: Test writers followed guidelines to avoid bias in the development and editing of items. Items were reviewed by of various ethnic and cultural groups. An extensive item analysis was conducted with the tryout items to compare responses of "Black" students and "other" students. A point biserial correlation was used to show the relation of items to category objective scores, and grade-to-grade growth as shown by item difficulties was also examined. The percent of biased items found in the trial items for the various subject areas ranged from 25 to 7 percent. After revision the percent of biased items was reduced to the 3-0 percent range.

317

CIRCUS
1976

1. ·Languages: English

2. Publisher's recommended in-level use:

Level	Grade
Circus A	Nursery School and Kindergarten - Fall
Circus B	Kindergarten - Spring
	First Grade - Fall
Circus C	First Grade - Spring
	Second Grade - Fall
Circus D	Second Grade - Spring
	Third Grade - Fall

3. Subtests:*

	Level			
	A	B	C	D
Pre-Reading		X		
Reading			X	X
Listen to the Story	X	X		
Listening			X	X
How Much and How Many	X	X		
Mathematics			X	X
Writing Skills				X

4. Norming: The Circus was administered to a national probability sample during the fall (October) only. Therefore, the comparison of a group to the national sample for pre- and posttesting can be done for a fall-to-fall evaluation design only. Information is also provided in sentence form describing what each range of scores means in terms of skills mastered. A fall to spring comparison of the proportion of students falling in each category could be made, but would require the use of a local comparison group to determine the normal growth expectation. Separate tables exist for comparing groups and for comparing individuals.

---

*Many other subtests are provided, but only these that coordinate with the STEP are listed here. No total scores are possible from any combination of subtests.

The subtests listed above provide coordination through content and expanded standard scores with the following subtests of STEP III, Level E-J; Reading, Listening, Math Concepts and Math Computation, and Writing Skills.

The normative data are very well suited to individual student
evaluation because the national sample is divided into subgroups
such as sex, geographic region, and SES.

5. Out-of-level testing: Expanded standard scores can be used for
    subtests that coordinate with STEP III.

6. Procedures for minimizing bias: No statistical procedures are
    reported. Separate norms are provided according to categories
    such as sex, geographic region, and SES.

3 19

Comprehensive Test of Basic Skills
English Version 1973, Spanish Version 1978
Form S

1. Languages: English and Spanish

The CTBS/Español is a direct translation of the English CTBS/S
with the exception of certain items which could not be translated
or which required different translations for dialects of Spanish.
In such cases equivalent items have been constructed.

Publisher's recommended (in-level testing):

	English CTBS/S	CTBS/Espanol
Level B	Grades K.6-1.9	Grade 1
Level C	Grades 1.6-2.9	Grade 2
Level 1	Grades 2.5-4.9	Grades 3 and 4
Level 2	Grades 4.5-6.9	Grades 5 and 6

3. Subtest components:

	Level			
Component	B	C	1	2
Reading				
Word Recognition	X			
Reading Vocabulary		X	X	X
Reading Comprehension	X	X	X	X
Mathematics				
Math Computations	X	X	X	X
Concepts & Applications	X	X	X	X

4. Norming: The norms for the Spanish version of the CTBS were
derived through a spring testing equating this version with the
nationally representative English language norms. The
no-treatment expectation obtained by their use is not referenced
to a Limited English Proficiency population but rather to the
English language performance that could be expected from the
bilingual/biliterate population on whom the equating was done.
The scoring patterns in both English and Spanish for limited
English proficiency students may be quite different; therefore,
the norms do not present a precise standard of comparison.
Empirical norms exist for the English CTBS for spring for grades
2-6, and for fall and spring for grades K and 1.

5. Out-of-level testing: An expanded standard score scale is
available for the CTBS/S norms.

6. Procedures for minimizing bias: Prior to standardization these items were reviewed by Black and Spanish-speaking consultants. In addition, trial items were administered to a sample of Black students and "other" students. Items with a point-biserial coefficient of less than .2 were rejected. A subsequent analysis was made of the test results of Black students, Spanish-speaking students, and other students. Although the mean scores were lower for the Black and Spanish-speaking group, the tests appeared to be functioning similarly for both groups.

351

Inter-American Series: Test of Reading, 1962-69
Forms CE, DE, CEs, DEs

1. Languages: English, Spanish, and French

   Spanish version is an exact translation of English version.

2. Publisher's recommended in-level use:

   Level 1   Grade 1.5-2.5
   Level 2   Grade 2.5-3.9
   Level 3   Grades 4, 5, 6

3. Subtest components:

	Level		
Components	1	2	3
Vocabulary	X	X	X
Comprehension	X		
Level of Comprehension		X	X
Speed of Comprehension		X	X

4. Norming: The Inter-American norms were not developed using a probability sample. They are based on data collected from test users. The test manual states that these norms "should be applied with caution until local norms can be developed." Although N's for some tests consist of more than a thousand students, others comprise less than a hundred students. For these reasons, the norms do not provide a convincing, precise standard of comparison.

5. Out-of-level testing: Norms are provided for out-of-level testing; however, above comments regarding norms should be taken into account.

6. Procedures for minimizing bias: Content was selected as being familiar to English and Spanish speakers of the Western Hemisphere. A semantic frequency list was consulted in wording the translation, but the manual states that frequency is not always an indication of difficulty level. Spanish trial items were administered to Spanish speakers, and English trial items were administered to English speakers. Item analysis and item selection were then performed on the basis of test results.

352

Inter-American Series: Test of General Ability, 1961-72
Forms CE, DE, CEs, and DEs

1. Languages: English and Spanish

Spanish version is an exact translation of English version.

2. Publisher's recommended in-level use:

Preschool Level   Ages 4 and 5
Level 1   Grades end K, Grade 1
Level 2   Grades 2, 3
Level 3   Grades 4, 5, 6

3. Subtest components:

Components	Level			
	Pre-School	1	2	3
Oral Vocabulary	X	X	X	
Number	X	X	X	
Association	X	X		
Classification	X	X	X	X
Analogies			X	X
Sentence completion				X
Computation				X
Word Relations				X
Number Series				X

4. Norming: The Inter-American norms were not developed using a probability sample; the norms are based on data collected from test users. The test manual states that these norms "should be applied with caution until local norms can be developed.". Although N's for some tests consist of more than a thousand students, others comprise less than a hundred students. For these reasons, the norms do not provide a convincing, precise standard of comparison.

5. Out-of-level testing: Norms are provided for out-of-level testing; however, the above comments regarding norms should be taken into account.

6. Procedures for minimizing bias: Content was selected as being familiar to English and Spanish speakers of the Western Hemisphere. A semantic frequency list was consulted in wording the translation, but the manual states that frequency is not always an indication of difficulty level. Spanish trial items were administered to Spanish speakers, and English trial items were administered to English speakers. Item analysis and item selection were then performed on the basis of test results..

353

IOWA Tests of Basic Skills, 1978
Forms 7 and 8

1. Languages: English

2. Publisher's recommended in-level use:

Level	Grade	Forms
Primary Battery 5	K.1-1.5	7
Primary Battery 6	K.8-1.9	7
Primary Battery 7	1.7-2.6	7
Primary Battery 8	2.7-3.5	7
Multilevel Battery 9	3	7 and 8
Multilevel Battery 10	4	7 and 8
Multilevel Bat ery 11	5	7 and 8
Multilevel Battery 12	6	7 and 8

3. Subtest components:

	Level 5	6	7	8	9	10	11	12
Reading								
Reading Comprehension					X	X	X	X
Pictures			X	X				
Sentences			X	X				
Stories			X	X				
Reading		X						
Vocabulary	X	X	X	X	X	X	X	X
Math								
Math Concepts			X	X				
Math Problems			X	X				
Math Computations			X	X	X	X	X	X
Math	X	X						
Language								
Spelling			X	X	X	X	X	X
Capitalization			X	X	X	X	X	X
Punctuation			X	X	X	X	X	X
Usage			X	X	X	X	X	X
Language	X	X						
Listening	X	X	X	X				

4. Norming: Empirical norms exist for 15 October and 15 April.

5. Out-of-level testing: An expanded standard score scale is provided.

6. Procedures for minimizing bias: Authors with diverse cultural backgrounds participated in writing of test.

Sequential Tests of Educational Progress
(STEP) III, 1979, Forms X and Y

1. Languages: English

2. Publisher's recommended in level use:

Level	Grade
Intermediate E	.3.5-4.5
Intermediate F	4.5-5.5
Intermediate G	5.5-6.5

3. Subtest components:

	Level		
	E	F	G
Reading Total			
Vocabulary	X	X	X
Comprehension	X	X	-X
Inference	X	X	X
Math			
Mathematics Basic Concepts	X	X	X
Mathematics Computations	X	X	X
Language: Writing Skills			
Spelling	X	X	X
Capitalization	X	X	X
Word Structure and Usage	X	X	X
Sentence and Paragraph Organization	X	X	X
Language: Listening			
Listening Comprehension	X	X	X
Following Directions	X	X	X

4. Norming: Empirical norms are available for fall and spring. Midpoints of the norming periods are 5 October and 10 May.

5. Out-of-level testing: Provides expanded standard score scale and also out-of-level norms. Has locator test.

6. Procedures for minimizing bias: Items were edited by in-house minority and women test specialists, and by an external minority review panel.

7. Additional comments: Can be used in conjunction with CIRCUS, 1978, because of the coordination of test content and an expanded standard score scale.

355

Metropolitan Achievement Tests
(MAT) 1978 Forms J1 and K1

1. Languages: English

2. Publisher's recommended in-level use:

Level	Primary
Primer	K.5-1.4
Primary 1	1.5-2.4
Primary 2	2.5-3.4
Elementary	3.5-4.9
Intermediate	5.0-6.9

3. Subtest components

	Primer	Primary 1	Primary 2	Elementary	Intermediate
Reading Comprehension*	X	X	X	X	X
Language					
Listening Comprehension	X	X	X	X	
Punctuation and Capitalization		X	X	X	X
Usage		X	X	X	X
Grammar and Syntax		X	X	X	X
Spelling	X	X	X	X	X
Study Skills	X	X	X	X	X
Math					
Numeration	X	X	X	X	X
Geometry and Measurement	X	X	X	X	X
Problem Solving		X	X	X	X
Operations: Whole Numbers	X	X	X	X	X
Operations: Laws and Properties				X	X
Operations: Fractions & Decimals					X
Graphs & Statistics					X

4. Norming: Empirical fall and spring norms have been developed with midpoints of 15 October and 20 April respectively.

---

* Additional reading subtests such as rate and auditory discrimination are available, but they are not part of the comprehension score.

5. Out-of-level testing: Provides an expanded standard score scale. Out-of-level testing should be no more than one level below that recommended for the grade.

6. A combination of objective and subjective methods was used to identify ethnically biased items on the MAT. Following review by a panel of ethnicaly diverse educators, test items were examined for bias using three conceptually different statistical methods. Items tagged as biased by either the subjective or objective procedures were subsequently revised or eliminated.

357

SRA Achievement Series, 1978
Forms 1 and 2

1. Languages: English

2. Publisher's recommended in level use:

Level	Primary
A	K.5-1.5
B	1.5-2.5
C	2.5-3.5
D	3.5-4.5
E	4.5-6.5

3. Subtest components:

Component	A	B	C	D	E
Reading					
Visual Discrimination	X				
Auditory Discrimination	X	X			
Letters/Sounds	X	X	X		
Listening Comprehension	X	X	X		
Vocabulary		X	X	X	X
Comprehension		X	X	X	X
Mathematics					
Concepts	X	X	X	X	X
Computation		X	X	X	X
Problem Solving					X
Language Arts					
Mechanics			X	X	X
Usage			X	X	X
Spelling			X	X	X

4. Norming: The norms are based on a nationally representative sample of students. Empirical spring norms are available with temporary fall interpolated norms. Empirical fall norms are currently being developed. Empirical fall and spring norming dates are: 7 October and 25 April.

5. Out-of-level testing: Out-of-level testing can be interpreted using the SRA expanded standard score scale known as GSV (Growth Scale Value).

6. Procedures for minimizing bias: Items were edited by representatives of minority and women's groups. The trial items were administered to a sample that included Black, Hispanic, American Indian, and non-minority subsamples. The items were then examined statistically and items which were easy for one group but difficult for another were eliminated.

353

Stanford Achievement Test, 1973
Forms A, B, and C

1. Languages: English

2. Publisher's recommended in-level use:

Level	Primary
Primary I	1.5-2.4
Primary II	2.5-3.4
Primary III	3.5-4.4
Intermediate I	4.5-5.4
Intermediate II	5.5-6.9

3. Subtest components:

	Primary I	Primary II	Primary III	Intermediate I	Intermediate II
Total Reading					
Reading Comprehension	X	X	X	X	X
Word Study Skills	X	X	X	X	X
Total Mathematics					
Concepts	X	X	X	X	X
Computation and Applications	X				
Computation		X	X	X	X
Applications		X	X	X	X
Total Auditory					
Vocabulary	X	X	X	X	X
Listening Comprehension	X	X	X	X	X

4. Norming: Empirical norms are available with a midpoint of 8 October for grades 2-9, and 8 May for grades 1-9, and 8 February for grades 1 and 2.

5. Out-of-level testing: Provides an expanded standard score scale. Testing more than one level out-of-level is not recommended.

6. Procedures for minimizing bias: Items were edited by a group of consultants with various minority backgrounds.

7. Other comments: The scaled score is continuous with Stanford Early School Achievement (SESAT) and Stanford Test of Academic Skills (TASK).

Test of Basic Experience II
(TOBE) 1978

1. Languages: English and Spanish

   The Spanish version is a direct translation from the English with the exception of items that would radically change in translation. In such cases equivalent items were constructed. The Spanish version of the test occasionally provides a choice of words so that the most common version of words can be used with Mexican, Cuban, and Puerto Rican students.

2. Publisher's recommended in-level use:

Level	Grade
K	Preschool, kindergarten, fall of first grade
L	Spring of kindergarten, first grade

3. Subtests:

	Level K	Level L
Mathematics	X	X
Language	X	X
Science	X	X
Social Sciences	X	X

4. Norming: Empirical norms exist only for the English version of the test; midpoints are October 19 and April 19.

5. Out-of-level testing: Provides expanded standard ore scales.

6. Procedures for minimizing bias: Test items were reviewed by a panel of women and minority consultants. The Spanish version of the test was reviewed by native speakers of Puerto Rican, Cuban, and Mexican Spanish.

367

Achievement Test Summaries

Test information summaries were developed by the Region V Technical Assistance Center (TAC) to serve as an information resource for evaluators. The information included in these summaries focuses on the use of norm-referenced tests in Reading, Language Arts, and Mathematics. Test Information Center staff of Region V TAC prepared 42 summaries (19 are included in this document) in response to requests for information about the tests. Preparation of these summaries was not intended to imply endorsement or approval of any test.

The test summaries, intended to serve as a guide to the use of the publishers' test and test publications, may be used for various purposes such as: test familiarization and/or selection; identifying the test publications which provide information on norming, reliability, and validity; selecting appropriate test levels for functional level testing; scheduling test administrations; and identifying the publications which contain the required norms tables as well as the names of the specific score conversion tables to be used.

Test summaries are revised on a periodic basis as new test information becomes available. The following summaries were most recently revised in early 1981 and were reviewed by the respective test publishers.

1. Test (Series)/Year: CALIFORNIA ACHIEVEMENT TESTS, 1970

    Forms: A and B

    Publisher/Distributor: CTB/McGraw-Hill

    Authors: E.W. Tiegs and W.W. Clark

    Description: A series of academic achievement test batteries
    designed for measurement, evaluation, and analysis
    of school achievement. The emphasis is upon
    content and objectives in areas of Reading,
    Language and Mathematics.

2. Test (Series)/Year: CALIFORNIA ACHIEVEMENT TESTS, 197,-78

    Forms: C and D

    Publisher/Distributor: CTB McGraw-Hill

    Authors: CTB/McGraw Hill Test Development Staff

    Description: Achievement test battery designed to measure
    knowledge of understanding in Reading,
    Mathematics, Language, Spelling, and
    Reference Skills. Levels 10-12 are
    available in Form C only.

3. Test (Series)/Year: COMPREHENSIVE TESTS OF BASIC SKILLS, 1968

    Forms: Q and R

    Publisher/Distributor: CTB McGraw-Hill

    Authors: Staff of CTB/McGraw-Hill

    Description: Achievement test battery designed to measure
    skills in Reading, Language, Arithmetic, and
    Study Skills.

4. Test (Series)/Year: COMPREHENSIVE TESTS OF BASIC SKILLS,
                          1973-75

   Forms: S (73) and T (75)

   Publisher/Distributor: CTB/McGraw-Hill

   Authors: Staff of CT8/McGraw-Hill

   Description: Achievement test battery designed to measure
                 skills prerequisite to studying and learning
                 in subject matter courses: Reading, Language,
                 Mathematics, Reference Skills; Science, and
                 Social Studies.

5. Test (Series)/Year: GATES-MACGINITIE READING TESTS, 1965-72

   Forms: 1, 2, 3

   Publisher/Distributor: Riverside Publishing Company,
                           Division of Houghton Mifflin

   Authors: A.I. Gates and W.H. MacGinitie

   Description: A series of tests designed to measure
                 group and individual reading achievement
                 from kindergarten through grade 12.

6. Test (Series)/Year: GATES-MACGINITIE READING TESTS, 1978

   Forms: 1, 2, 3

   Publisher/Distributor: Riverside Publishing Company,
                           Division of Houghton Mifflin

   Author: W.H. MacGinitie

   Description: A series of tests designed to measure reading
                 achievement of children in Grades 1 through 12.

363

7. Test (Series)/Year: IOWA TESTS OF BASIC SKILLS, 1971

   Forms: 5 and 6

   Publisher/Distributor: Riverside Publishing Company,
                           Division of Houghton Mifflin

   Authors: A.N. Hieronymus and E.F. Lindquist

   Description: Provides for comprehensive measurement in
                 fundamental skills: Vocabulary, Reading,
                 Language, Work-Study Skills (maps, graphs
                 and tables, and references), and
                 Mathematics

8. Test (Series)/Year: IOWA TESTS OF BASIC SKILLS, 1978

   Forms: 7 and 8

   Publisher/Distributor: Riverside Publishing Company,
                           Division of Houghton Mifflin

   Authors: A.N. Hieronymus, E.F. Lindquist, and H.D. Hoover

   Description: A new test edition designed to provide
                 comprehensive assessment of student
                 achievement in important areas of basic
                 skills. Normed concurrently with Tests of
                 Achievement and Proficiency, Form T, 1978,
                 and the Cognitive Abilities Test, Form 3,
                 1978. The expanded standard score of ITBS,
                 1978 is continuous with that of TAP, Form T.

9. Test (Series)/Year: METROPOLITAN ACHIEVEMENT TESTS, 1970

   Forms: F, G, H

   Publisher/Distributor: Psychological Corporation

   Authors: W.N. Durost, H.H. Bixler, J.W. Wrightstone,
             G.A. Prescott, and I.H. Balow

   Description: Designed to assess achievement in the
                 important skill and content areas of the
                 school curriculum in kindergarten through
                 junior high.

364

10. Test (Series)/Year: METROPOLITAN ACHIEVEMENT TESTS, 1978, INSTRUCTIONAL BATTERY

Forms: JI and KI

Publisher/Distributor: Psychological Corporation

Authors: I.H. Balow, R. Farr, T.P. Hogan and G.A. Prescott

Description: A nationally normed, criterion-referenced test battery in Reading, Mathematics, and Language. Each subject area includes major learning strands, each of which is represented by a test. Empirical norm-referenced scores are available for each subtest and for Total Mathematics and Total Language.

11. Test (Series)/Year: METROPOLITAN ACHIEVEMENT TESTS, 1978, SURVEY BATTERY

Forms: JS and KS

Publisher/Distributor: Psychological Corporation

Authors: I.H. Balow, R. Farr, T.P. Hogan, and G.A. Prescott

Description: Norm-referenced survey tests in Reading Comprehension, Mathematics, Language, Social Studies, and Science.

12. Test (Series)/Year: METROPOLITAN READINESS TESTS, 1974-76

Forms: P and Q

Publisher/Distributor: Psychological Corporation

Authors: J.R. Nurss and M.E. McGauvran

Description: The MRT is designed to assess readiness to being formal learning by measuring pre-reading and pre-mathematics skills. It is a readiness test battery only and has no provision for testing beyond the beginning of grade 1.

13. Test (Series)/Year: PEABODY INDIVIDUAL ACHIEVEMENT TEST, 1970

    Forms: One only

    Publisher/Distributor: American Guidance Service, Inc.

    Authors: L.M. Dunn and F.C. Markwardt, Jr.

    Description: A wide-range, individually administered
                 screening test of achievement.

14. Test (Series)/Year: STANFORD ACHIEVEMENT TEST, 1973

    Forms: A and B

    Publisher/Distributor: Psychological Corporation

    Authors: R. Madden, E.F. Gardner, H.C. Rudman,
              B. Karlsen, and J.C. Merwin

    Description: Academic achievement test battery whose
                 content areas include: Reading, Language
                 Arts, Mathematics, Science, and Social
                 Science. Expanded standard (scaled)
                 scores are continuous with Stanford Early
                 School Achievement Test (SESAT) and
                 Stanford Test of Academic Skills (TASK).

15. Test (Series)/Year: STANFORD EARLY SCHOOL ACHIEVEMENT TEST,
                         1967-70

    Forms: One only

    Publisher/Distributor: Psychological Corporation

    Authors: R. Madden and E.F. Gardner

    Description: A group-administered test designed to
                 measure children's cognitive abilities upon
                 entrance to kindergarten and during kinder-
                 garten and first grade. Expanded standard
                 (scaled) score is continuous with Stanford
                 Achievement Test, 1973.

366

16. Test (Series)/Year: TESTS OF ACHIVEMENT AND PROFICIENCY, 1978

    Forms: T

    Publisher/Distributor: Riverside Publishing Company,
                             Division of Houghton Mifflin

    Authors: D.P. Scanell, O.M. Haugh, A.H. Schild, and G. Ulmer

    Description: Designed to provide comprehensive appraisal
                   of student achievement for widely accepted
                   secondary-school goals in basic skills and
                   curricular areas. Normed concurrently with
                   Iowa Test of Basic Skills, Forms S and T, 1978,
                   to provide extended measurement in grades
                   9-12. TAP was also normed concurrently with
                   the Cognitive Abilities Test, Form 3.

17. Test (Series)/Year: TESTS OF BASIC EXPERIENCES, 1970-75,
                          FIRST EDITION

    Forms: One only

    Publisher/Distributor: CTB/McGraw-Hill

    Authors: M.H. Moss

    Description: TOBE measures children's acquisition of
                   the concepts and experiences considered
                   necessary for participation in the early
                   years of school. TOBE has two overlapping
                   levels (K and L) which span preschool
                   through grade 1. Each level has five tests:
                   Language, Mathematics, Science, Social
                   Studies, and General Concepts. Each test
                   item consists of a verbal stimulus and
                   four picture responses. As the examiner
                   reads the stimulus aloud, the child makes
                   a mark over the picture (or inside a bubble)
                   that he/she believes is the correct response.

18. Test (Series)/Year: TESTS OF BASIC EXPERIENCES, 1978,
                          SECOND EDITION

    Forms: One only

    Publisher/Distributor: CTB/McGraw-Hill

    Authors: M.H. Moss

Description: TOBE 2 measures children's acquisition of
the concepts and experiences considered
necessary for participation in the early
years of school. TOBE 2, which spans
preschool through grade 1, has two over-
lapping levels--K and L. Each level has
four tests--Language, Mathematics, Social
Studies, and Science. Each test item
consists of a verbal stimulus and four
picture responses. As the examiner
reads the stimulus aloud, the child fills
in an answer space indicating the picture
that he/she believes is the correct
response.

19. Test (Series)/Year: WIDE RANGE ACHIEVEMENT TEST, 1978

Forms: One

Publisher/Distributor: Jastak Associates, Inc.

Authors: J.F. Jastak, S.W. Bijou, and S.R. Jastak

Description: A two-level wide range test comprised of
three subtests: Reading, (recognizing and
naming letters and pronouncing words out of
context); Spelling, (copying marks
resembling letters, writing the name, and
writing single words to dictation); and
Arithmetic (counting, reading number
symbols, solving oral problems, and
performing written computations). The
test, basically a clinical type test,
consists of one four page test booklet
which includes both levels.

369

List of Test Publishers

Publisher	Tests
Addison-Wesley Testing Service 2725 Sand Hill Road Menlo Park, CA 94025 Telephone: 415-854-0300	CIRCUS Cooperative Primary Tests Sequential Tests of Educational Progress (STEP)
American Guidance Service, Inc. Publisher's Building Circle Pines, MN 55014 Telephone: 612-786-4343	Key Math Diagnostic Arithmetic Test Peabody Individual Achievement Test Woodcock Reading Mastery Tests
CTB/McGraw-Hill Del Monte Research Park Monterey, CA 93940 Telephone: 408-649-8400	California Achievement Tests Comprehensive Tests of Basic Skills Diagnostic Mathematics Inventory Prescriptive Reading Inventory Tests of Basic Experiences
Education Progress, Division of Educational Development Corporation 4235 South Memorial Tulsa, OK 74145 Telephone: 918-622-4522	Individualized Criterion Referenced Testing
Jastak Associates, Inc. 1526 Gilpin Avenue Wilmington, DE 19806 Telephone: 302-652-9990	Wide Range Achievement Test
Psychological Corporation 757 Third Avenue New York City, NY 10017 Telephone: 212-888-3500	Durrell Listening-Reading Series Iowa Silent Reading Tests Metropolitan Achievement Tests Metropolitan Readiness Test Stanford Achievement Test Stanford Diagnostic Mathematics Test Stanford Diagnostic Reading Test Stanford Early School Achievement Test Stanford Test of Academic Skills (TASK)
Riverside Publishing Company 1919 S. Highland Avenue Lombard, IL 60148 Telephone: 312-629-9700	Gates-MacGinitie Reading Tests Iowa Tests of Basic Skills Nelson Reading Skills Tests Tests of Academic Progress Tests of Achievement and Proficiency

369

Science Research Associates,
Inc.
155 North Wacker Drive
Chicago, IL 60606
Telephone: 312-984-2195

Iowa Tests of Educational
  Development
National Educational Development Tests
SRA Achievement Series

Scholastic Testing Service
480 Meyer Road
Bensenville, IL 60106
Telephone: 312-766-7150

STS Educational Development Series

Scott, Foresman and Company
1900 E. Lake Avenue
Glenview, IL 60025
Telephone: 312-729-3000

Comprehensive Assessment Program
  Achievement Series

Teaching Resources Corporation
50 Pond Park Road
Hingham, MA 02043
Telephone: 617-749-9461

Woodcock-Johnson Psycho-Educational
  Battery

371

Annotated List of Language Proficiency Tests

This annotated list of eight language proficiency tests is intended to provide project directors and evaluators with the information necessary to make a well-informed choice in selecting a language proficiency test. The criterion used for including tests in the list is the following: each test is recommended (at the time of printing) by at least one of the three states having the largest number of bilingual education programs.

The tests are primarily in Spanish and English and range from kindergarten level to high school. A brief description is offered of each test as well as comments on the linguistic and technical properties of the tests. The comments are points that evaluators and project directors should be well aware of in selecting a test or in interpreting test results. The comments were drawn from several sources including the experience of districts in the bilingual PIP field test study, and published articles and critiques. Each publisher was given an opportunity to respond to the review and to include "Publisher's Comments." This information has been incorporated into the reviews.

371

Basic Inventory of Natural Language (BINL)

Languages: English and Spanish (can be used for other languages)

What It Tests: Speaking

Levels and Grades: K-12

Administration: Individually administered. Requires 10-15 minutes.

Pictures are used to elicit natural speech and ten sentences are tape recorded for later analysis.

Scoring: Hand or machine scored.

Interpretation: Yields raw scores that can be coverted to one of four levels: NES, LES, FES, PES ("proficient").

Age is taken into account in determining levels.

Comments: Pictures are large, attractive, with multicultural content. It is difficult to standardize administration procedures since there is no set of "items" but rather an elicitation technique. Complex to score by hand. Scored on the basis of linguistic complexity and length of sentences. These criteria may not always be valid indicators of proficiency.

No information is provided on the validity of the proficiency categories. Information on validity is limited to correlations of sentence length with complexity, and correlations of complexity scores with an oral reading test. Reliability data is limited to correlations between the first half and the second half of the test. These correlations were high. Some districts have found that the test classifies fluent speakers as "limited" (see Gilmore and Dickerson, 1979).

Publisher's Comment: Standardization is facilitated by adequate training and close adherence to BINL procedures. Machine scoring procedures: reports of five different types, from classroom listings to district summaries, including pre-post averages, minimum, maximum and average scores by grade levels. A recent study establishes averages for grades K-12 based on a sample of 125,000 students. Standard error allows for valid adjustment of scores. The format of the test permits retest on invalid tests which have been reported to be less than 4% of tests submitted for machine scoring. Percentile rank of scores is now included in reports.

372

Bilingual Syntax Measure (BSM)

Languages: English and Spanish

What It Tests: Speaking

Levels and Grades: Level I, K-2 (ages 4 to 9); Level II (not available for review)

Administration: Individually administered.
Requires 10-15 minutes
Students respond orally to questions based on pictures.

Scoring: Hand scored

Interpretation: Provides language dominance (when both English and Spanish tests are administered), level of second language acquisition, and degree of maintenance or loss of the first language. Assigns students to one of five proficiency levels in each language. Additionally, provides instructional suggestions for reading and ESL which correspond to each of the five English proficiency.

Comments: Attractive, colorful pictures are used to elicit speech through structured conversation. Responses are scored strictly on he correctness of specific grammatical structures. The choice of grammatical structures is based on research studies on the sequence of acquisition of morphemes. Allows for regional language variation. A number of discussions of this test have been published including Hernandez-Ch., 1978(1) and Rosansky, 1979(2).

Both test-retest reliability and inter-scorer reliability are reported in the Technical Handbook. Although the reported reliability is low, the authors attempt to explain why this is so (TH, p. 45).

---

(1) Hernandez-Chavez, Eduardo. Critique of a critique: Issues in Language Assessment. Journal of the National Association for Bilingual Education, March 1978, Vol. II, No. 2.

(2) Rosansky, E.J. A Review of the Bilingual Syntax Measure. In: Advances in Language Testing, edited by B. Spolsky. Arlington, VA: Center for Applied Linguistics, 1979.

Comprehensive English·Language Test for Speakers of English as a
Second Language (CELT)

Language: English

What It Tests: Listening comprehension, grammar, and vocabulary.
Contains three subtests: (1) Listening, (2) Structure, and (3)
Vocabulary..

Levels and Grades: High school, college, and adult.

Designed for intermediate to advanced ESL students.

Administration: Group administered.

Listening requires 40 minutes; Structure requires 45 minutes;
Vocabulary requires 35 minutes. A recording can be used to administer
the listening test.

All test items are multiple choice. Students respond to oral and
written stimuli by marking an answer sheet.

Scoring: . Scored with a key.

Interpretation: Yields percent correct for each test.

Percentile scores are available (but see Comments).

Does not provide proficiency classifications. No cutoff score is
provided for classification of students as limited in English
proficiency, since test was not designed for this purpose.

Comments: Oral production is not tested.

All test items on each subtest are multiple choice items that require
reading; therefore, the measures of listening comprehension,
structure, and vocabulary are each confounded with literacy skills.
The authors recommend the Vocabulary subtest for use with students who
have had advanced training in reading.

The three subtests had moderate to high internal consistencies with
four groups of foreign students and, therefore, very reasonable
standard errors of measurement. No information is given on predictive
validity. Tentative evidence of concurrent validity is offered based
on correlations with other standard ESL tests. Tentative norms for
five different groups, based on small samples, are provided. The
norms are not appropriate for use in most bilingual programs, however,
since the students in the norming sample are not similar to most
students in bilingual programs.

3 7

Ilyin Oral Interview Test

Languages: English

What It Tests: Speaking

Levels and Grades: Secondary and adult.

Forms: There are two forms (BILL and TOM) and each has a long version (50 items) and a short version (30 items).

Administration: Individually administered. Requires up to 30 minutes.

The students respond to pictorial stimuli and questions by responding orally. Items are ordered in difficulty and interview is terminated when a frustration level is reached.

Scoring: Hand scored.

Interpretation: Yield raw scores. No cutoff score is given to identify students as "limited" in English proficiency; however, suggestions are given for placement levels in adult ESL programs, and a range is suggested as the degree of proficiency required for jobs in which oral communication with the public is limited.

Comments: The requirement to answer in a complete sentence is an unnatural one and may depress scores of students who fail to do this. The long version can become monotonous since many pictures are repeated.

Internal consistency reliabilities are high. No information is given for test-retest reliability or interrater reliability. Validity information is limited to correlations with other tests, and based on very small samples.

375

Language Assessment Battery (LAB)

Languages: English and Spanish

What It Tests: Listening, speaking, reading, and writing.

Level I has three subtests: (1) Listening and Speaking, (2) Reading, and (3) Writing. Levels II and III have four subtests: (1) Listening, (2) Reading, (3) Writing, and (4) Speaking.

Levels and Grades: Level I, grades K-2; Level II, grades 3-6; Level III, grades 7-12.

Administration: Level I: Individually administered, requires 5-10 minutes.

Levels II and III: Part is individually administered; requires 41 minutes.

Students respond to verbal, written, and pictorial stimuli by pointing, by giving oral responses, by writing, and by marking answer sheets (on Levels II and III only).

Scoring: Hand scored; parts scored with a key.

Interpretation: Yields raw scores and stanines and percentiles by grade. Students scoring below the 20th percentile may be classified as limited in English proficiency.

Comments: The speaking section of Level I, Test 1, contains only 6 items, all of which may be answered with one word. The writing tests measure reading skills in addition to writing skills.

The test went through all the stages of preparation by expert and experienced item writers, pilot studies, item- and test-analyses, and norming on substantial samples (20 schools, and about 500 students at each level from K through 12). The technical manual is a model.

One study(1) has shown that the Level I English test does not discriminate well in the range near the cutoff point for classifying students as limited in English. This reduces its value for use as a pre-post measure.

_____

(1) Hubert, J. "An Investigation of the Language Assessment Battery (English, Level 1) for Title VII students in Hartford." Unpublished manuscript, 1978.

Language Assessment Scales (LAS)

Languages: English and Spanish

What It Tests: Listening comprehension and speaking. Five subtests form the total score for both levels: (1) discrimination of minimal phonemic pairs, (2) vocabulary production, (3) phoneme production, (4) syntax comprehension, and 85) story production.

Levels and Grades: Level I, grades K-5. Level II, grades 6-12.

Administration: Individually administered.

Requires.20 minutes.

Stimuli consists of tape recorded speech and pictures. Students respond orally, and by pointing.

Scoring: Hand scored.

Interrater reliability should be obtained on storytelling task.

Age is taken into account in scoring.

Interpretation: Yields a score of 1 to 100 which can be converted to a level, 1 to 5.

Students who score at level 3 or below are classified as "Limited English (or Spanish) speakers."

Comments: This is a fairly comprehensive overall aural-oral proficiency test. There are problems with the phonemic discrimination section since this task requires a kind of metalinguistic awareness students may not have. The story retelling task measures not only production, but also comprehension.

Interrater reliability coefficients for the story retelling task are moderately high. Coefficients of internal item consistency for discrete-point items range from .36 to .96.

Validation consisted of one-way analyses of variance of relatively small samples (one- to two hundred) of students dichotomized into English-dominant and Spanish-dominant on the basis of teacher judgment.

Several studies of reliability were done on small samples (21 English and 35 Spanish) using various approaches. The sample sizes were too small to justify some of the analyses and the conclusions drawn from them.

3͞7

Primary Acquisition of Language (PAL) Oral Language Dominance Measure (CLDM) Oral Language Proficiency Measure (OLPM)

Languages: English and Spanish

What It Tests: Listening comprehension and speaking

Levels and Grades: PAL CLDM. K-3; OLPM. 4-6

Administration: Individually administered.

Requires 15 minutes for each language.

Students respond orally to oral and pictorial stimuli.

Scoring: Hand scored.

Interpretation: Yields raw scores ("G scores") that are converted to proficiency levels. 1 to 5. Also yields dominance categories.

Students who score at level 4 or below are classified as "Limited English (or Spanish) speakers."

Comments: Simple to use and score. Scored on the basis of grammaticality and appropriateness of responses a well as quantity of speech.

The test was developed "as a result of research by the El Paso Public Schools."

Item analyses were used in the construction of the tests although samples were somewhat small (about 20C drawn from three grades in high schools). Validity is quoted in terms of the test's ability to grade schools in correct order, and of correlations with a reading test. The latter were fair being around 0.3 to 0.5.

3~8

Shurtt Primary Language Indicator Test (SPLIT)

Languages: English and Spanish

What It Tests: Listening comprehension, speaking, reading, and grammar.

There are three subtests: (1) Listening Comprehension, (2) Verbal Fluency, and (3) Reading Comprehension and Grammar.

Levels and Grades: Listening Comprehension, Verbal Fluency, K-6; Reading Comprehension and Grammar, 3-6.

Administration: Listening Comprehension: Group administered; requires 35 minutes, tape recording available.

Verbal Fluency: Individually administered; requires 15 minutes.

Reading Comprehension and Grammar: Group administered; requires 30 minutes.

Instructions are provided in both languages and are available on tape. Stimuli are oral, pictorial, or written. Students respond orally, by marking pictures in answer book, or by marking an answer sheet.

Scoring: Hand scored; parts scored with a key.

Interpretation: Yields raw scores, percentile ranks, and age and grade equivalents.

Yields a dominance classification.

Comments: Yields no cutoff point to classify students as limited in English proficiency (independent of Spanish/Portuguese score). A proficiency classification is given based on the dominance classification. This wrongly assumes that students are highly proficient in the dominant language. A student whose English score is very low can be classified as "English Adequate" if the student's Spanish score is also very low, but higher than the English score. Districts should establish their own cutoff points for classifying students in English.

Grade equivalent scores should not be used.

MEASURING SELF-CONCEPT

Research investigating the relationship between ethnicity and student self-concept is mixed and inconclusive. The mixed findings may be attributed to many factors, including the fact that different researchers have measured different dimensions of self-concept and compared them as if they were the same. The end result is that the relationship between ethnicity and self-concept is still vague and needs more careful study.

The concept of self is basically derived from (1) the responses made toward the individual by significant people in his immediate environment, (2) his perception of their behavior towards him, (3) the interrealization of his perception into a coherent set of self-views, (4) the resultant self which he perceives as reflected back into the eyes of others, (5) the reinforcement of that self as seen by him and by others, as well as by his view of their concepts of him and (6) his responses to the challenges and pressures of living.

The need to address cultural differences as a factor when measuring the self-concept of minority students was addressed by Whiting (1974) when he developed a series of self-concept measures. In stressing the need for culturally sensitive instruments, Whiting pointed out that tests should be designed with a particular population in mind, taking into account that population's values and concerns in order to

measure self-concept more accurately. Whiting goes on to describe a
battery of self-concept instruments developed around the
multi-dimensionality of self-concept. He described self-concept in
the following manner:

o   Self-esteem refers to how an individual evaluates
        himself and indicates the extent to which he
        believes himself to be capable, significant,
        successful, and worthy. Within the context of
        this definition, the individual arrives at an
        evaluation of his own worthiness by examining his
        performance, capacities, and attributes in light
        of his own personal standards and values. Thus,
        self-esteem is a "personal judgement of worthiness
        that is expressed in the attitudes the individual
        holds toward himself."

o   Sense of control refers to how much an individual
        accepts responsibility for his own actions, or
        whether he attributes power and control to various
        external agents, such as adults, peer, luck,
        brothers or sisters - the "system," or fate. If a
        child has little sense of control, he also has
        little sense of responsibility, since the two are
        so closely related.

o   Academic self-concept refers to how an individual
        evaluated his ability to function succesfully in a
        school environment.

o   Social self-concept refers to how an individual
        thinks the people who are significant in his life
        perceive him.

Special Problems of Measuring Self-Concept of Culturally Different
Children

The measuring of self-concept and attitudes is complicated by the need

to consider the distinct cultural background of many of the students

participating in the program. Thus, a published instrument with

acceptable levels of validity and reliability may not be appropriate

3 9 1

because it was normed on middle class Anglo-American students. It is common among some American Indian tribes, for example, to expect children to be quiet and passive when adults are conversing. In a school setting, however, a child who is non-talkative and passive may be considered shy and withdrawn.

Measuring self-concept is complicated by questions such as the appropriateness of norm groups, as well as the difficulty of accurately assessing an area whose definition and sophistication are questionable. The qualifications of persons administering the self-concept measurement instruments, the attitudes of teachers and administrators toward the use of school time to test, and the views of both teachers and administrators on the importance of self-concept as a program goal also present measurement problems. Still another problem is the change in self-concept as children grow older. Studies have shown without exception that the self-concept actually deteriorates as children grow older. One is reminded of Alice who is told she must run twice as fast to remain in the same place. By selecting positive growth in self-concept as a goal in bilingual education programs, teachers must work twice as hard to show positive changes in a phenomenon that normally shows negative changes.

Measuring Self-Concept

Because of the issues above, the best means for measuring desired change in self-concept would be 1) pre- and posttesting with published tests normed on minority groups similar to the students in the

bilingual program and (2) to use informal, teacher-developed formative measures using a variety of approaches, including paper and pencil and picture instruments, teacher observation guides, and parent questionnaires.

The reason for monitoring changes in self-concept centers around the relationship of positive changes in self-concept and success in school, as well as success in social relationships. Since it is important to obtain an accurate assessment of what is happening to a student's self-concept while he is a participant in the program, it is necessary to take readings on self-concept fairly often and in a variety of ways. An adequate system for measuring self-concept would include teacher-made tests, as well as published instruments. Picture tests for non-readers, as well as paper and pencil tests for older children, should be a part of the testing program. Projective techniques where students determine concurrence between self-ratings and the ratings of others are powerful means of helping students become aware of themselves in terms of how "others see me." However, certain cautions should be taken when utilizing self-reporting instruments. Results should be kept confidential, instruments should be administered in a non-threatening manner, testers should point out that there are no right or wrong answers and testers should read items to very young students.

Teacher-Developed Self-Concept Instruments

Teacher-developed instruments offer several advantages in that (1) they may include items unique to the community, program, school or

classroom, (2) the instrument may include items that take into account any cultural variables and (3) teachers will be more positive toward instruments that are locally developed.

Among the behavioral indicators a teacher might look for would be:

o   How does the student react to a new situation?

o   How does the student react to new material?

o   Does he trust his teacher (especially when new in class)?

o   Is he cooperative and does he follow directions reasonably well?

o   Does he control his own behavior?

o   Does he have his own ideas?

o   Does he talk freely about his ideas?

o   Does he operate on his own with a minumum of direction from the teacher?

o   Is he generally a happy person?

The preceding are only suggestions, but if adapted, they will serve as the basis for developing an observation guide to assess self-concept.

Published Self-Concept Instruments

Published instruments, with acceptable levels of reliability and validity, are an essential element in evaluating the self-concept of students in the program. However, because the students are sometimes not as mature as test publishers would like them to be, self-reporting instruments may lack desirable levels of reliability and validity.

This is especially true when the instruments are used with very young children. Nevertheless, they do provide insights into student behavior, especially when used in conjunction with other measuring instruments.

Samples of published instruments should be examined, if possible prior to their purchase. If not, there are several reference works available which provide descriptions of self-concept instruments. Among them are R.C. Wylie's The Self Concept: A Review of Methodological Considerations and Measuring Instruments. Revised edition. Lincoln University of Nebraska Press 1974; or R. Shavelson's "Self Concept: Validation of Construct Considerations" Review of Educational Research 46 (1976) pp. 407-441.

Due to the built-in difficulties of showing positive growth, a word of caution concerning pre- and posttest timing of self-concept instruments is needed. In order to avoid pretesting students that are already in the program and considering the fact that self-concepts are relatively stable, testing should be done as early as possible. Subsequent posttesting should be spaced as far from the pretesting as is feasible.

The message for the evaluator is clear. Unless self-concept measures are carefully developed to reflect the unique characteristics of the students in the program and carefully selected from reputable publishers, there is a distinct possibility that, in spite of a

positive effort to show positive changes in self-concept as a result

of the project, negative changes may actually result.

A list of published tests to measure self-concepts is attached.

3ˁ6

Published Self-Concept Scales

1. ## Self Concept Picture Inventory (Wiseman & Adams)

   The Self Concept Picture Inventory was designed to evaluate grades one to three of Title I programs in Alton, Illinois. The test is appropriate for younger students and is relatively free of racial and sex biases.

   Wiseman, E.D. & Adams, J.H. "Self Concept Picture Inventory". Atton, Ill. 1972. (see ERIC ED 170-299)

2. ## The Florida Key: A Scale to Infer Learner Self Concept (Purkey)

   The Florida Key is a learner self-concept scale that, with adaptations, may be used with students of all ages. The scale is designed to aid teachers in evaluating students' self-concepts as learners, as well as attitudes toward school.

   Purkey, W.W. et. al., "The Florida Key: A Scale to Infer Learner Self Concept." Educational and Psychological Measurement. 33 (1973) pp. 979-984.

3. ## Thomas Self-Concept Values Test (Thomas)

   The Thomas Self-Concept Values Test measures fourteen self-value dimensions, such as sociability, ability, attractiveness, and independence. The 14-item test is designed to be used with young children from 4 to 6 years old. However, some caution should be exercised in interpreting test results given the problems of self-concept measurement in young children.

   Thomas, W.L., The Thomas Self Concept Values Test. Combined Motivation Education Systems, Inc., 1959. Rosemont, Illinois.

4. ## Self-Esteem Inventory (Coopersmith)

   This 58-item scale was designed to measure general self, social self-peers, home/parents, and school academic self in addition to self-esteem. It is worded to be used with children from 8 to 10 years old, but has been used successfully with students in grades three through twelve.

   Coopersmith, S., The Antescendents of Self-Esteem. San Francisco, California, W.H. Freeman & Co., 1967

5. Self-Concept of Ability Scale (Brookover)

The eight items contained in Form A are designed to
measure self-concept of general academic ability; and
the eight items in Form B are designed to measure
self-perceptions of ability regarding science,
mathematics, social studies, and English. The scale is
most suitable for use with high school aged students.

Brookover, W.B. et. al., Relationship of Self-Concept
to Achievement in High School, 1967. Michigan State
University, Lansing, Michigan.

6. Piers-Harris Children's Self-Concept Scale (Piers and
Harris)

The 80-item instrument measures general self-concept
and may be used for both research and diagnostic work.
The simple descriptive statements are designed to
measure ten self-concept dimensions and the scale is
appropriate for use in grades three or above.

Piers, E.V. & Harris, D.B. Piers-Harris Children's Self
Concept Scale, Nashville, Tenn. Counselor Recording
And Tests, 1969.

7. How I See Myself Scale (Gordon)

The How I See Myslf Scale consists of 40 (elementary
form) or 42 (secondary form) items developed for use
with children ages 3 to 12 years. The scale has been
found to measure five self-concept dimensions: physical
appearance, interpersonal, teacher-student, academic
ability, and autonomy.

Gordon, I.J. A Test Manual for the How I See Myself
Scale. Florida Research and Development Council,
Tampa, Florida.

8. About Me by James Parker; Not Dated; Grades 4-6; James
Parker*.

A five-point self-rating scale assessing five areas of
self-concept which are expressed in behavior in the
school setting. Subscores included are: Self, Self in
Relation to Others, Self as Achieving, Self in School,
and the Physical Self.

*Included in Parker, James. The Relationship of
Self-Report to Inferred Self-Concept, Educational and
Psychological Measurement, 26 pp. 691-700; 1966.

9. The Behavior Cards: A Test-Interview for Delinquent Children by Ralph M. Stogdill; c1941-50; Grades 3-10; Stoelting Company.

    Use of the Cards provides the child with an opportunity to face his problems and provides an insight into the child's attitudes toward his delinquent behavor. The test is individually, administered, employing the card-sort technique. Any child who scores grade 4.5 or higher on a standardized reading test should be able to sort the cards with little assistance. Cards can be read to subjects with reading disabilities. At times an abbreviated version of the test can be given by eliminating fifty specified cards.   This eliminates the more serious delinquent behaviors.

10. Behavior Rating Form by Stanley Coopersmith; Not Dated; Grades Kindergarten-9; Stanley Coopersmith*.

     A 13-item five-point rating scale devised for appraising assured and confident behavior. Items refer to such behavior as the child's reaction to failure, self-confidence in a new situation, sociability with peers, and the need for encouragement and reassurance. The form yields two scores: Esteem Behavior and Confidence Behavior.

     *Data is available in: Coopersmith, Stanley; Antecedents of Self-Esteem; San Francisco; W.H. Freeman, 1967.

11. Children's Self-Conception Test: Form II by Marjorie B. Creelman; c1954-55, Grades 3-6; Marjorie B. Creelman.

     Designed to assess the relationship of self-concept to adjustment or maladjustment. Employs a series of pictures depicting situations commonly experienced by children in Western culture. Test provides indications of self-esteem and moral standards.

12. Children's Self-Social Constructs Test: Primary Form by Edmund H. Hendersen, Barbara H. Long, and Robert C. Ziller; 1967; Grades 1-6; Edmund H. Hendersen.*

     A measure of social self-concept from which certain aspects of the child's conceptions of himself are inferred. Subscores include: Self Esteem, Social Interest or Dependency, Identification, Group Identification, Individuation or Minority Identification, Power, Egocentricity, Complexity, Realism for Size, and Preference for Others.

*Listed as available from Edmund H. Henderson, may now be obtained from The Office of Special Tests, Educational Testing Service, 17 Executive Park Drive, NW, Suite 100, Atlanta, Georgia 30329.

13. **Columbus Sentence Completion for Children** by Jack A. Shaffer and Arthur S. Tamkin; Not Dated; Ages 4-Adolescence; Jack A. Shaffer.

   A general projective test covering the following topics: Self-Concept, Wishes and Plans, Self-Concept (Problems), Family, Social, School, and Picture of Self. The test provides an indication of the child's adjustment level.

14. **Coopersmith Self-Esteem Inventory: Form A** by Stanley Coopersmith; Not Dated; Ages 9-Adults*; Stanley Coopersmith.

   Designed to provide a general assessment of self-esteem. The 58 items are arranged into five subscales: General Self, Social Self-Peers, Home-Parents, Lie Scale, and Home Academic.

   *Can be used with children younger than age 9 if individually administered. Technical information is available in: Coopersmith, Stanley. Antecedents of Self-Esteem; San Francisco; W.H. Freeman, 1967.

15. **Coopersmith Self-Esteem Inventory: Form B** (Short Form) by Stanley Coopersmith; Not Dated; Ages 9-Adults; Stanley Coopersmith*.

   Designed to measure self-esteem from the perspective of the subject, Emphasis is placed on the subject's self-attitudes in four areas: peers, parents, school and personal interest.

   *Additional information is available in: Coopersmith, Stanley, Antecedents of Self-Esteem; San Francisco; W.H. Freeman, 1967.

16. **Expanded Test Anxiety Scale for Children** (Feld and Lewis 1969) by Sheila C. Feld and Judith Lewis; 1969; Grades 1-9; Sheila C. Feld*.

   A modification of the Sarason Test Anxiety Scale for Children which includes the original and revised questions and two neutral items about dreams and achievement. Subscales include: Test Anxiety, Remote School Concern, Poor Self-Evaluation, and Somatic Signs of Anxiety.

391

*Included in Feld, S. and Lewis, J. "The Assessment of Achievement Anxieties in Children." In C.P. Smith (Ed.). Achievement Related Motives in Children. New York; Russell Sage Foundation, 1969, pp. 151-199.

17. How I See Myself Scale: Elementary Form by Ira J. Gordon; 1968; Grades 3-6; Ira J. Gordon (Manual is available from the Florida Educational Research and Development Council).

Factors assessed are Teacher-School, Physical Appearance, Interpersonal Adequacy, and Academic Adequacy.

18. How Much Like Me? by Dale W. Dysinger; Not Dated; Grades 3-5; Dale W. Dysinger.

A self-administered measure of general self-concept.

19. Inferred Self-Concept Judgment Scale by Elizabeth McDaniel; 1965-69; Grades 1-9; Elizabeth McDaniel.

Designed to measure the student's self-concept as it is generated and in the school setting.

20. Inferred Self-Concept Scale: Experimental Form by Elizabeth L. McDaniel; 1969; Grades 1 and above. Western Psychological Services, 12031 Wilshire Blvd., Los Angeles, California.

Scale is based upon the assumption that self-concept can be inferred from manifest behavior. Scale purports to be appropriate for assessing and comparing self-concepts of culturally different groups. Test may also be used with adults and juveniles.

21. Instructional Objectives Exchange: Measures of Self-Concept, Kindergarten-Grade 13, Revised Edition; 1972; Grades Kindergarten-12; Instructional Objectives Exchange.

A series of affective objectives concerning the learner's self concept. Dimensions employed are peer, scholastic, family, and general, self-report inventories (direct and indirect) and observational inventories are provided to assess the attainment of each objective.

22. Instructional Objectives Exchange: Objective Collection in Attitude Toward School, Kindergarten-Grade 12, Revised Edition: 1972; Grades Kindergarten-12; Instructional Objectives Exchange.

A collection of affective objectives dealing with the
learner's self-concept as reflected in attitudes toward
teacher, school subjects, learning, peers, social
structure and climate, and general attitudes. An
observational indicator and both direct and inferential
self-report measures are provided to assess the
attainment of each objective.

23. Morgan Punishment-Situation Index by Patricia K. Moran;
     Not Dated (Test is copyrighted); Ages Children 9-12 and
     their mochers; Eugene L. Gaier.

     A projective device specifically concerned with the
     perception of the direction of aggression in the
     punishment situation. The Index yields four concepts
     operating in the punishment situation: the child's
     self-concept, his concept of his mother, the mother's
     self-concept, and her concept of the child. Employs
     scoring procedures developed for Rosenzweig
     Picture-Frustration Test.

24. Rogers' Personal Adjustment Inventory by Carl R.
     Rogers; 1961; Ages 9-13; Western Psychological
     Services.

     Designed to assess a child's attitude toward himself,
     his family, and his peers. Subscores include:
     Personal Inferiority, Social Maladjustment, Family
     Maladjustment, and Daydreaming.

25. Sears Self-Concept Inventory: Abbreviated. Form by
     Pauline S. Sears; 1966; Grades 3-6; Pauline S. Sears.

     The child rates himself in terms of: Physical Ability,
     Attractive Appearance, Convergent Mental Ability,
     Social Relations with Same Sex, Social Virtues,
     Divergent Mental Ability, Work Habits, Happy Qualities
     and School Subjects.

26. Self-Concept Adjective Checklist by Alan J. Politte; c
     1971; Grades Kindergarten-8; Psychologists and
     Educators, Inc.

     Enables the student to project his personal feelings
     related to self-concept phenomena and provides indices
     of his general levels of self-concept feelings. The
     adjectives cover the following: Physical Traits,
     Social Values, Intellectual Abilities, and
     Miscellaneous (emotional feelings, group behaviors, and
     habits): As a result of the scoring, the child is
     identified as "self-confident," poor self-concept, or
     "aggressive."

27. Self-Concept and Motivation Inventory: Later Elementary Form by George A. Farrah; c1968; Grades 3-6; Person-O-Metrics.

Measures academic self-concept in terms of the child's perception of his role as a learner. The inventory yields scores for role expectations, self-adequacy, goal and achievement needs, and failure avoidance.

28. Self-Concept As A Learner Scale-Elementary by John K. Fisher; Not Dated; Grades 3-6; John K. Fisher.

The SCALE is a modification of the secondary scale developed by Walter B. Waetjen. Subscores include: Motivation, Task Orientation, Problem Solving, and Class Membership. The Motivation factor is designed to determine the degree to which the respondent perceives himself motivated to do school work and to participate in learning activities. Task Orientation refers to the way a student sees himself relating to learning activities. Problem Solving determines the view that a pupil has of himself as a problem solver. The Class Membership factor is designed to find out how the student sees himself in relation to other members of the class.

29. Self-Concept Instrument-A Learner Scale by Gordon P. Liddle; 1967; Grades 3-6; Gordon P. Liddle.

Variables assessed are self-concept in reference to motivation, intellectual ability, task orientation, and class membership.

30. Self-Concept of Ability Scale; 1963-68; Grades 2-6; University of Maryland Research and Demonstration Center of the Interpersonal Research Commission on Pupil Personnel Services.

Designed to assess change in self-reported attitudes of groups of students toward themselves as learners. Covers six academic content areas: arithmetic, English, social studies, science, music, and art. The bases of comparison are the class, the grade level, close friends, future high school class, future college associates, other students in general, and one's own ability. The scale was adapted from Brookover, Paterson, Thomas' Self-Concept of Ability.

31. Self-Concept Target Game by Ann Fitz-Gibbon; 1970; Ages 9-10; Ann Fitz-Gibbon.

Designed for use with children who have participated in the Responsive Model Follow Through Program. It is a

measure of self-concept in terms of the child's willingness to take reasonable risks of failure, make positive estimates of his ability to perform a task, make realistic statements about the probability of being right or wrong, learn from errors and corrections, use failure in a productive manner, and take credit for accomplishments and acknowledge failure, individually administered.

32. Self Profile Q-Sort by Alan J. Politte; c1970; Grades 3-8; Psychologists and Educators, Inc.

Aids in elementary school counseling by providing a means for eliciting self-evaluation from a student, for investigating changes in a student's self-concept through the course of counseling sessions, and for stimulating group interaction in the counseling setting.

33. A Semantic Differential for Measurement of Global and Specific Self-Concepts by Lois Stillwell; Not Dated; Grades 1-3 and 4-6; Lois Stillwell.

Test can be modified to assess attitudes towards self in a variety of specific roles or conception of self from the point of view of a stated referent. The Primary Form is appropriate for Grades one through three and the Upper Grades Form is for the fourth grade and beyond. Test can be group administered easily to those in grade three or higher. First and second graders may have difficulty and will require several assistants to provide close observation. Subscores include: Myself, Myself As a Student, Myself As a Reader, Myself As an Arithmetic Student.

34. Tennessee Self-Concept Scale: Clinical and Research Form by William H. Fitts; c1964-70; Ages 12 and Above; Counselor Recordings and Tests.

Yields 30 profiled scores: Self Criticism, Self Esteem (Identify, Self-Satisfaction, Behavior, Physical Self, Moral-Ethical Self, Personal Self, Family Self, Social Self, Total), Variability of Response (Variation across First Three Self-Esteem Scores, Variation across Last Five Self-Esteem Scores, Total), Distribution, Time, Response Bias, Net Conflict, Total Conflict, Empirical (Defensive Positive, General Maladjustment, Psychosis, Personality Disorder, Neurosis, Personality Integration), Deviant Signs, and five scores consisting of counts of each type of response made.

35. Tennessee Self-Concept Scale: Counseling Form by William H. Fitts; c1965-70; Ages 12 and Above; Counselor Recordings and Tests.

Yields 15 profiled scores: Self-Criticism, Self-Esteem (Identity), Self-Satisfaction, Behavior, Physical Self, Moral-Ethical Self, Personal Self, Family Self, Social Self, Total, Variability of Responses (Variation across First Three Self-Esteem Scores, Variation across Last Five Self-Esteem Scores, Total), Distribution, and Time.

36. What I Am Like; Not Dated; Grades 4-10; Cincinnati Public Schools, Division of Psychological Services and Division of Program Development.

A five-point, bi-polar, self-rating scale based on Osgood's concept of the semantic differential. Subjects are: What I Look Like, What I Am Like When I Am With My Friends, and What I Am. The test is for research only and is to be used only ir group assessment.

37. When Do I Smile? by Dale W. Dysinger; Not Dated; Grades 1-5; American Institutes for Research.

Variable assessed in self-concept in reference to the school setting.

List of Publishers and Developers

American Institutes for Research
P.O. Box 1113
Palo Alto, CA 94302

Cincinnati Public Schools
Division of Psychological Service
608 East McMillan Street
Cincinatti, OH 45206

Stanley Coopersmith
University of California
Davis, CA 95616

Combined Motivation Education Systems
6300 River Road
Rosemont, IL 60018

Counselor Recordings and Tests
Vanderbilt University
Box 6184, Acklen Station
Nashville, TN 37212

Marjor e B. Creelman
Cleveland State University
College of Arts and Sciences
Cleveland, OH 44115

Dale W. Dysinger
American Institutes for Research
135 North Belleview Avenue
Pittsburg, PA 15213

Educational Testing Services, Inc.
Princeton, NJ 08541

Sheila C. Feld
University of Michigan
Frieze Building
Ann Arbor, MI 48104

John K. Fisher
Department of Psychology
Edinboro State College
Edinboro, PA 16512

Ann Fitz-Gibbon
Far West Laboratory for Education
  Research and Development
1855 Folsom Street
San Francisco, CA 95103

Florida Research and
  Development Council
901 East Kennedy Boulevard
Tampa, FL 33602

W.H. Freeman & Co.
1736 Stockton
San Francisco, CA 94133

Eugene L. Gaier
Faculty of Educational Studies
State University of New York
  at Buffalo
Buffalo, NY 14214

Ira J. Gordon
College of Education
Gainesville, FL 32601

Edmund H. Henderson
McGuffey Reading Center
School of Education
University of Virginia
Charlottesville, VA 22093

Instructional Objectives Exchange
Box 24095
Los Angeles, CA 90024

Gordon P. Liddle
Associate Dean
University of Kentucky
Lexington, KY 40506

Michigan State University
Lansing, MI 48824

James Parker
Department of Education
Georgia Southern College
Statesboro, GA 30458

Person-O-Metrics
203-04 Williamsburg Road
Dearborn Heights, MI 48127

Psychologists and Educators, Inc.
Suite 212
211 West State Street
Jacksonville, IL 62750

111-103

Pauline E. Sears
Research and Development Education
Stanford University
Stanford, CA 94305

Jack A. Shaffer
Department of Psychology
Humboldt State College
Arcata, CA 95521

Lois Stillwell
Office of Child Study and Guidance
Akron Public Schools
70 North Broadway
Akron, OH 44308

Stoelting Company
1350 South Kostner
Chicago, IL 60623

University of Maryland Research
  and Demonstration Center of
  the Interprofessional Research
Committee on Pupil Personnel Service
University of Maryland
College Park, MD 20910

Western Psychological Services
Box 775
Beverly Hills, CA 90213

ETHNOGRAPHIC METHODS OF PROGRAM DESCRIPTION

Introduction

This document contains a discussion relative to the collection of data
and eventual development of a program description. The ethnographic
procedures outlined fulfill the basic requirements of the program
description needed for a local school district·bilingual education
evaluation report. Use of ethnographic methods for program
description are not common in bilingual education evaluations.
However, such procedures have long been suggested for use in examining
the efficacy of bilingual education programs. These procedures are
relatively easy to implement and require a low expenditure of time and
energy. Furthermore, they necessitate minimal skill development on
the part of the evaluator and/or other data collectors.

While necessary components of program evaluations, psychometric and
quantitative descriptions of bilingual education programs are limited
in the breadth of useful information provided. The demographic
numbers, the enumeration of staff rolls, the description of physical
features, and the indications of time allocations for various
instructional components, etc. are all included as necessary parts of
most evaluation reports of school programs. This information,
however, seldom provides all the necessary useful insights needed to
comprehend the actual process of schooling that occurs in the
educational program evaluated. Very often, nonmeasureable aspects of

schooling such as classroom climate, teacher's attitude, instructional
interactions, etc. are ignored. Yet this subjective characterization
can provide an account of the features of a program that may well be
responsible for its success or failure in meeting educational goals.
This document proposes an addition to the traditional information
included in evaluation reports; namely: Ethnography of classrooms and
program.

School programs are complex. They change over time. They do not
always conform to a priori theorizing or standardization of procedure
or to goals that have been previously set. They cannot be
characterized as consisting of isolated and discrete occurrences of
phenomena each having meaning only in a strictly defined, contained,
or denotative sense. All aspects of a program are related to each
other and to the participants taking part in the program. Because the
programs are not homeostatic, with variables that can be isolated,
each with a singular independent effect, they do not lend themselves
to manipulation by evaluators. The process of schooling and indeed of
bilingual education has multiple realities and as such, events must be
understood from the perspective of the total program.

Quantitative evaluations and their program descriptions attempt to
discover, verify, or identify causal relationships among concepts
derived from a theoretical scheme that may or may not reflect the
reality of the program. Frequently, interrelationships among the
various aspects of the program may not be clear. As a result,

309

replication of the program may not be possible. At times, a program
description in purely quantitative evaluations represents an
"outsider's" cursory comprehension of a program's operation that lacks
verification by the participants as to its accuracy and explanatory
input from "insiders" as to its content. Qualitative descriptions,
such as are suggested in this document, can reflect both what the
"insiders" (the managers and the teachers) believe is occurring as
well as represent what is actually seen to occur by the evaluator
("outsider").

This document will be organized into four parts: Definition of
ethnographic methods, types of information to acquire, data collection
and instrumentation, and information usage and reporting.

Ethnography

Ethnographies attempt to accurately describe what is occurring in a
given situation. They define or redefine reality. In ethnography,
the evaluator attempts to understand what is happening in that
setting, how it is occurring, how the participants view that
occurrence, and how members of various groups participate within and
across these occurrences. With an ethnography, the evaluator does not
judge what occurs as either good or bad, as effective or ineffective;
rather, ethnographies describe the relevant information in a situation
and examine the recurrent patterns of behaviors. From this
characterization, the ethnographer defines the rules and processes for

the participants and for participation in the context described. The process of constructing an ethnography involves the development of a picture that explains the reality observed. Ethnographies, then, involve a three-step process: First, data is collected that helps describe what is occurring; second, a typology or model is developed that reflects these occurrences; and third, the validity of the model is tested and implications are drawn.

Ethnographies add to the information usually obtained by other traditional methodologies of program description in the following ways:

1. Ethnographies are concerned with the culture of the situation observed.

2. Ethnographies necessitate direct, on-site observations occurring over time and at times necessitate participation by the evaluator in the activities taking place.

3. The instruments used are field-based and are used to determine reality.

4. Ethnographies are holistic and characterize how various 'parts of a programmatic puzzle fit together.

5. Patterns and hypotheses developed result from an immersion in the field by determining what actually occurs in the field and not from predetermined theorizing of what should occur.

Ethnographies are not brief or selected samplings. Rather, they involve complete descriptions of the interrelationships of recurring

variables in a society under specified conditions as they affect or produce certain results and outcomes in that society. As such, ethnographies add a needed dimension to evaluative program descriptions.

Types of Information to Obtain

In addition to providing students with access to the skills and knowledge expected from traditional education of monolingual children, bilingual education programs are by definition different from traditional school programs in two important ways. First, two languages are used as media of instruction and language development is an integral part of the educational program. Secondly, the culture of the children (including attention to self-concept, learning styles, motivational styles, etc.) is an important consideration in the process of schooling. As a result, it is important in the program description section of the evaluation report to characterize the bilingual classroom and/or program in each of these three important areas: Use of language(s), incorporation of culture, and the instructional focus.

Since the quantity of information that can be otained about a particular classroom and/or program is very great, it may be necessary to develop and keep in mind organizers and categories of the kind of data needed. In each of the following three subsections, some salient questions for and suggested areas of data collection are provided as a

general guide to data collection. The suggestions included in each subsection are not intended to be all inclusive or totally complete. The evaluator/ethnographer, in consultation with the program manager, should be the best judge of what information should be obtained in order to adequately and accurately depict the processes operant in the bilingual program.

Language -- A description of how participants use language to communicate information in bilingual school settings, how they influence and persuade others, how they negotiate using language, etc. is necessary if one is to understand the bilingual classroom. This subsection includes suggested information that can be useful in describing language used in a given and specific situation. Additionally, this subsection provides general guidelines for information procurement that can characterize the use of language in bilingual classrooms and/or programs.

A language-use mapping technique (Green and Wallat, 1981)* has been suggested for describing the language used by both teachers and students in a bilingual classroom. Coding of information that can be gathered in a given language-use situation can be as follows:

---

* Green, J.L. and Wallat, C. "Mapping Instructional Conversations--A Sociolinguistic Ethnography." In: Ethnography and Language in Educational Settings, edited by J.L. Green and J. Wallat. Norwood, New Jersey: Ablex Publishing Corporation, 1981, 161-195.

4:13

1. Source = This category identifies the speaker involved in the interaction. Possible individuals may include the teacher, the student(s), or some other person(s). The interlocuters can be identified by a number code (i.e.; teacher=T, student(s)=S1, S2, etc.).

2. Form = Two forms of language used during instruction can be identified: The question and the response. (It is assumed that most interactions in an instructional period are either questions or responses to questions.) Three subcategories of responses can be expected:

    a. Type A response (+) = This type of response is both expected and consistent with the sociolinguistic content. This predictable response includes those that meet the social, cultural, psychological, and semantic aspects of the situation.

    b. Type B response (O) = These responses are not predictable given the preceding linguistic, topical, or social context. These responses may be spontaneous production of language by an involved student or they may be responses by a student not previously designated in the interaction.

    c. Type C response (-) = All nonresponses are included in this category.

3. Strategies - The purpose of the communication unit in its sociolinguistic context is mapped in this category. The various types of strategies include:

    a. Focusing = This occurs when an attempt is made to initiate or change the content of a discussion. A shift or focus function results.

    b. Ignoring = This is a nonverbal action resulting when no response occurs when one is required.

    c. Confirming (+) = The acceptance of a preceding response is indicated either verbally or nonverbally.

    d. Confirming (-) = Here the previous response is not accepted. A "no" may be indicated as a response to a request for confirmation.

4'14

e. Continuance = A verbal or nonverbal message suggesting that the listener is following the speaker's communication.

f. Extending = This category includes those messages that are designed to provide additional information about a topic.

g. Raising = This category of communication raises the level of a discussion.

h. Clarifying = Here, messages that explain or redefine are included.

i. Editing = In this strategy, shifts or changes in content, form, or strategy are signaled. Internal mediating of a message occurs.

j. Controlling = Messages that control an interaction or behavior of individuals are included in this category.

k. Refocusing = This type of language strategy reestablishes a previous line of thinking.

l. Restating = In this category are included those messages that repeat or refer to previous information.

4. Levels = The level of functioning of the interaction can be categorized into three groups. They are:

a. Factual = Literal recall of facts from memory are relayed in the message.

b. Interpretive = Inferential comprehension providing information not previously discussed is indicated in the interaction.

c. Applicative = This level of communicative interaction requires the information to be used in new ways or in novel contexts.

5. Ties = The basis of the message is often tied to some behavior or message of participants in an interaction. This relationship is indicated by the "ties" described. The four sources of ties include:

a. Teacher = Here, the message is tied to a teacher's goal or may be in response to the teacher's message.

4'5

b. Student = If the message is feedback to the student or if it extends to the student's purpose or it permits the student to build on his/her original message then it has a "student tie".

c. Instructional (media aide) = If the text, material or media aide triggers the message unit, then it is recorded as an instructional tie.

d. Context = A context tie occurs when the situation serves as the basis of the message.

Many situations may be mapped in obtaining information regarding the use of language by students, teachers, and others in bilingual programs. Some relevant contexts include: A typical lesson, the use of language in informal play situations, the use of language in formal non-structured classroom situations or other contexts as may be mutually agreed to by both the program managers and the evaluator.

An ethnography of the use of language in a bilingual classroom or program may provide insights regarding the following pedagogical questions:

1. How does the teacher use language in instructing children?

2. How do children use language with each other?

3. How do children use language with teachers, aides, parents, etc.?

4. How does the teacher respond to L2 language attempts of the bilingual children?

5. What is the language development climate in the classroom?

6. Which language (L1 or L2) is used with which interlocuters?

4''6

7. Which language (L1 or L2) is used with various topics?

8. Which language (L1 or L2) is used with which situations?

9. What language use opportunities exist in the bilingual classroom?

10. What opportunities exist for the student to experiment using language with different participants, on different topics, and for different purpose?

11. How does the teacher deliberately attempt to develop the language skills of the the bilingual children (either L1 or L2)?

12. How does the teacher respond to "nonadult-like" speech (grammar and phonology) from bilingual children?

13. What is the teacher's instructional registry?

14. How are children of varying linguistic and communicative proficiencies accommodated in the classroom?

15. To what extent and under what circumstances do students with different language dominances interact?

Instruction -- The volume of information that can be obtained in observing instruction that takes place in a bilingual program can be overwhelming. Oftentimes, the focusing of the perceptions of the observers may be necessary. Questions such as those posed below contain no preconceived hypotheses and can be used as a guide for the gathering of data as well as the interpretation of patterns perceived.

4''7

MAJOR CATEGORIES IN HENRY'S

CROSS-CULTURAL OUTLINE OF EDUCATION*

1. On what does the educational process focus?

2. How is the information communicated (what are the teaching methods employed)?

3. Who educates?

4. How does the person being educated participate?

5. How does the educator participate? What is his/her attitude?

6. Are some things taught to some and not to others?

7. Discontinuities in the educational process.

8. What limits the quantity and quality of information a child receives from a teacher?

9. What forms of conduct control (discipline) are used?

10. What is the relation between the intent and the results of education?

11. How long does the process of formal education last?

In addition, information and conclusions relating to the general climate of the classroom and/or program can be obtained by the evaluator.

---

* J. Henry. "The Cross Cultural Outline of Education." In: <u>Current Anthropology</u>, 4, 1960, 269-305.

Culture -- Information regarding culture is probably the most
difficult to obtain. As used in this context, culture does not mean
the surface trapping or artifacts associated with a group of people
like sarapes, clothing, festive or national holidays, and foods. It
would involve however, the recurrent behavior patterns, thinking,
perceptual, and learning styles. In addition, culture would include
information regarding the attitudes, values, communication styles as
well as the expected and the manifest norms found in a classroom or
found to be true of a program. Essentially, the evaluator attempts to
find out what is the culture of the bilingual component being
evaluated. The following is suggested as guides to the types of
information that may be obtained in conducting an ethnography of the
culture of a classroom/program.

A.   Behavior patterns

1.   What are the expected and manifest norms?

2.   What is considered by teachers and students to be
     acceptable behavior?

3.   What is the community's expectation of acceptance
     behavior?

4.   How do students respond to stress?

5.   How do students respond to instruction?

6.   How do students respond to independence as well as
     to structure or to the lack of structure?

7.   What discipline is imposed by the school? How is
     it imposed?

8.   What games do students play? (Game theory)

4'''9

9. How do students interact with elders and act toward peers?

10. What is the pattern of behavior characteristic of cross-ethnic or cross-linguistic groupings?

B. Perception and thinking

1. What topics are of concern to students?

2. Do students personalize or depersonalize topics of instruction?

3. Is instruction related to the student's personal milieu?

4. Are the following reasoning styles evidenced: Difference, magnitude, relationship, or appraisal?

C. Learning styles and motivation

1. Do students prefer to work as groups or individually?

2. Do students prefer visual or auditory presentations?

3. Is a preference shown for deductive or inductive presentations?

4. What motivates the students?

5. What reward systems are used in the classroom? How successful are they?

6. Do students impose structure in learning situations or do they need structure to be imposed for them?

7. What is the effect of peer pressure on motivation?

8. Is there a preference for personalized or depersonalized instruction?

9. On what tasks do students prefer to work?

D. Values

1. What value statements are heard from teachers or students?

2. Are student's values accepted?

410

3. What value prejudices are manifested? (Tastes, preferences, and individual choices)

4. What universal values are encouraged? (These are broad moral values covering such general topics as fair treatment, individual rights, equal opportunity under the law, acceptance of diversity of sex and race, and the respect for individual expression of diversity.)

E. Attitudes

1. What are the students' attitudes toward the school? the program? the instruction or instructors?

2. What are the teachers' attitudes toward the students? the program? the school?

3. What are the teachers' attitudes toward the cultural or linguistic groups represented in the school bilingual program?

4. What are the taboo topics not to be covered or discussed in school?

5. What status does the L1 (native language) have? the L2 second language)?

6. What is the intellectual climate of the school?

7. Can a temperament, mood, nature, etc., of the program be determined?

F. Communication styles

1. What language is used in informal situations? by whom?

2. What language is used in formal situations? by whom?

3. Are formal and informal codes of language used by students? when? with whom?

4. To what extent is the L1 (native language) or L2 (second language) encouraged or allowed throughout the day?

5. With whom can and do students use L1 and L2?

6. Which language is used in discussing which topics?

G.   Accommodation by the school

1.   To what extent is the classroom and program
     organization flexible in accommodating the
     differences found among the students?

2.   Are norms of behavior, values, attitudes, etc,
     imposed?

3.   What is the school's written or unwritten policy
     regarding the cultural and lingustic differences
     found in the bilingual program?

4.   How is the bilingual education program viewed by
     nonbilingual teachers and administrators?

Data Collection

Data collection involves a three step process. First sources of
information must be secured. Secondly, the information must be
organized. Finally, hypotheses must be verified.

Sources -- Information can be obtained from many sources in bilingual
programs. The four main sources include printed matter, participant
observation, non-participation, and the use of an informant.

Most programs have developed a brochure or some other printed matter
which describes the intent, operation, size, etc. of bilingual
schooling which have been used in dissemminating information
throughout the local community. Often, this printed matter is used
with parents in recruiting students, with school boards in discussing
approval of the bilingual program, and in the recruitment of teachers.
In addition, some local newspapers may have written articles to inform
the public of the local school's bilingual programs. The project's
proposals either to the state or federal funding agency can provide

other preliminary information that may be useful in understanding the nature of the bilingual program. With long standing programs, previous evaluation reports may provide some data that is still current and which may be useful in the initiation of the data collection period and the categories of data collection. As with all sources of information, the accuracy should be determined as a result of direct observation and by confirmation by program managers and teachers and as a result of direct observation.

In the participant observation strategy, the evaluator becomes involved in the normal activities of a classroom or program in order to gain direct access to information regarding classroom instruction and the normal operations of the bilingual programs. By becoming an "insider", the evaluator often becomes privy to information that otherwise might be kept from him/her. As participant observer, the evaluator seeks to blend into the operation of the program. At no time should the evaluator attempt to change or judge what is seen to occur.

One caution must be mentioned here. Participant observers need to deliberately remain intellectually separate from the program. Too close an identification with the participants in the program or its philosophy or structure may result in the evaluator sharing the biases of the group involved in the program. As a result, objectivity may be lost.

413

The nonparticipant observation strategy requires that the evaluator take note of occurrences without participating in them. The nature of this noninvolvement, however, may influence the behaviors of those who are observed. Frequently, those who are observed will change behaviors to reflect what they value as teachers and managers. As altered as the primary information may be, however, it still provides the evaluator with useful data regarding the ideals of behavior as viewed by the participants in the program.

In the informant strategy, the evaluator seeks out some knowledgeable person(s) from whom to secure information regarding the program. Structured interviews in which specific questions are asked of the informant may be used. With this strategy, specific information regarding unclear perceptions may be obtained. Care must be taken, of course, to avoid adopting the biases of the informant. Verification should always be part of any data gathering period.

Information Organization -- Information gathered from the various sources may be collected in field notes taken by the evaluator. These, of course, may be written either at the time of the occurrence and observation or soon after. Little time should lapse between the procurement of the information and the recording. Otherwise, memory and perceptions may become somewhat hazy.

In compiling and interpreting notes, every attempt should be made to accurately describe the situation. In this depiction, no prejudgment, bias, deletion, or predrawn conclusions should affect the information that is obtained.

414

Information	Analysis and Reflection	Synthesis Determination of Patterns	Verification	Conceptualization of Model
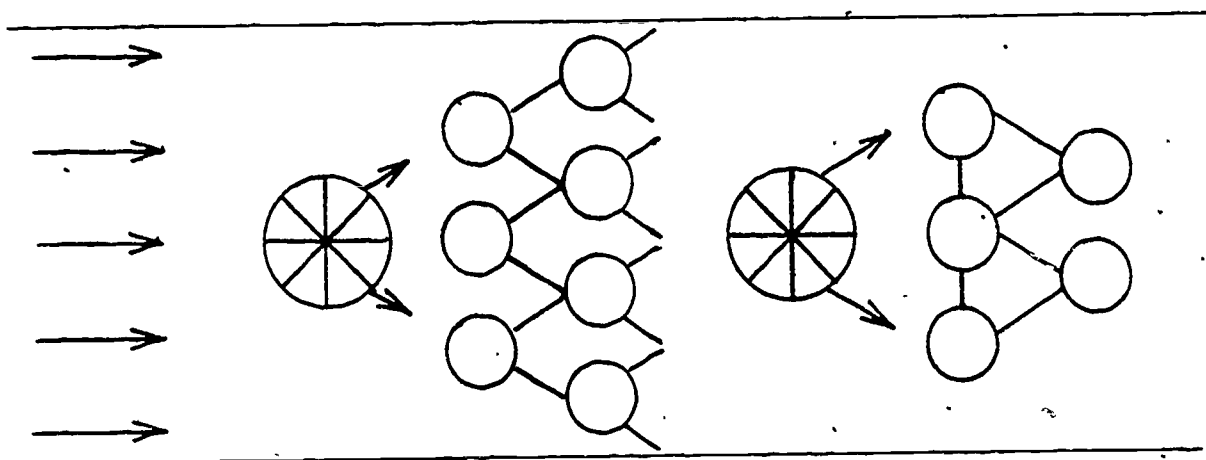
With the gathering of data, reflection should occur. In this examination, exisiting information should be analyzed and recurrent patterns, concepts, and common themes from natural grouping of informatiion are sought. Consistent feedback and verification of the perceived themes and patterns should occur along with the compilation of the data.

Upon completion of. the gathering of data and its analysis, a conceptual interpretation resulting in a model describing the program should result. This model should accurately reflect the actual program that was examined in that it was grounded in the data obtained. Finally, hypotheses and policy statements that may be requested from the program managers can be developed from the model that was constructed.

415

<u>Verification</u> -- The final step in the ethnographical process is the verification of the model. This can occur either in a written or oral form with the managers or other participants of the program taking part. The purpose for an ethnographical description of a bilingual program is to accurately depict the character of that program. This can be accomplished by the fusion of the perceptions of the ethnographers with those intents, understandings, and conceptualizations of the participants in the program.

The evaluator can validate perceptions used in the construction of the model by directly asking the participants in the program to review the draft of the program description part of the evaluation report. Indirectly, the same verification can be obtained by the evaluator by soliciting information related to the model by asking the following types of questions of the participants in the program:

1. Reportorial = these are literal questions of a who, where, what, how, and why nature. These questions are used primarily to verify the facts included in the description.

2. Posing = These questions challenge or act as devil's advocates by determining the strength of the participants' convictions and consistent use of various procedures. The model developed by the evaluator/ethnographer is true only as it is consistently accurate under various conditions.

3. Hypothetical = These questions are of a "what if..." or "what would happen if..." nature. Use of these questions can help the evaluator/ethnographer determine the strength of the model under unknown or novel circumstances.

4. Posing the ideal = Here, the program participant is asked to describe the ideal situation or the evaluator/ethnographer often solicits information regarding the aspirations and goals of the

participants as well as perceived faults with the existing program.

5. Offering interpretations or testing propositions on respondents = This allows the evaluator to tell the program participants about the propositions or patterns that are being used° in the construction of the model. If the program participants disagree with the conclusions drawn, then new information can be. secured and/or new patterns conceptualized.

Instrumentation

The following general outline or instrument is suggested for the gathering of field notes, their compilation into patterns, and their verification. Notes are recorded in four columns or sections. The first section, labeled observations, should contain the most information about what occurred in the field.

This section is followed by two "code" columns. In the first column, the evaluator can record whether the information is related to the use of L1 (native) or L2 (second) languages, and/or if the information provided insights regarding culture (C), or instruction (I). In the second column, more specific information regarding these three categories can be recorded. For example, information related to use of language on topics related to schooling (S), relationships (R), or home (H) can be indicated. Additionally, information related to values (V), behavior (B), learning styles/motivation (L), or perception/thinking (P) can be noted.

417

Observations	Code 1	Code 2	Patterns	Verification Implications Model

In the next section, preliminary ideas regarding recurrent patterns or perceived relationships can be recorded. This column records information regarding the meaning or interpretation of the data. Finally, in the last section, verifications that were secured regarding patterns perceived can be recorded. Implications for policy statements can be listed. Modification warranted from new information can also be indicated. This section can contain the preliminaries of the model describing the bilingual program.

419

In mapping classroom language use, the chart presented below can be used. Language from the classroom should be recorded (taped or written) and each line of interaction can then be numbered to correspond to the analysis represented on the chart.

TABLE 1
Descriptive Analysis of Instructional Conversations

Transcript line	Source		Form				Strategies													Level			Tie				Unit Resolution				
	Teacher	Student	Question	Response +	Response 0	Response -	Focusing	Ignoring	Confirming*	Confirming-	Continuance	Extending	Raising	Clarifying	Editing	Controlling	Refocusing	Restating	Factual	Interpret	Applicative	Teacher	Student	Lesson		Resolved Interaction	Unresolved Interaction	Instruction Sequence	Phase	Lesson
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19.	20	21	22	23	24	25	26	30	31	50	60	70

(Green and Wallat. _Ethnography and Language in Educational Settings_, 1981, 169.)

Final analysis of this language-use information can result from a seeking of recurrent patterns and the drawing of implications that help characterize a realistic model of instruction occurring in the bilingual classroom(s).

Information Usage and Reporting

Two types of reports can result from the evaluation suggested in this section. First, case studies depicting individual classrooms, and second, an ethnography of an entire program. Both types of reports are similar in content but differ in scope. Obviously, the ethnography involves the synthesis of information from several classrooms and, as a result, may be the more difficult to develop.

Similar to other types of programmatic descriptions, case studies and ethnographies include a discussion of the history of the program, a discussion of the student population in terms of language and ethnicity, a description of the program's facilities, number of students in the program, teachers involved, time allocations for various instructional components, and enumeration of the goals of the bilingual program. In addition, case studies and ethnographies include the following types of information:

1. Discussion of entry procedures and site selection.

2. Characterization of the procedures and site selection.

3. Description of the encounters (contacts) with the students, teachers, and managers in the program.

4. Discussion of the classroom(s) in terms of its culture, the use of language, and the organization of instruction.

5. Perceived patterns and the model resulting from the synthesis of these patterns.

6. Implications, conclusions, and policy statements.

SECTION III

WORKSHEETS AND FORMS USED WITH THE

DESIGNER'S MANUAL

This section provides program directors and evaluators with a complete

set of worksheets which are recommended in the Designer's Manual.

These worksheets are included in this volume in order to facilitate

their reproduction, dissemination and use. An index, by title and

worksheet number (when appropriate), follows this brief introduction.

422

INDEX TO SECTION II

Worksheet #1

DETERMINE AUDIENCE AND INFORMATION REQUIREMENTS FOR THE EVALUATION

Audience	Type of Information Needed	Reason Information is Needed	Date Information is Needed	Type of Report and Section to Emphasize in Cover Letter

III-133

Worksheet #2

SETTING PRIORITIES

Put a "1" by components which will receive a maximum emphasis, a "2"

by components receiving moderate emphasis, a "3" by components receiving

minimum emphasis, and an "x" by components which will not be evaluated.

Evaluation Components	Done last year (19__)	This Year (19__)	Next Year (19__)	Following Year (19__)
A. Program Description Information				
1. Project Overview	___	___	___	___
2. Instructional Approach	___	___	___	___
3. Project Management	___	___	___	___
B. Program Operations				
1. Instructional Program Implementation	___	___	___	___
2. Staff Development	___	___	___	___
3. Parent Involvement	___	___	___	___
C. Student Effects				
1. English Language Component	___	___	___	___
2. NonEnglish Language Component	___	___	___	___
3. NonEnglish Academic Component	___	___	___	___
4. Nonacademic Student Effects	___	___	___	___

426

Worksheet #3

TIMETABLE FOR THE EVALUATION

Year_____

Tasks	MONTHS											
	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	July
A. Plan Evaluation Design* 1. Determine which goals and objectives in each component to focus on 2. Cost out evaluation 3. Summarize design for administrator												
B. Project Description 1. Collect data – divide up 2. Summarize data 3. Review & analyze data for purposes of plan- ning its use in analyzing evaluation data												
C. Monitoring of Program Operations 1. Instructional Program Implementation a. Develop/select instruments b. Administer instruments c. Analyze data d. Interpret data e. Draft Report Section 2. Staff Training a. Develop/select instruments b. Administer instruments c. Analyze data d. Interpret data e. Prepare report section												

Worksheet #3

Year_____

Tasks	MONTHS											
	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	July
3. Parent Involvement a. Develop instruments b. Administer instru- ments c. Analyze data d. Interpret data e. Draft report section												
D. Evaluation of Language Components 1. Develop/select instruments 2. Administer instru- ments 3. Analyze data 4. Interpret data 5. Draft report section												
E. Evaluation of Non- language Academic Components 1. Select instruments 2. Administer instru- ments 3. Analyze data 4. Interpret data 5. Draft report section												
F. Evaluation of Non- academic Components 1. Develop/select in- struments 2. Administer instruments 3. Analyze data 4. Interpret data 5. Draft report section												
G. Report a. Compile report sections b. Review report c. Prepare final report												

428

Worksheet #3

TIMETABLE FOR THE EVALUATION
(Completed Sample)

Year_____

Tasks	MONTHS											
	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	July
A. Plan Evaluation Design*												
1. Determine which goals and objectives in each component to focus on	━*											
2. Cost out evaluation	━*											
3. Summarize design for administrator	━*											
B. Project Description												
1. Collect data - divide up	━━*											
2. Summarize data	━*											
3. Review & analyze data for purposes of planning its use in analyzing evaluation data	━*											
C. Monitoring of Program Operations												
1. Instructional Program Implementation												
a. Develop/select instruments												
b. Administer instruments												
c. Analyze data												
d. Interpret data												
e. Draft Report Section												
2. Staff Training												
a. Develop/select instruments												
b. Administer instruments												
c. Analyze data												
d. Interpret data												
e. Prepare report section												

*Last possible time to do this. Ideally this would also be done the previous spring.

Year_____

Tasks	MONTHS											
	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	July
3. Parent Involvement												
a. Develop instruments								—				
b. Administer instruments									—			
c. Analyze data									—			
d. Interpret data									—			
e. Draft report section										—		
D. Evaluation of Language Components												
1. Develop/select instruments												
2. Administer instruments	—									—		
3. Analyze data			—	**──						—		
4. Interpret data			—	**──								
5. Draft report section				— ── **								—
E. Evaluation of Non-language Academic Components												
1. Select instruments	—											
2. Administer instruments		—								—		
3. Analyze data			—**							—		
4. Interpret data			— **								—	
5. Draft report section			— **								—	
F. Evaluation of Non-academic Components												
1. Develop/select instruments	—											
2. Administer instruments		—								—		
3. Analyze data										—		
4. Interpret data										—		
5. Draft report section											—	
G. Report												
a. Compile report sections											—	
b. Review report											—	
c. Prepare final report											—	

**Partial analysis interpretation and reporting is done at this point.

430

Worksheet #4

OPERATING CHECKLIST FOR BILINGUAL EDUCATION
PROGRAM EVALUATION

EVALUATION STEPS

	Initiated	Completed

1. Planning, Managing, and Staffing the Evaluation

  1.1 Determination of audience for the evaluation

  1.2 Determine the focus of the evaluation

  1.3 Allocation of resources for evaluation activities

  1.4 Setting timelines for evaluation activities

  1.5 Develop overall management plan of evaluation

  1.6 Hire outside evaluator

  1.7 Assigning evaluation responsibilities to staff

2. Planning Data Collection for the Evaluation

  2.1 Description of program

  2.2 Description of students

  2.3 Description of program's goals

3. Planning Monitoring of Program Operations

  3.1 Description of program in operation

  3.2 Description of staff development activities

  3.3 Description of parent involvement

4. Planning Evaluation of Student Outcomes

  4.1 Selection of evaluation questions

  4.2 Selection of evaluation design for English, non-English, and other areas

  4.3 Selection of assessment instruments

431

Worksheet #4

	Initiated	Completed

4.4 Scheduling the testing for the evaluation

4.5 Designing procedures and scheduling data collection

4.6 Planning the analysis of the data

4.7 Reporting the results

5. Reporting the Results and Writing the Evaluation Report

5.1 Identification of audiences and reporting requirements

5.2 Establishing timelines

5.3 Outline for report

5.4 Analysis of the data

5.5 Selection of convening the interpretative panel for analyzing the data

5.6 Writing the evaluation report and planning presentations of results

432

Worksheet #5

EVALUATION SUMMARY GUIDE

Evaluation Questions	Evaluation Instruments	Source of Information	Data Collection		Data Analysis		Data Interpretation		Reporting	
			Who does it	When	Who	When	Who	When	Who	When
Program Description										
Monitoring Program Operations										
Student Outcomes										

433   434

ERIC

Worksheet #6

Part A

ESTIMATING LEVEL OF EFFORT REQUIREMENTS
FOR
DESCRIBING THE PROGRAM AND THE STUDENTS

Estimates are provided for three levels of evaluation activity for a given
year: (Different activity levels may occur each year).

a) Minimum - collect information from project proposal, school records,
    and project director.

b) Moderate - collect information from project proposal, school records,
    project director, and a sample (one to three people in each category)
    of project staff, bilingual teachers, district administrators and
    parents using structured interviews or questionnaires (For estimation
    purposes below, assume total number of people interviewed or
    receiving a questionnaire is eight).

c) Major - same as that described for "moderate," except more people in
    each category are interviewed or sent questionnaires plus classroom
    observations are conducted. (For estimation purposes below, assume
    the total number of people interviewed or receiving questionnaires is
    fifteen and that three classrooms are observed).

Task	Minimum	Moderate	Major	Your Estimate*
Level of Effort for a Given Year (in Days)				
1. Prepare, discuss with and obtain support of project director for proposed plan	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{2}$	_____
2. Prepare data collection instruments (using samples provided in Designer's Manual)	$\frac{1}{2}$	1	2	_____
3. Identify specific people or records from whom to collect data and make arrangements	1	3	5	_____
4. Collect data	$1\frac{1}{2}$	5	12	_____

*   Circle the estimate for any tasks which will be done by project staff
    instead of the external evaluator. Do not include these amounts in
    the total for the evaluator.

435

Part A

	Level of Effort for a Given Year (in Days)			
Task	Minimum	Moderate	Major	Your Estimate
5. Analyze and organize data for use in report and analysis of evaluation data collected for later components	2	4	6	_____
Total Days	(5½)	(13½)	(25½)	(____) Evaluator
				(____) Project Staff

436

Worksheet #6

Part B

ESTIMATING LEVEL OF EFFORT REQUIREMENTS
FOR
EVALUATING PROGRAM OPERATIONS

Estimates are provided for two levels of activity to be conducted during

a given year for each of three components - instructional methods, staff

development, parent involvement (Different levels of activity may occur

each year):

Instructional Methods

   a) Minimum - Conduct observations and interviews twice/year in only
                  two classrooms and have evaluator do interpretation.*

   b) Major -   Conduct observations and interviews three times/year in
                  all classrooms (for estimation purposes below, assume
                  total number of classrooms equal five) and have inter-
                  pretative panel.*

Staff Training

   a) Minimum - Same questionnaire given to trainees following each train-
                  ing session. Knowledge test not used and evaluator does
                  interpretation. (For estimation purposes below, assume
                  fifteen trainees and three training sessions).

   b) Major -   Same as for minimum, plus a knowledge test given pre and
                  post training, an end of project summary questionnaire
                  given and an interpretative panel is used. (For estima-
                  tion purposes below, assume fifteen trainees and three
                  training sessions).

Parent Involvement

   a) Minimum - Address only the issue of the extent to which the level of
                  parent involvement matched the planned level); evaluator
                  interprets data.

   b) Major -   Address all four proposed evaluation questions given on
                  page 81. (For estimation purposes below, assume ten
                  parents and eight staff members interviewed); have inter-
                  pretative panel.

---

* The alternative methods of interpreting the data are discussed in the
   staffing chapter which follows.

437

Part B

Level of Effort (in Days)

Task	Minimum	Major	Your Estimate*
Instructional Method			
1. Prepare, discuss with and obtain support of project director for proposed plan	1	1	_____
2. Prepare data collection instruments (using samples provided in Designer's Manual)	1	2	_____
3. Identify who to observe and interview and make arrangements to do so	1	1	_____
4. Collect data	½	5	_____
5. Analyze data	1	5	_____
6. Interpret data	1	3	_____
7. Write report section	1	2	_____
Total days	(6½)	(20)	( ) Evaluator
			( ) Project Staff
Staff Training			
1. Prepare, discuss with and obtain support of project director for proposed plan	½	½	_____
2. Prepare data collection instruments (using samples provided in Designer's Manual)	½	⅙	_____
3. Make arrangements for data collection	½	½	_____
4. Collect data - minimum (have trainer collect all data); major (have trainer collect all data except end of year questionnaire)	0	1½	_____

* Circle the estimate for any tasks which will be done by project staff instead of the external evaluator. Do not include these amounts in the total for the evaluator.

438

Worksheet #6

Part B

Task	Level of Effort (in Days)		Your Estimate
	Minimum	Major	
5. Analyze data	1½	7	_____
6. Interpret data and develop recommendations	1	3	_____
7. Write report section	1	1½	_____
Total days	(5)	(19)	(_____) Evaluator
			(_____) Project Staff

Parent Involvement

Task	Minimum	Major	Your Estimate
1. Prepare, discuss with and obtain support of project director for proposed plan	¼	½	_____
2. Prepare data collection instruments (using samples provided in Designer's Manual)	½	1	_____
3. Make arrangements for data collection	¼	1	_____
4. Collect data	½	6	_____
5. Analyze data	1	3	_____
6. Interpret data and develop recommendations	½	2	_____
7. Write report section	1	2	_____
Total days	(4)	(15½)	(_____) Evaluator
			(_____) Project Staff

439

Worksheet #6

Part C

ESTIMATING LEVEL OF EFFORT REQUIREMENTS
FOR
EVALUATING STUDENT OUTCOMES

Estimates are provided for two levels of activity to be conducted during a given year for each of four components--English language component, nonEnglish language component, nonlanguage academic component, and nonacademic student effects.

English Language Component

a) Minimum - Use norm-referenced evaluation design only; analyze by grade, subject, language used in instruction, and student proficiency; evaluator does interpretation.

b) Major - Use time series, norm-referenced and comparison group evaluation designs; analyze by grade, subject, language used in instruction, student proficiency factors; use interpretative panel.

NonEnglish Language Component

a) Minimum - Use existing test and do norm-referenced evaluation design only; analyze by grade, subject, language used and student proficiency; evaluator does interpretation.

b) Major - Develop own test; use time series, norm-referenced and comparison designs; analyze by grade, subject, language used in instruction and student proficiency; use interpretative panel.

Nonlanguage Academic Component

a) Minimum - Use existing test, compare to national norms; analyze only by grade; evaluator does interpretation.

b) Major - Use existing test, compare to national norms; analyze by grade and two other key factors; use interpretative panel.

Nonacademic Student Effects

a) Minimum - Use only a published self concept measure; analyze only by grade and student proficiency; evaluator does interpretation.

b) Major - Use all proposed evaluation questions and data collection instruments; analyze by grade and student proficiency; use interpretative panel.

440

Part C

Task	Level of Effort (in Days)		
	Minimum	Major	Your Estimate*

English Language Component

Task	Minimum	Major	Your Estimate*
1. Prepare, discuss with and obtain support of project director for proposed plan	½	½	_____
2. Select appropriate tests	1	5	_____
3. Train test administrators and make arrangements for testing	1	2	_____
4. Supervise testing - minimum (one day each, pre- and post-testing); major (monitor all testing)	2	14+	_____
5. Analyze data - minimum (prepare achievement data for standard computer analysis); major (prepare data for standard computer analysis, for several analyses)	3	8+	_____
6. Interpret results	2	10	_____
7. Write report section	3	10+	_____

	Minimum	Major	
Total days	(12½)	(50+)	(_____) Evaluator
			(_____) Project Staff

NonEnglish Language Component

Task	Minimum	Major	Your Estimate*
1. Prepare, discuss with and obtain support of project director for proposed plan	½	½	_____
2. Select appropriate tests	1	5+	_____
3. Train test administrators and make arrangements for testing	1	2	_____

* Circle estimate for any tasks which will be done by project staff instead of the external evaluator. Do not include these amounts in the total for the evaluator.

Part C

Task	Level of Effort (in Days)		
	Minimum	Major	Your Estimate*
4. Supervise testing - minimum (one day each, pre- and post-testing); major (monitor all testing)	2	10+	_____
5. Analyze data - minimum (prepare achievement data for standard computer analysis); major - (prepare data for standard computer analysis for several analyses)	2	8	_____
6. Interpret results	2	10	_____
7. Write report section	2	10	_____
Total days	(10½)	(45½)	(_____) Evaluator
			(_____) Project Staff

Nonlanguage Academic Component

1. Prepare, discuss with and obtain support from project director for proposed plan	½	½	_____
2. Select appropriate tests - minimum (become familiar with district tests); major (review commercial achievement tests and match to curriculum)	1	5	_____
3. Train test administrators and make arrangements for testing	1	2	_____
4. Supervise testing - minumum (one day each, pre- and post-testing)	2	10+	_____
5. Analyze data - minimum (prepare achievement data for standard computer analysis); major (prepare data for standard computer analysis for several analyses)	2	8	_____

412

Worksheet #6

Part C

Task	Level of Effort (in Days)		Your Estimate*
	Minimum	Major	
6. Interpret results	2	10	_____
7. Write report section	2	8	_____
Total days	(10½)	(45+)	(_____) Evaluator
			(_____) Project Staff

Nonacademic Component

Task	Minimum	Major	Your Estimate
1. Prepare, discuss with and obtain support from project director for proposed plan	½	½	_____
2. Select or develop appropriate instruments	1	4	_____
3. Train test administrators and make arrangements for testing and other data collection	1	4	_____
4. Analyze data - minimum (prepare for standard computer analysis)	2	8	_____
5. Interpret results	2	10	_____
6. Write report section	2	8	_____
Total days	(8½)	(34½)	(_____) Evaluator
			(_____) Project Staff

Worksheet #6

Part D

SUMMARY OF ESTIMATED LEVEL OF EFFORT
REQUIREMENTS AND
ASSOCIATED COSTS

Summary of Days	Evaluator	Project Staff
Program Description	_____	_____
Monitoring Program Operations	_____	_____
Instructional Methods	_____	_____
Staff Training	_____	_____
Parent Involvement	_____	_____
Evaluating Student Effects		
English Language Component	_____	_____
NonEnglish Language Component	_____	_____
Nonlanguage Academic Component	_____	_____
Nonlanguage Student Effects	_____	_____
	(_____)	(_____)

Total days X evaluator cost per day = Total evaluator cost per year

_____ X _____ = _____

411

Worksheet #6

Part D

Additional Cost Items	Program Description	Costs (in Dollars) Monitoring Program Operations	Evaluating Student Effects
1. Secretary time	_____	_____	_____
2. Printing	_____	_____	_____
3. Mailing	_____	_____	_____
4. Other			
a._____	_____	_____	_____
b._____	_____	_____	_____
c._____	_____	_____	_____
d._____	_____	_____	_____
e._____	_____	_____	_____
Totals	_____	_____	_____
Total Evaluator Costs	_____	_____	_____
Total Additional Costs	_____	_____	_____
Total Costs for Evaluation	_____	_____	_____

445

Worksheet #7

DATA COLLECTION FORM FOR INFORMATION FROM THE

PROJECT PROPOSAL AND OTHER RECORDS

The project proposal and various project or school records should be reviewed to obtain the indicated information.

(C) 1. What are the major project goals?

Linguistically_____

_____

_____

Culturally_____

_____

_____

Academically_____

_____

_____

(S) 2. What is the pattern of predominant languages among the student population?.

_____

_____

_____

(S) 3. What is the approximate achievement level (in languages, other academic and nonacademic areas) of students within the various language categories? Report separately for each language group.

Language achievement_____

_____

_____

Other academic achievement_____

_____

_____

C = refers to program context   G = refers to program goals
S = refers to program students   P - refers to instructional programs

Nonacademic achievement_____

_____

_____

(P) 4. What grade levels and how many classrooms are served by the project?

_____

_____

_____

(P) 5. What portion of the school day is covered?_____

_____

_____

·(C) 6. Describe the following community characteristics

   a. Languages spoken (approximate percentage speaking each language)

_____

_____

_____

   b. Ethnicity (approximate percentage of each)_____

_____

_____

   c. Socioeconomic status (general description based on type of
      employment)

_____

_____

_____

   d. Size of community_____

_____

_____

447

(C) 7. Describe the local education agency as follows:

a. Size_____

b. Financial status of district_____

_____

c. Facilities available for project_____

_____

_____

(C) 8. Describe the following school characteristics

a. Number of bilinguals in school by language group_____

_____

_____

b. Number of students in bilingual program_____

_____

_____

c. Bilingual mix in the classrooms_____

_____

_____

(P.) 9. Describe the project staff and its organization. List each member of the staff, the percentage of time committed to the project and their qualifications

Title	Name	Percentage time	Qualifications

448

_____

_____

_____

b. Describe the organizational structure of the project_____

_____

_____

_____

_____

c. What selection procedures are used in selecting staff members?

_____

_____

_____

(P)   10. Describe the project director's role with respect to the following items:

a. Funds and budgets_____

_____

_____

b. Public relations_____

_____

_____

c. Administration_____

_____

_____

d. Overseeing instruction_____

_____

_____

449

e. Staff training_____

_____

_____

f. Developing and ordering materials and equipment_____

_____

_____

g. Staff recruiting and hiring_____

_____

_____

Worksheet # 8

PROGRAM DIRECTOR INTERVIEW SCHEDULE

(G)   1. The goals of the program as stated in the proposal are as follows:
         (Present the goals orally or in writing as obtained from the pro-
         posal.)

         _____

         _____

         _____

         What evidence will show that these goals have been met?_____

         _____

         _____

         _____

         Which goals have the highest priority?_____

         _____

         _____

         _____

(G)   2. How would you define the program as to the extent which it is a
         maintenance, transitional or partial bilingual program?

         _____

         _____

         _____

(C)   3. Describe the mobility of the community including any specific data
         available

         _____

         _____

         _____

C = refers to program content
G = refers to program goals
S = refers to students
P = refers to instructional program

451

III-163

(P)   4. How are students assigned to classrooms?_____

_____

_____

_____

(S)   5. Describe the student entry and exit criteria and procedures. Do the
          actual procedures conform to the planned procedures?

_____

_____

_____

(P)   6. Describe the scheduling of instruction including daily schedules and
          grouping and regrouping across and within classes

_____

_____

_____

(P)   7. Describe the staff and its organization in terms of the following
          dimensions

          a. Staff members' time commitments_____

_____

_____

          b. Staff organizational structure_____

_____

_____

          c. Staff qualifications_____

_____

_____

          d. Staff selection procedures_____

_____

_____

(P) 8. What is your general leadership style as program director?

_____

_____

_____

_____

_____

(P) 9. What is your role as program director with respect to each of the following areas?

a. Funds and budgets_____

_____

_____

b. Public relations_____

_____

_____

c. Administration_____

_____

_____

d. Overseeing instruction_____

_____

_____

e. Staff training_____

_____

_____

e. Developing and ordering materials and equipment_____

_____

_____

f. Staff recruiting and hiring_____

_____

_____

(P) 10. What is the teacher's role in the following areas?

a. Planning instruction_____

_____

_____

b. Implementing instruction_____

_____

_____

c. Noninstructional responsibilities_____

_____

_____

(P) 11. What is the role of the aides in the program?_____

_____

_____

_____

(P) 12. What is the role of other staff members such as the following?

a. Instructional coordinator_____

_____

_____

b. Community coordinator_____

_____

_____

c. Evaluator_____

_____

_____

d. Other (please specify)_____

_____

_____

e. Other (please specify)_____

_____

_____

(P) 13. Describe the program's staff development activities related to the
following aspects.

a. Needs assessment_____

_____

_____

b. Structure of training (pre-service and in-service)_____

_____

_____

c. Characteristics of training

· (1) Appropriateness for staff of differing levels of knowledge
and experience

_____

_____

_____

(2) Practicality_____

_____

_____

(3) Coordination with degree programs_____

_____

_____

455

(4) Integration with other training _____

_____

_____

d. Audiences trained (program and/or nonprogram staff) _____

_____

_____

_____

(P) 14. Describe the involvement of the community and parents with respect
         to the following items.

         a. Parent involvement in school affairs _____

         _____

         _____

         b. Community input in program planning _____

         _____

         _____

         c. Evidences of community support for the program _____

         _____

         _____

         d. Parent education _____

         _____

         _____

         e. Parent conferences/counseling _____

         _____

         _____

456

(P) 15. Describe the means of communication of the following groups.

a. Among program staff_____

_____

_____

b. Program staff with the following nonproject staff:

(1) Principals_____

_____

_____

(2) Other district administrators_____

_____

_____

(3) Nonprogram teachers_____

_____

_____

(4) School board_____

_____

_____

(P) 16. What means are used to disseminate project information to school personnel, parents and community?

_____

_____

_____

457

Worksheet #9

PROGRAM STAFF INTERVIEW SCHEDULE

- (Check one) _____Program staff   _____Bilingual teacher

(G)   1. What is the intended content of instruction (i.e: the <u>theoretical</u> curriculum) with respect to the following matters?

        a. Content areas covered_____

        _____

        _____

        b. Relationship of content to program goals_____

        _____

        _____

        c. Who determines the content?_____

        _____

        _____

        d. What articulation is there between program content and the extant district curriculum?

        _____

        _____

        _____

(P)   2. Describe the presentation of content with respect to the following items.

        a. Type of instructional model or theory (e.g. concurrent, alternate week/day, preview-review, half day, resource room, and/or bilingual aide)

        _____

        _____

        _____

        _____

C = refers to program content
S = refers to students
G = refers to program goals   III-171
P = refers to instructional program

458

b. Organizational practices (e.g. individualized, large group,
   learning centers, peer tutoring, small group instruction, and/or
   team teaching)

_____

_____

_____

_____

(P)   3. Describe the methodologies employed for bilingual education with
           respect to the following items.

        a. Language of instruction

           (1) General language use plan of teacher and student over length
               of program

_____

_____

_____

           (2) Daily instructional time in each language_____

_____

_____

_____

           (3) Variations for different student groups_____

_____

_____

_____

           (4) Criteria for establishing language of instruction_____

_____

_____

_____

459

b. Approach to nonstandard forms

   (1) Acceptance_____

_____

_____

   (2) Form of corrections_____

_____

_____

c. Approach to second language instruction

   (1) Formal instruction_____

_____

_____

   (2) Functional use of second language for content instruction
      and other activities

_____

_____

d. Approach to reading instruction

   (1) Language in which students learn to read_____

_____

_____

   (2) Criteria for beginning reading in second language_____

_____

_____

_____

(P) 4..Describe the specific instructional methodologies used in each
     subject area

_____

_____

_____

<center>460</center>

_____

(P)   5. Describe those aspects of the program that are intended to motivate
         students and improve their self-concept with respect to the follow-
         ing matters:

      a. Appropriate content and language of instruction

         (1) Using L_1 for instruction_____

         _____

         _____

         (2) Accepting language of the student_____

         _____

         _____

         (3) Content that relates to experiences of students_____

         _____

         _____

         _____

         (4) Culturally relevant material_____

         _____

         _____

      b. Improved affective climate

         (1) Placing equal value on both languages and cultures____

         _____

         _____

         _____

         (2) Insuring student success_____

         _____

         _____

(3) Involving parents _____

_____

_____

(4) Teacher as a role model _____

_____

_____

c. Discipline approach

(1) Philosophy _____

_____

_____

(2) Guidelines/approach to control _____

_____

_____

(3) Special reward systems (e.g. prizes and privileges) _____

_____

_____

(P)   6. What materials are used within each of the following categories?

a. Core materials in use

(1) Commercial _____

_____

_____

(2) Locally developed _____

_____

_____

b. Appropriateness

    (1) Linguistic _____

_____

_____

    (2) Cultural _____

_____

_____

(P)   7. Describe the role of each of the following personnel in the classroom

    a. Teachers _____

_____

_____

    b. Aides _____

_____

_____

    c. Parents _____

_____

_____

    d. Peers _____

_____

_____

    e. Resource staff _____

_____

_____

463

(P)   8. Describe the program director's work with respect to the following:

a. Leadership style_____

_____

_____

b. Role or responsibilities in connection with each of the following

(1) Funds and budgets_____

_____

_____

(2) Public relations_____

_____

_____

(3) Administration_____

_____

_____

(4) Overseeing instruction_____

_____

_____

(5) Staff training_____

_____

_____

(6) Developing and ordering materials and equipment_____

_____

_____

_____

(7) Staff recruiting and hiring_____

_____

_____

.464

Worksheet #10

LOCAL AND DISTRICT ADMINISTRATORS INTERVIEW SCHEDULE

(G)   1. Describe the school district's general goals_____

_____

_____

(C)   2. What is the school district's philosophy toward language and cul-
tural diversity?

_____

_____

(P)   3. To what extent is there articulation of program content with the
existing district curriculum?

_____

_____

(P)   4. What is the relationship between the program staff and each of the
following categories of district personnel? Comment specifically
on program acceptance.

a. Principals_____

_____

_____

b. Central office administrators_____

_____

_____

c. Nonproject teachers_____

_____

_____

d. The school board_____

_____

_____

C = refers to program context
S = refers to students   III-179 495
G = refers to program goals
P = refers to instructional program

(P)   5. Describe the dissemination of program information to the following
          two groups.

          a. School personnel _____

          _____

          _____

          b. Parents and the community _____

          _____

          _____

Worksheet #11

CLASSROOM OBSERVATION SCHEDULE

Date:_____ Instructor:_____

Class Hour:_____ Observer:_____

I. List the content areas covered during the class period as they occur.

1. time started:_____
    time ended:_____
2. time started:_____
    time ended:_____
3. time started:_____
    time ended:_____
4. time started:_____
    time ended:_____
5. time started:_____
    time ended:_____
6. time started:_____
    time ended:_____
7. time started:_____
    time ended:_____
8. time started:_____
    time ended:_____
8. time started:_____
    time ended:_____

II. List the instructional methodologies employed as they occur during the period:

Summary statement (enter at end of period):

III. The beginning and ending time for each of the instructional components of the close period can be indicated in item I above. In addition the observer can indicate here estimates of how much time fell within each of three categories during each three minute segment of the class period.

Three Minute Period	Off-task Time	On-task Students Active*	On-task Students Passive	Three Minute Period	Off-task Time	On-task Students Active*	On-task Students Passive
1				11			
2				12			
3				13			
4				14			
5				15			
6				16			
7				17			
8				18			
9				19			
10				20			

* One or more students engaged in behavior for which they get feedback from the teacher.

IV. Describe any variations ·in
teaching approach used for
different student groups
(include any variations in
pace of instruction for in-
dividuals or groups)

V. Describe any evidence of self-
concept development and mo-
tivation including indicators
of (a) accepting the language
of the student and (b) con-
tent that relates to the
experience of the students

Summary statement (enter at end of
period)

Summary statement (enter at end of
period)

VI. Describe the role of all of the following personnel who were present
in the classroom.

(1) Teachers:

(2) Aides:

(3) Parents:

(4) Parents:

(5) Resource staff:

488

Worksheet #12

PROGRAM OPERATIONS INTERVIEW SCHEDULE FOR TEACHERS

1. What are the major instructional methods that you employ?

_____

_____

_____

2. Why do you use these particular methods, i.e. are these particular methods directed to particular instructional objectives?

_____

_____

_____

3. Are there other instructional methods that you would prefer to employ if it were not for various circumstantial constraints that you face?

_____

_____

_____

4. If so, what are these constraints?_____

_____

_____

_____

5. What program changes would you recommend that would facilitate your efforts to provide the best instruction possible?

_____

_____

_____

469

6. How typical would you say the class period that we observed was in terms of the instructional approach used and the nature and amount of interaction with students? How was it atypical?

_____

_____

_____

7. How do the entry and exit criteria and procedures actually used differ from those planned for the project? (Interviewer: Be prepared to describe the planned procedure. This information can be obtained through W #13.)

_____

_____

_____

470

Worksheet #13

STAFF DEVELOPMENT QUESTIONNAIRE

Name of training activity_____

Date of training_____

Name of person completing questionnaire (optional)_____

1. In general, what expectations did you have for the staff training provided as part of this project?

   _____

   _____

   _____

2. To what extent were these expectations met?_____

   _____

   _____

   _____

3. Based on your knowledge of the objectives for this staff training, which objectives do you think have been met?

   _____

   _____

   _____

4. Which objectives do you think have not been met?_____

   _____

   _____

   _____

471

Worksheet # 14

INTERVIEW SCHEDULE FOR LEADER OF PARENT ACTIVITIES

1. What is the general scope of parent involvement which was planned for the project this year?

_____

_____

_____

2. To what extent have these goals changed since the beginning of the project year?

_____

_____

_____

3. To what extent have these goals been met?_____

_____

_____

_____

4. Are you satisfied with the level of parent involvement? Is the staff as a whole satisfied?

_____

_____

_____

5. To what extent and in what ways has parent involvement changed over the life of the project?

_____

_____

_____

6. What are the most positive aspects of parent activities?_____

_____

7. What aspects of the parent involvement have the most potential for improvement?

_____

_____

_____

8. What changes are you recommending be made in parent activities in the future?

_____

_____

_____

473

Worksheet # 15

PARENT INTERVIEW SCHEDULE

(C) 1. To what extent have you been involved in school affairs?

_____

_____

_____

(P) 2. To what extent are you aware that the school has gotten suggestions and reactions from the community in planning its bilingual education program?_____

_____

_____

(C) 3. How much community support do you believe there is for the bilingual education project?_____

_____

_____

(P) 4. How much education has the school district provided for you as a parent as part of the bilingual education project?

_____

_____

_____

(P) 5. To what extent are you aware that the school has provided parent counseling or conferences?_____

_____

_____

(P) 6. What information have you received about the bilingual education project from the school district?

_____

_____

_____

P = refers to instructional programs
C = refers to program context

474

(P) 7. The bilingual program has as one of its goals (fill in the goals related to parent involvement). To what extent do you think this goal has been met? What evidence do you know of that indicates this goal has been met?_____

_____

_____

P = refers to instructional programs
C = refers to program context

4.5

Worksheet #16

EVALUATION DESIGN WORKSHEET

I. Subject Area and Language:

Tests: NRT: _____

Other: _____

II. Program Student Description:

Grade Level(s): _____

Language Skills: English: _____ Other: _____

Other Descriptors: _____

III. Comparison Data (Groups and Years)

Student Groups	Test Code	Current Year	Earlier Years
A.			
B.			
C.			

IV. Evaluation Questions

Student Performance

1. Relative Standards
   of Performance: _____

   _____

   _____

   _____

2. Absolute Standards
   of Performance: _____

   _____

   _____

476

Worksheet #17

BILINGUAL PROGRAM EVALUATION REPORT OUTLINE

	Section	Check When Done
I. Executive Summary (3-5 pages)	_____	_____
A. Overview of project goals, numbers and types of students served, instructional approach and evaluation design	_____	_____
B. Summary of findings	_____	_____
1. Instructional methods	_____	_____
2. Parent involvement component	_____	_____
3. Staff development	_____	_____
4. Student outcomes	_____	_____
a. English language	_____	_____
b. NonEnglish language	_____	_____
c. Nonlanguage academic	_____	_____
d. Nonacademic student effects	_____	_____
C. Recommendations	_____	_____
II. Program Overview and Background (2 pages)	_____	_____
A. Context of program including community characteristics, LEA, and school description	_____	_____
B. Student description and needs	_____	_____
C. Program's major goals	_____	_____
D. Program methods	_____	_____
E. Size, scope. and definition of the program	_____	_____

	Section	Check When Done
III. Description of Evaluation (3 pages)		
A. Purposes and audiences		
B. Evaluation staff and roles		
C. Design		
1. Questions addressed (includes standards for comparison)		
2. Constraints and questions not addressed		
D. Relationship to past and future years' evaluations		
IV. Program and Student Description		
A. Target students		
1. Definition of project student		
2. Student selection criteria and method		
a. Tests and cut-off scores used		
b. Role of teacher judgment		
c. Role of parent wishes		
d. Method of combining criteria		
3. Exit criteria and follow-up		
4. Student turnover		
5. Student characteristics at beginning of year		
a. Language proficiency		
(1) English		
(2) NonEnglish language		
b. Achievement level		
c. Biographic data		

4.5

III-194

	Section	Check /When Done
B. Instructional Approach	———	———
1. Self-concept and cultural emphasis	———	———
2. Content of instruction	———	———
3. Presentation of content	———	———
a. instructional model or theory	———	———
b. methodologies for bilingual education	———	———
c. specific methodologies for each subject area	———	———
d. role of presentation	———	———
e. self concept development and motivation	———	———
f. materials	———	———
a. personal role in classrooms	———	———
4. Scheduling	———	———
C. Program Management		
1. staff organization	———	———
2. staff roles	———	———
a. Project Director	———	———
b. Teachers	———	———
c. Aides	———	———
d. Other staff	———	———
3. staff development	———	———
a. needs assessment	———	———
b. structure of training	———	———
c. characteristics of training	———	———
d. audiences trained	———	———
4. parents and community	———	———
5. communication	———	———
6. dissemination of project information	———	———

	Section	Check When Done
VII. Parent Involvement Component	_____	_____
A. Goald and objectives	_____	_____
B. Description of activities to be evaluated	_____	_____
C. Evaluation procedures	_____	_____
1. Measures used	_____	_____
2. Data collection procedure	_____	_____
3. Analysis procedures	_____	_____
D. Evaluation Outcomes	_____	_____
1. Results (including unanticipated outcomes)	_____	_____
2. Interpretations	_____	_____
VIII. Staff Development	_____	_____
A. Goals and objectives	_____	_____
B. Description of activities to be evaluated	_____	_____
C. Evaluation Procedures	_____	_____
1. Measures used	_____	_____
2. Data collection procedures	_____	_____
3. Analysis procedures	_____	_____
D. Evaluation	_____	_____
1. Results (including unanticipated outcomes)	_____	_____
2. Interpretation	_____	_____
V. Student Effects	_____	_____
A. English language component	_____	_____
1. Goals and objectives	_____	_____

450

	Section	Check When Done

V.   Continued

    2. Evaluation procedures

       a. Measures used

       b. Data collection procedures

       c. Analysis procedures

    3. Evaluation Outcomes

       a. Results (including unanticipated outcomes)

       b. Interpretation

       c. Recommendations

B. NonEnglish language component

    1. Goals and objectives

    2. Evaluation procedures

       a. Measures used

       b. Data collection procedures

       c. Analysis procedures

    3. Evaluation Outcomes

       a. Results (including unanticipated outcomes)

       b. Interpretation

       c. Recommendations

C. Nonlanguage academic component

    1. Goals and objectives

    2. Evaluation Procedures

       a. Measures used

       b. Data collection procedures

       c. Analysis procedures

    3. Evaluation Outcomes

       a. Results (including unanticipated outcomes)

       b. Interpretation

       c. Recommendations

	Section	Check When·Done

VI. Continued

   D. Nonacademic component

     1. Goals and objectives

     2. Evaluation procedures

       a. Measures used

       b. Data collection procedures

       c. Analysis procedures

     3. Evaluation Outcomes

       a. Results (including unanticipated outcomes)

       b. Interpretation

Recommendations

A. Program Operations

   1. Instructional approach

   2. Program management

B. Parent involvement

C. Staff Development

D. Student Effects

VII. Program Operations Evaluation

   A. Instructional Approach

     1. Goals and objectives

     2. Description of activities to be evaluated

     3. Evaluation procedures

       a. Measures used

       b. Data collection procedures

       c. Analysis

     4. Evaluation outcomes

       a. Results

       b. Interpretations

452

	Section	Check When Done
B. Program Management	———	———
1. Goals and objectives	———	———
2. Description of activities to be evaluated	———	———
3. Evaluation procedures	———	———
a. Measures used	———	———
b. Data collection procedures	———	———
c. Analysis	———	———
4. Evaluation Outcomes	———	———
a. Results	———	———
b. Interpretations	———	———

PROGRAM INFORMATION ACQUISITION ·FORM

Type of Information	Available Instruments	Should it Be Done? (Yes or No)	If Yes When? (List Date)
A. Program Overview			
1. Grades and number of classrooms served	W #7	_____	_____
2. Portion of day covered	W #7	_____	_____
3. Definition of program – maintenance, transitional, partial bilingual	Proposal W #8	_____	_____
B. Instructional Approach			
1. Self concept and cultural emphasis	W #9	_____	_____
2. Content of instruction	W #9	_____	_____
a. Content areas covered	W #9	_____	_____
b. Who determines content	W #9	_____	_____
c. Other content features	W #9	_____	_____
(1) Relationship of content to goals	W #9	_____	_____
(2) Articulation of project content with existing district curriculum	W #9	_____	_____
3. Presentation of content	W #9	_____	_____
a. Instructional model or theory	W #9	_____	_____
(1) Type, e.g., concurrent, alternate day/week, preview-review, half day, resource room, and/or bilingual aide	W #9	_____	_____
(2) Organizational practices, e.g., individualized, large group, learning centers, peer tutoring, small group instruction, and/or team teaching	W #9	_____	_____

484

Type of Information	Available Instruments	Should it Be Done? (Yes or No)	If Yes When? (List Date)
b. Methodologies for bilingual education	W #9		
(1) Language of instruction	W #9		
(a) General language use pla.. of teacher and student over length of project	W #9		
(b) Daily instructional time in each language	W #9		
(c) Variations for different student groups	W #9		
(d) Criteria for establishing language of instruction	W #9		
(2) Approach to nonstandard forms	W #9		
(a) Acceptance	W #9		
(b) Form of corrections	W #9		
(3) Approach to second language instruction	W #9		
(a) Formal instruction	W #9		
(b) Functional use of second language for content·instruction and other activities	W #9		
(4) Approach to reading instruction	W #9		
(a) Language in which students learn to read	W #9		
(b) Criteria for beginning reading in second language	W #9		

Type of Information	Available Instruments	Should it Be Done? (Yes or No)	If Yes When? (List Date)
c. Specific methodologies for each subject area	W #9		
d. Rate of presentation	W #11		
(1) Variation in pace of instruction for individuals or groups	W #11		
(2) Time on task	W #11		
(a) Minutes per day per content area (see scheduling, 5.b.)	W #11		
(b) Proportion of time student is actively engaged in producing responses for which s/he gets feedback	W #11		
e. Self-Concept Development and Motivation (aspects of program that may motivate students and improve their self-concept	W #9 W #11		
(1) Appropriate content and language of instruction	W #9 W #11		
(a) Using L_1 for instruction	W #9 W #11		
(b) Accepting the language of student	W #9 W #11		
(c) Content that relates to experience of students	W #9 W #11		
(d) Culturally relevant material	W #9 W #11		
(2) Improved affective climate	W #9 W #15		
(a) Placing equal value on both languages and cultures	W #9		
(b) Insuring student success	W #9		
(c) Involving parents	W #9 W #15		

Type of Information	Available Instruments	Should it Be Done? (Yes or No)	If Yes When? (List Date)
(3) Discipline approach	W #9	_____	_____
(a) Philosophy	W #9	_____	_____
(b) Guidelines approach to control	W #9	_____	_____
(C) Special reward systems, e.g., prizes and privileges	W #9	_____	_____
f. Materials	W #9		
(1) Core materials in use	W #9		
(a) Commercial	W #9	_____	_____
(b) Locally developed	W #9	_____	_____
(2) Appropriateness	W #9		
(a) Linguistic	W #9	_____	_____
(b) Cultural	W #9	_____	_____
g. Personnel roles in Classroom	W #9		
(1) Teachers	W #9 W #11	_____	_____
(2) Aides	W #9 W #11	_____	_____
(3) Parents	W #9 W #11	_____	_____
(4) Peers	W #9 W #11	_____	_____
(5) Resource staff	W #9 W #11	_____	_____
4. Scheduling	W #8	_____	_____
a. Grouping and regrouping	W #8	_____	_____
(1) Across classes	W #8	_____	_____
(2) Within classes	W #8	_____	_____
b. Daily schedules	W #8	_____	_____

497

Type of Information	Available Instruments	Should it Be Done? (Yes or No)	If Yes When? (List Date)
C. Management	W #7 W #8	_____	_____
1. Staff Organization			
a. List of staff members and time commitment	W #7 W #8	_____	_____
b. Organizational structure	W #7 W #8	_____	_____
c. Qualifications	W #8	_____	_____
d. Selection procedures	W #8	_____	_____
2. Staff Roles (describe responsibilities)	W #8 W #9	_____	_____
a. Project Director	W #8 W #9	_____	_____
(1) Style of leadership as determined by project and LEA	W #8	_____	_____
(2) Funds and budgets	W #8 W #9	_____	_____
(3) Public relations	W #8 W #9	_____	_____
(4) Administration	W #8 W #9	_____	_____
(5) Overseeing instruction	W #8 W #9	_____	_____
(6) Staff training	W #8 W #9	_____	_____
(7) Developing and ordering materials and equipment	W #8 W #9	_____	_____
(8) Staff recruiting and hiring	W #8 W #9	_____	_____

488

Type of Information	Available Instruments	Should it Be Done? (Yes or No)	If Yes When? (List Date)
b. Teachers	W #8		
(1) Planning instruction	W # 8		
(2) Implementing Instruction	W # 8		
(3) Non-instructional responsibilities	W # 8		
c. Aides	W # 8		
d. Other staff	W # 8		
(1) Instructional coord-inator	W # 8		
(2) Community coordinator	W # 8		
(3) Evaluator	W # 8		
3. Staff Development (Describe)			
a. Needs assessment	W # 8		
b. Structure of training	W # 8		
(1) Pre-service	W. # 8		
(2) In-service	W # 8		
c. Characteristics of Training			
(1) Appropriateness for staff of differing levels of knowledge and experience	W # 8		
(2) Practicality	W # 8		
(3) Coordination with degree programs	W # 8		
(4) Integration with other	W # 8		
d. Audiences Trained			
(1) Project staff included	W # 8		
(2) Inclusion of non-project staff	W # 8		

Type of Information	Available Instruments	Should it Be Done?. (Yes or No)	If Yes When? (List Date)
4. Parents and Community	W # 8 W #.15	_____	_____
a. Parent involvement in school affairs	W # 8 W # 15	_____	_____
b. Community input in program planning, e.g., through advisory group	W # 8 W # 15	_____	_____
c. Community support for project	W # 8 W # 15	_____	_____
d. Parent education	W # 8 W # 15	_____	_____
e. Parent conferences/counseling	W # 8 W # 15	_____	_____
5. Communication	W # 8	_____	_____
a. Staff relations	W # 8	_____	_____
b. Relations with nonproject staff	W # 8 W # 8	_____	_____
(1) District administrators	W # 8	_____	_____
(2) Principals	W # 8	_____	_____
(3) Nonproject teachers	W # 8	_____	_____
(4) School board	W # 8	_____	_____
6. Dissemination of project information	W # 8 W # 15	_____	_____
a. School personnel	W # 8	_____	_____
b. Parents and community	W # 8 W # 15	_____	_____

490