DOCUMENT RESUME

ED 221 053 FL 013 188

AUTHOR Rodriguez, Victor, Ed.
TITLE Language Proficiency Assessment: What Does That Mean? A Report of the NABE Pre-Conference Workshop (Los Alamitos, California, April 19, 1980). Professional Papers.
INSTITUTION National Center for Bilingual Research, Los Alamitos, Calif.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
PUB DATE Apr 80
NOTE 112p.; Workshop conducted pursuant to cooperative agreement 00-CA-80-0001.

EDRS PRICE MF01/PC05 Plus Postage.
DESCRIPTORS *Bilingual Education; Bilingual Students; *English (Second Language); Korean; *Language Proficiency; *Language Tests; Oral Language; Reading Skills; Spanish; Testing Problems; Writing Skills

ABSTRACT

Materials are presented from a workshop designed to provide an opportunity for bilingual education researchers and practitioners to share knowledge, experiences, and concerns related to assessing language proficiency. The sessions included: (1) "Formal and Informal Evaluation of Oral English Language Skills," by William Russell; (2) "Assessing English Literacy Skills: Writing," by Ann Humes; (3) "Reading Assessment and the Bilingual Teacher," by Laila Fiege-Kollman; (4) "Some Considerations in Constructing and Administering Language Proficiency Tests," by Bonita Ford; and (5) "Adaptation of English Proficiency Instruments for Korean," by Kenneth K. Kim. The workshop evaluation form, along with the participants evaluations are appended. (EKN)

LANGUAGE PROFICIENCY ASSESSMENT: WHAT DOES THAT MEAN? A REPORT OF THE NABE PRE-CONFERENCE WORKSHOP

Edited by: Víctor Rodríguez

April, 1980

R-1

[redacted]

national center for bilingual research, 4665 lampson avenue, los alamitos, CA 90720, telephone (213) 598-0481

The material in this publication was prepared under Cooperative Agreement 00-CA-80-0001 from the National Institute of Education, Department of Education. Its contents do not necessarily reflect the views of the National Institute of Education or of any other agency of the United States of Education or of any other agency of the United States government, and no official endorsement should be inferred

WORKSHOP REPORT

Language Proficiency Assessment:

What Does That Mean?

April 19, 1980

FUNDING INFORMATION

PROJECT TITLE:	National Center for Bilingual Research (NCBR)
GRANT NUMBER:	00-CA-80-0001
SOURCE OF GRANT:	National Institute of Education Washington, DC
CONTRACTOR:	Southwest Regional Laboratory for Educational Research & Development (SWRL) Los Alamitos, CA
PROJECT STAFF:	Cándido Antonio de León (Director), Víctor Rodríguez, Robert Berdan, Maryellen García, Kenneth Kim, Bonita Ford, Angelina Veyna, David Thrift
DISCLAIMER:	The Workshop presented or reported herein was conducted pursuant to a Cooperative Agreement from the National Institute of Education, Department of Education. However, the opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by the National Institute of Education should be inferred.

National Center for Bilingual Research
4665 Lampson Avenue
Los Alamitos, CA 90720

FOREWORD

The National Center for Bilingual Research was created by the National Institute of Education (NIE) and Southwest Regional Laboratory for Educational Research and Development (SWRL) in November, 1979. The Center, which represents the first federally funded major research unit in bilingualism and bilingual schooling in the United States, conducts a range of activities, including research, dissemination, and collaborative activities in research, training, and technical assistance.

As part of its activities, the Center conducted a pre-conference workshop in conjunction with the Ninth Annual Conference of the National Association for Bilingual Education. The workshop, "Language Proficiency Assessment: What Does That Mean?", was held at SWRL on April 19, 1980. This report was developed in an effort to record some of the experiences and knowledge that were shared during the half-day workshop.

The National Center for Bilingual Research anticipates sponsoring other workshops and conferences to provide more opportunities for researchers and practitioners to communicate and share their knowledge, experiences, and concerns.

Víctor E. Rodríguez
Workshop Coordinator

INTRODUCTION

"Language Proficiency Assessment: What Does That Mean?" was the title of a workshop held at SWRL Educational Research and Development in Los Alamitos, CA, on April 19, 1980. The workshop was co-sponsored by SWRL and by the National Center for Bilingual Research in conjunction with the pre-conference activities of the Ninth Annual International Bilingual Bicultural Education Conference of the National Association for Bilingual Education.

The workshop was designed to provide an opportunity for educational researchers and practitioners to share knowledge, experiences, and concerns related to one particular area of bilingual education: what language proficiency is and how one measures language proficiency. A primary focus of the workshop was to be the assessment of non-English languages. As planning progressed, however, it became apparent that, although bilingual educators are primarily concerned with assessing oral language proficiency, the implications of language proficiency in assessing writing and reading skills also had to be addressed.

In coordinating the conference presentations, the National Center for Bilingual Research was fortunate in being able to draw upon its staff and the staff resources of SWRL, which over the past 13 years has distinguished itself in conducting research in language skills assessment and in developing curriculum materials for bilingual students.

Members of the SWRL staff were asked to provide a general overview of formal and informal assessment of language proficiency, and to discuss the assessment of reading and writing skills. The staff of the National Center for Bilingual Research addressed the issue of assessing non-English languages, focusing primarily on Spanish and on Korean.

William Russell, SWRL Member of the Professional Staff, prepared a presentation on "Formal and Informal Assessment of Language Proficiency," based on his work with the SWRL project to develop a Student Placement System for Bilingual Programs. Laila Fiege-Kollmann, SWRL Member of the Professional Staff, spoke on "Reading Assessment and the Bilingual Teacher" and drew on her experience with the development of the reading component of the Survey of Essential Skills Project for the Los Angeles Unified School District. Ann Humes' topic was, "Assessing English Literacy Skills: Writing." Ms. Humes, SWRL Member of the Professional Staff, has extensive experience with the assessment of literacy skills as a result of her work in the SWRL Proficiency Verification System, the LAUSD Survey of Essential Skills Project, and the Project to Develop a Student Placement System for Bilingual Programs.

Bonita Ford, Member of the Center's Research Staff, worked on the development of SWRL's Resources for Assessing Language Proficiency in Spanish (RALPS) and on the Diagnostic Assessment System (DAS) for assessing both English and Spanish language proficiencies prior to joining the staff of the National Center for Bilingual Research. Her presentation was on "Some Considerations in Constructing and Administering Language Proficiency Tests." Kenneth Kim, Member of the Center's Research Staff, chose to discuss the problems that one might encounter when attempting to adapt existing language assessment instruments for use with Asian and non-Indoeuropean languages. His consulting experience with Los Angeles Unified School District, ABC Unified School District, and the California State Department of Education, has convinced him that this practice is more often based on expediency than on sound linguistic principles and can cause problems with the interpretation of scores and, subsequently, with the classification of students.

Seventeen participants attended the half-day workshop. Nine states were represented and participants included Bilingual Program Coordinators, Resource Aides, Teachers, and Teacher Trainers. The agenda for the workshop is reproduced on the following page. The report of the workshop includes papers which were developed subsequent to the workshop presentations and a workshop evaluation based on questionnaires completed by the participants.

NCBR/SWRL
NABE Workshop
April 19, 1980

AGENDA

NABE Pre-Conference Workshop
"Language Proficiency Assessment--What Does That Mean?"

9:30 - 9:45	Coffee	
9:45 - 10:00	Introduction:	Dr. Víctor Rodríguez Associate Director National Center for Bilingual Research
10:00 - 10:20	Session #1:	Assessing Oral English Language Use. Mr. William Russell Member of the Professional Staff SWRL Educational Research & Development *"Formal and Informal Evaluation of Oral English Language Skills"*
10:30 - 11:30	Session #2:	Assessing English Literacy Skills. Ms. Ann Humes Member of the Professional Staff SWRL Educational Research & Development *"Assessing English Writing Skills"* Dr. Laiia Fiege-Kollmann Member of the Professional Staff SWRL Educational Research & Development *"Assessing English Reading Skills"*
11:30 - 11:45	BREAK	
11:45 - 12:45	Session #3.	Assessing Non-English Language Skills. Dr. Bonita Ford Associate Member of the Research Staff National Center for Bilingual Research *"Assessing Spanish Language Proficiency"* Dr. Kenneth Kong-On Kim Member of the Research Staff National Center for Bilingual Research *"Adapting Language Assessment Instruments for Testing Asian Students"*
12:45	ADJOURN--Bus returns to Convention Center	

NCBR/SWRL
NABE Workshop
April 19, 1980

Introduction
NABE Pre-Conference Workshop

I. Objectives

A. Learn what language proficiency assessment is.

1. Learn approaches which are used to assess language proficiency.
2. Learn why approaches work or not.
3. Learn about assessing non-English language proficiency.

B. Plan to use the ideas.

1. Apply them in your situation
2. Seek materials and people to learn more about approaches to assessing language proficiency.
3. Seek additional training in the use of approaches mentioned.

C. Be willing and able to share with others what was learned.

II. Reasons

A. Share the ideas available from various sources which are usually not shared.

B. Speed up the process of dissemination of research information and ideas.

C. Go beyond talking about ideas and being aware of approaches to trying to apply them.

III. Roles

A. Be inquisitive and willing to share your experience.

B. Take responsibility to make your involvement worthwhile. Ask yourself: How can I use this information?

FORMAL AND INFORMAL EVALUATION OF ORAL ENGLISH LANGUAGE SKILLS

NABE Pre-Conference Workshop

William Russell

April 19, 1980

FORMAL AND INFORMAL EVALUATION OF ORAL ENGLISH LANGUAGE SKILLS

NABE Workshop Presentation

William Russell

We have a wide range of bilingual educators present. A number of you identify yourselves as school level coordinators and resource teachers, some are classroom teachers, a few are paraprofessionals and at least one is a district level coordinator. I will attempt to address this presentation to your various needs, especially to the needs of those persons who work directly with students.

The title of this workshop is "Language Assessment: What Does That Mean"? Since language assessment measures language proficiency, I will first address the question, "Language Proficiency: What Does That Mean"?, and then I will suggest some ways that language proficiency can be assessed. I will distinguish between formal and informal assessment of language proficiency and explain why I believe both are important. Throughout this presentation I will emphasize oral language assessment as contrasted with reading and writing.

Here at SWRL, we recently completed a large scale analysis of English proficiency to be used in developing an assessment system for limited English proficiency (LEP) pupils. Figure 1 is a diagram of the English proficiency analysis produced by that project.

The specific purpose of this analysis was to identify the English skills minimally necessary for functioning in an all-English language classroom. It was designed to be one of the tools that school districts and others could use for selecting or constructing tests to be used in the placement of LEP pupils. The analysis is included in a set of documents called Resources for Developing a Student Placement System for Bilingual Programs, which will be mailed to all school districts in the United States that have Title VII bilingual programs.

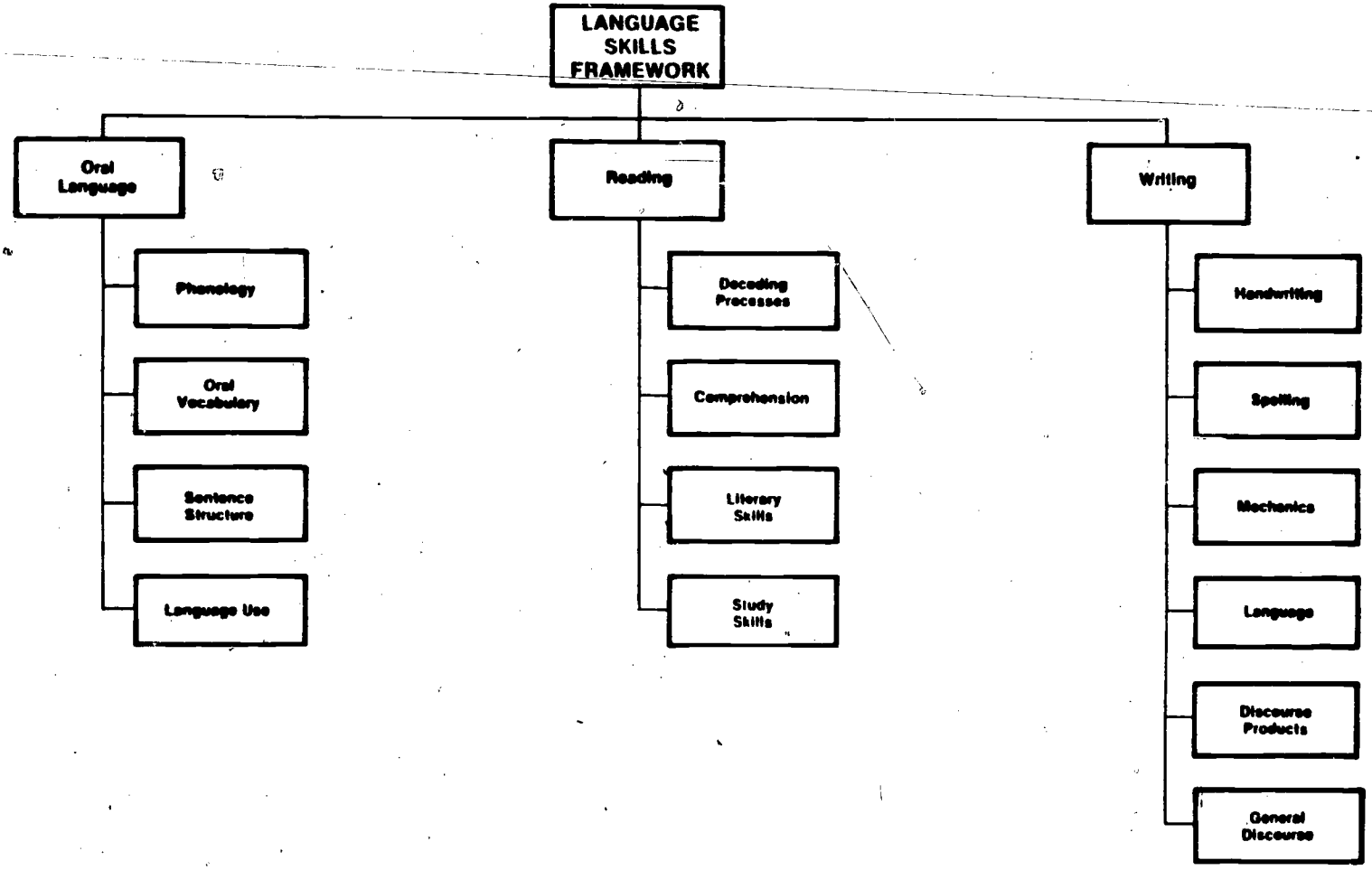


Figure 1. Organization of the SWRL English Proficiency Analysis by Component, Area and Subarea.

You see in Figure 1 that the first cut we made in the SWRL English proficiency analysis was to divide it into Reading, Writing and Oral components of English proficiency. These in turn were divided into areas and these areas further divided into subareas. Separate tree-graphs for each of the three components are shown in Figures 2, 3, and 4.

The principal sources for the Reading and Writing components were actual textbook series used in classrooms across the country. Page by page analyses of these texts, together with other information about classroom practices, revealed the skills that children at particular levels are expected to know. But oral proficiency is not determined by formal instruction as is reading and writing. So, in place of instructional materials, we consulted studies of actual classroom language behavior, psycholinguistic studies of school age native English development, and a SWRL study of vocabulary use.

Now, let's look again at Figure 4, the diagram of the Oral component. At the lowest level of this figure, oral language proficiency is divided into twelve subareas. This analysis was derived from information extracted from the sources I just mentioned.

Let's pursue the oral language proficiency analysis a step further than Figure 4. Take, for example, the Complex Sentences subarea. It was further divided into Relative Clauses and Verb Complements. Below this level of analysis, every box on the oral language diagram branches into actual pupil behaviors. These last two levels of analysis are not shown on Figure 4, but they are listed in the table shown as Figure 5.

In reality there are an indefinite number of behaviors for every subdivision of the analysis, but, for assessment purposes, we are interested in only those few behaviors that are practically testable. Figure 5 shows the behaviors (skills) listed for the Complex Sentences subarea of the Sentence Structure area of the Oral Language component. Figure 5 also shows the grade levels to which each of these skills is assigned.

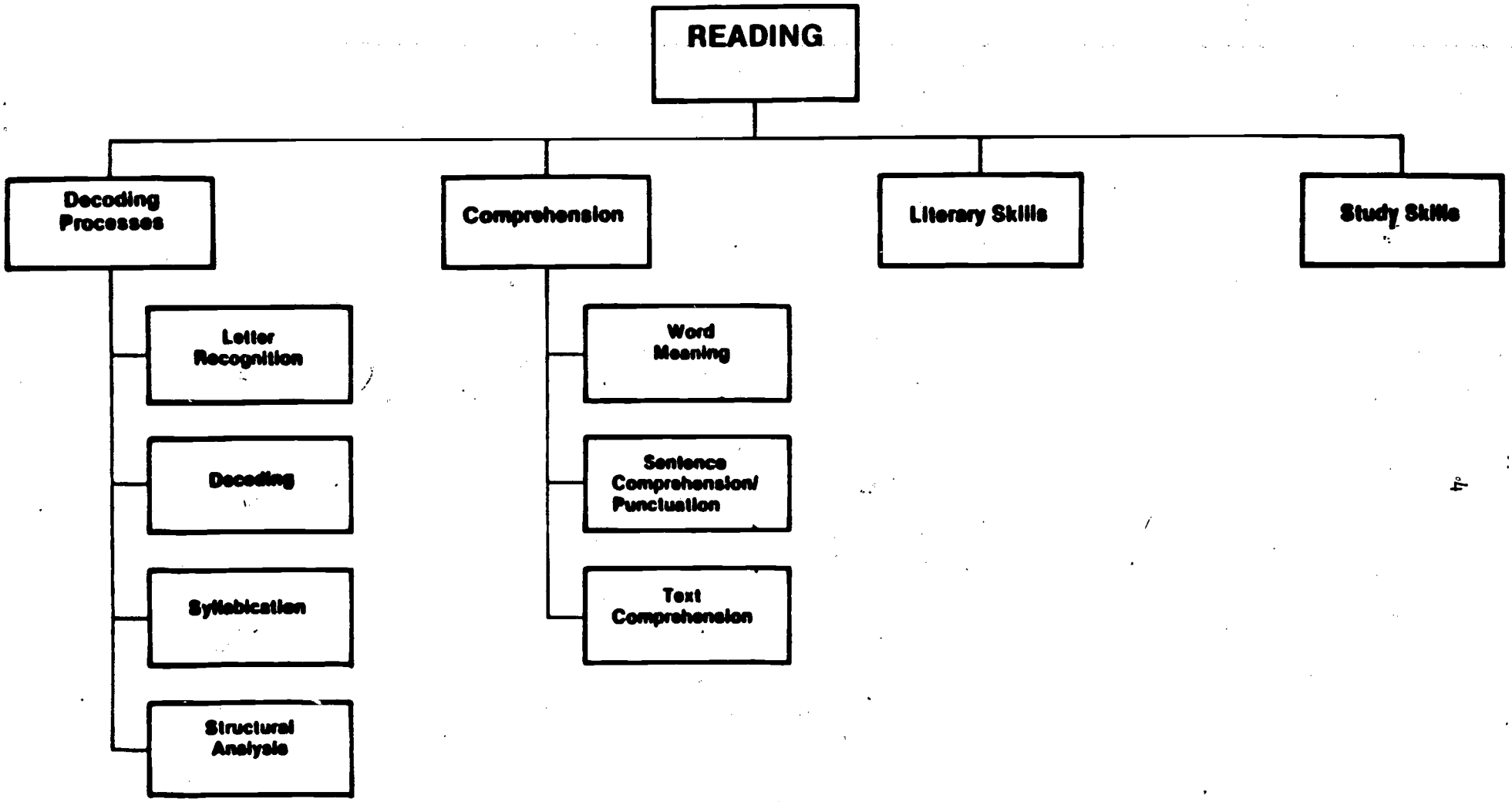


Figure 2. Proficiency Analysis for reading.

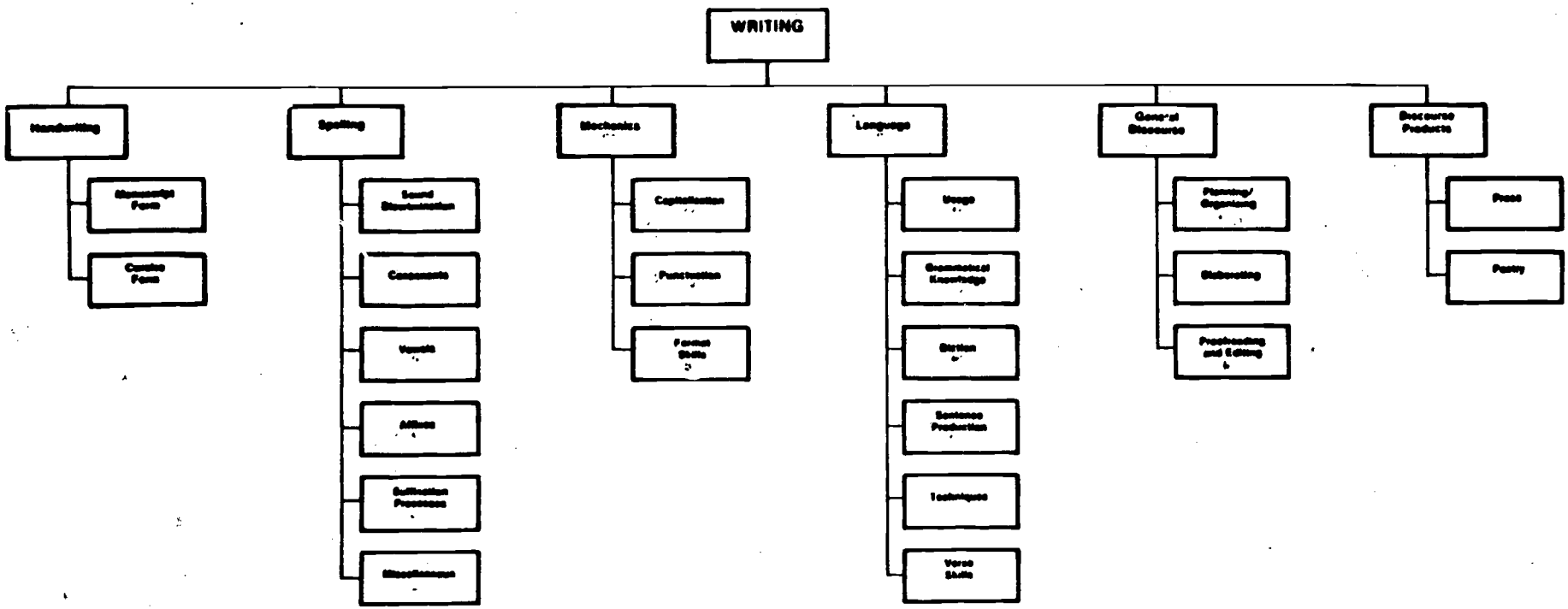


Figure 3. Proficiency Analysis for writing.

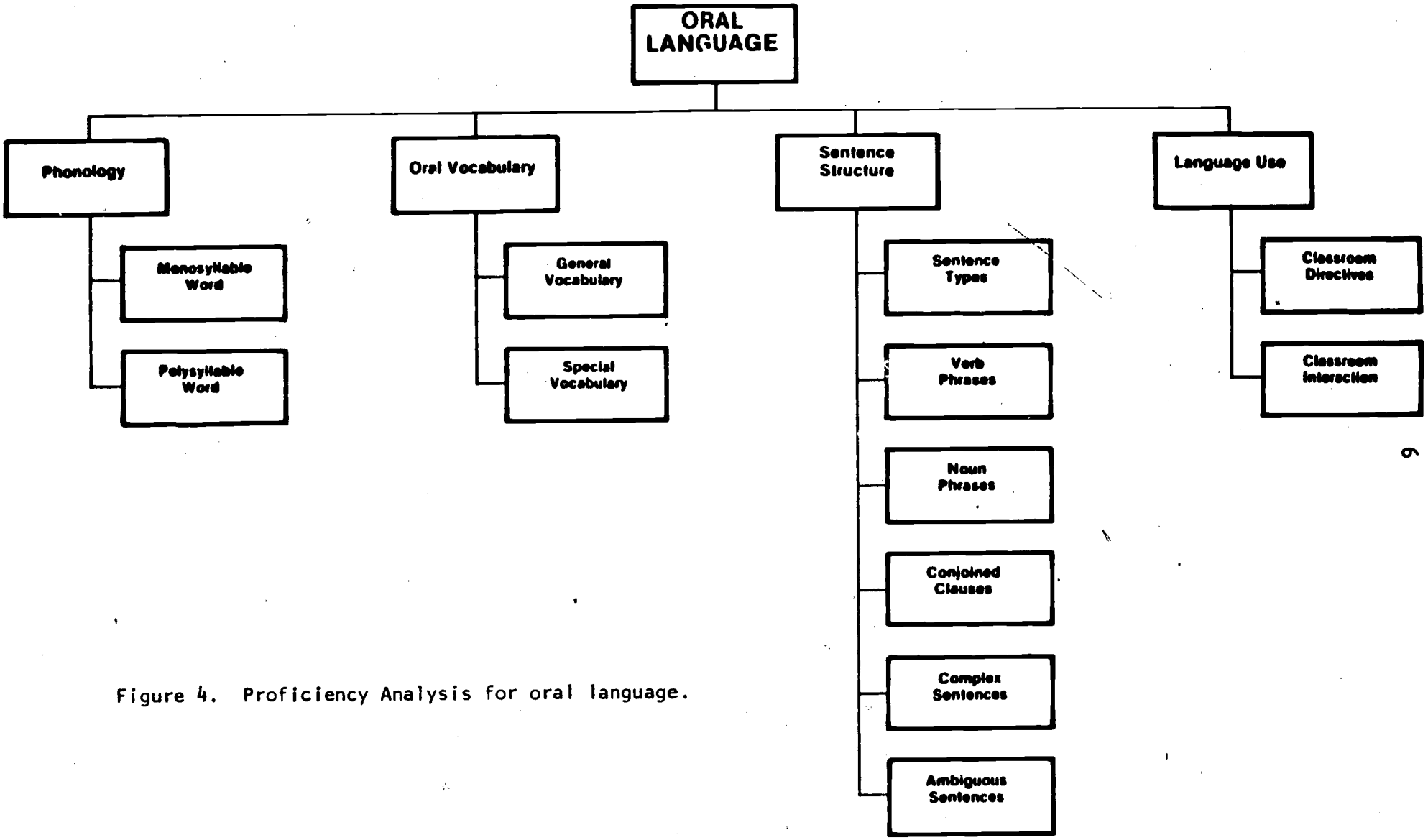


Figure 4. Proficiency Analysis for oral language.

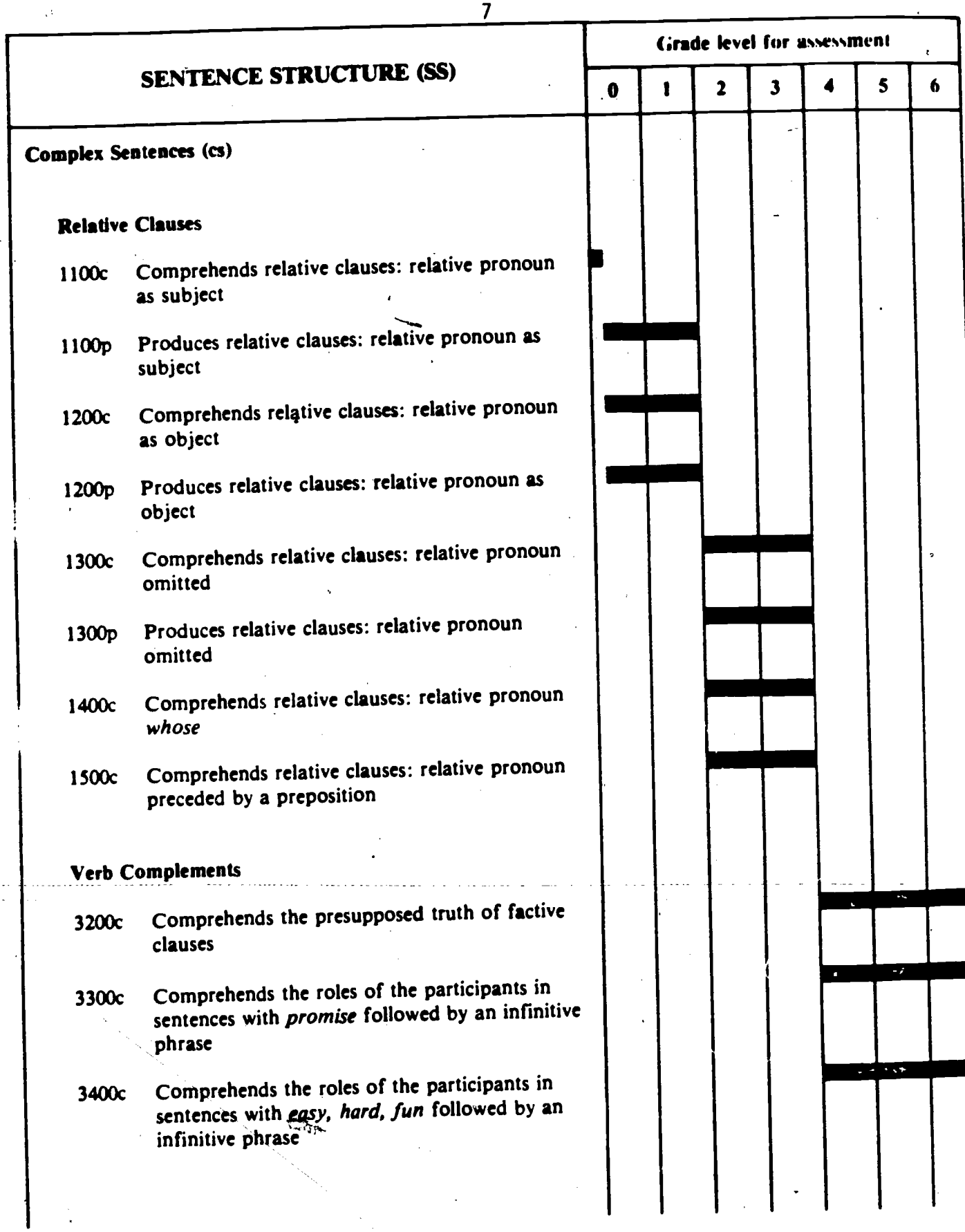


Figure 5. A Portion of the Oral Language Analysis at the Skills Level.

Each of the skills in the SWRL analysis is extensively specified. For example, let's look at just one of the skills from Figure 5, skill number 1200p, "Produces relative clauses: relative pronoun as object." The full specification of this skill is shown in Figure 6. The best way to understand a skill in this analysis is to look at actual examples of the behavior in question. Such examples are given in the full specification shown here. Also, advice is given here about how to assess this skill.

Up to this point, I have briefly shown how the SWRL analysis of English proficiency progresses from a first cut into Reading, Writing and Oral Language until the actual skills level is reached. We have been looking at a small sample of the analysis of English proficiency which was done at SWRL for the Resources for Developing a Student Placement System for Bilingual Programs. These Resources provide a substantial basis for the selection or construction of tests to assess the English proficiency of LEP pupils. Similar analyses could be done for other languages, but I do not know of any. Because of the need for such proficiency frameworks for other languages, there may be a temptation to translate some of the assessment items that appear in the Resources, but this would be a serious mistake. You have seen the many levels of analysis that were used to organize the skills of the Resources. A similar analytic framework for another language would be quite different, and the skills that it would subsume would only occasionally be similar to those for English.

As an example of an assessment item in another language, let me show you two relative clause skills in Spanish together with tasks for assessing proficiency in performing these two skills. These materials are shown in Figure 7.

For the Spanish examples, I purposely chose skills that are very similar to the English skills we just looked at, so that the Spanish examples would be more understandable, but keep in mind the caveat regarding translating skills. Even these very similar skills relate

00 SScs 1200p (continued)

To ease processing requirements at this level, the relative clause is located at the end of the sentence; i.e., modifying either a direct object or a predicate noun; e.g.:

Direct object: Mary caught the ball that Jim threw.
Predicate noun: This is the ball that Jim threw.

In addition, at this level, the relative clause is restrictive (not set off by pauses).

Relative pronouns that can be used with this type of relative clause are that, which, who/whom. Not all are recommended, however, for the reasons enumerated below.

a) That: That is the most commonly used pronoun in this type of relative clause. That can be used with either human or non-human nouns although use with non-human nouns is more common.

I saw the ball that John hit.
I saw the boy that the dog bit.

b) Which: Which is used only with non-human nouns in the modified noun phrase. Some authorities do not approve of its use in restrictive clauses.

I saw the ball which John hit.

c) Who/whom: This type of relative clause takes the whom form in writing and very formal speech; e.g., I saw the boy whom the dog bit. However, who is commonly used, especially in speech. Students should not be expected to discriminate usage of who/whom.

For production, students may use any of these pronouns. Although which should be distinguished from who/whom, students should not be required to use whom rather than who.

Assessment

Sample Items

1. A. [picture: girl, wrench in hand, standing next to bicycle she has just fixed]

Janet fixed the bicycle.

Figure 6 (page 2)

00 SScs 1200p (continued)

B. [picture: boy, wrench in hand, standing next to bicycle he has just fixed]

Tom fixed the bicycle.

(Point to picture.) Which bicycle is this?

2. A. [picture: man standing next to tree he has chopped down]

The man chopped down the tree.

B. [picture: woman standing next to tree she chopped down]

The woman chopped down the tree.

(Point to picture.) Which tree is this?

3. A. [picture: monster standing next to a chair with a large bite taken out of the chair]

The monster ate this chair.

B. [picture: a monster sitting on a chair that is crushed by its weight]

The monster sat on this chair.

(Point to picture.) Which chair is this?

Item Description

The examiner displays two pictures and, pointing to each picture, says the corresponding descriptive sentence. Then the examiner points to one of the pictures and asks, "Which _____ is this?" The student responds orally with an answer containing a relative clause. (If desired, the examiner may then point to and query the other picture for an additional student response.)

Each pair of pictures depicts the same object acted upon by two different people or animals. The two objects are distinguishable solely on the basis of who or what is acting on them.

The sentences describing the pictures are in the form desired for the relative clause to be elicited; e.g., "The monster ate this chair," is equivalent to "that the monster ate."

Figure 6 (page 3)

00 SScs 1200p (continued)

There are a number of possible responses for each item, all containing relative clauses; for example, the following (plus other variations) should be considered correct for item 1:

This is the bicycle that Janet fixed.
This is the one that Janet fixed.
The bicycle that Janet fixed.
The one that Janet fixed.

Moreover, the student need not describe the picture exactly as the examiner does, so long as an appropriate relative clause is used; e.g., the one that Janet has. Any appropriate relative pronoun (i.e., which, who; see Skill Description) may be used instead of that.

Comment: A problem with assessing this skill is that the student can give an accurate response without using a relative clause; e.g., the one on the bicycle, the one with the bicycle. Or the student may use a relative clause in which the relative pronoun is not the object; e.g., the one that was eaten by the monster, the one that has a hungry monster. Moreover, the student may use the appropriate relative clause without a relative pronoun; e.g., the one Janet fixed. (Use of this more complex structure is assessed by Skill 02 SScs 1300p.) If the student seems to understand the pictures and the questions, but does not respond with the appropriate relative clause, two other approaches are possible: modeling and imitation.

Modeling. For modeling, the examiner supplies the appropriate response for a few items and then asks the student to respond to others; e.g.:

A. Janet fixed the bicycle.
B. Tom fixed the bicycle.
(Point to picture A.) Which bicycle is this?
This is the bicycle that Janet fixed.

Imitation. If modeling does not produce the desired response, the examiner may have the student imitate his/her use of the correct structure; e.g.:

I'm going to tell you about these pictures.
When I'm done, you will say what I say.
A. This is the bicycle that Janet fixed.
B. This is the bicycle that Tom fixed.
(Point to picture A.) What is this picture?

Figure 6 (page 4)

to their respective language systems quite differently and carry somewhat different functions in each language.

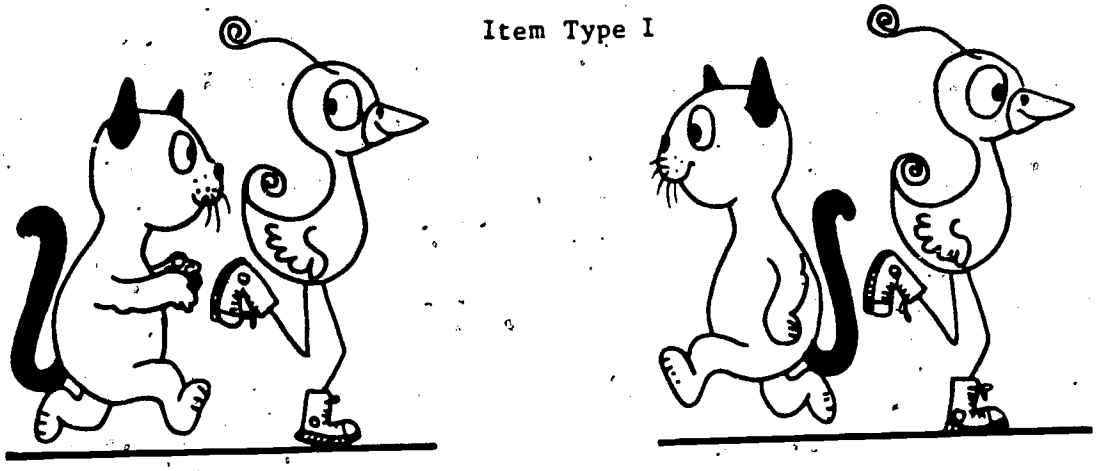
Notice in Figure 7 that the pupil does not respond exactly as expected, especially to the second item. The last paragraph of Figure 7 makes the poing that it is the successful production of the relative clause form that is being elicited, so the reponse counts as correct even though vocabulary, style, dialect or other features of the response may not be expected shcool usage. One must be clear about what is being tested and then evaluate responses on that basis. An analytical framework like the one used in the Resources is helpful in making clear exactly what category of behavior a particular skill belongs to, and therefore exactly what is being tested by the assessment items that test this skill.

So far, we have been looking at a basis for making formal assessments of language proficiency. But, as a person experienced in working with bilingual pupils, you are in a position to make educated informal judgements as well as administer formal tests. Such informal judgements are most sensitive to "global" aspects of successful communication as distinguised from the discrete aspects that were sorted out by the SWRL analysis. In the present state of the art of language assessment, I belive that you can do the best job of placing LEP pupils by using a balance of experienced intuition and formal language assessment to capture both global and discrete aspects of communication behavior.

Let me suggest in conclusion two areas of caution when informally judging the language proficiencies of a LEP pupil. With these cautions in mind, your experience with LEP pupils should produce excellent intuitions about the "global" language proficiencies of these pupils.

Interview sessions consisted of the presentation of a series of pictures to the pupils. Each picture contained two frames, with two animal characters in each: a cat and a bird or a cat and a dog. After describing the situations, the interviewer asked the child to distinguish one cat from the other by the question: "¿Cuál gato es éste?" Interview items were divided evenly into two basic situation types according to the subject-object relation of the characters. An example of each of these types is given below, along with the elicitation procedure and typical pupil responses.

Item Type I



(a) (b)

Interviewer: ___ (Points to the cat in frame (a))
"Este gato está persiguiendo a un pájaro."
___ (Points to the cat in frame (b))
"Pero este gato no está persiguiendo al pájaro."
___ (Points to one of the cats as specified in the written script, in this example the cat in frame (a))
"¿Cuál gato es éste?"

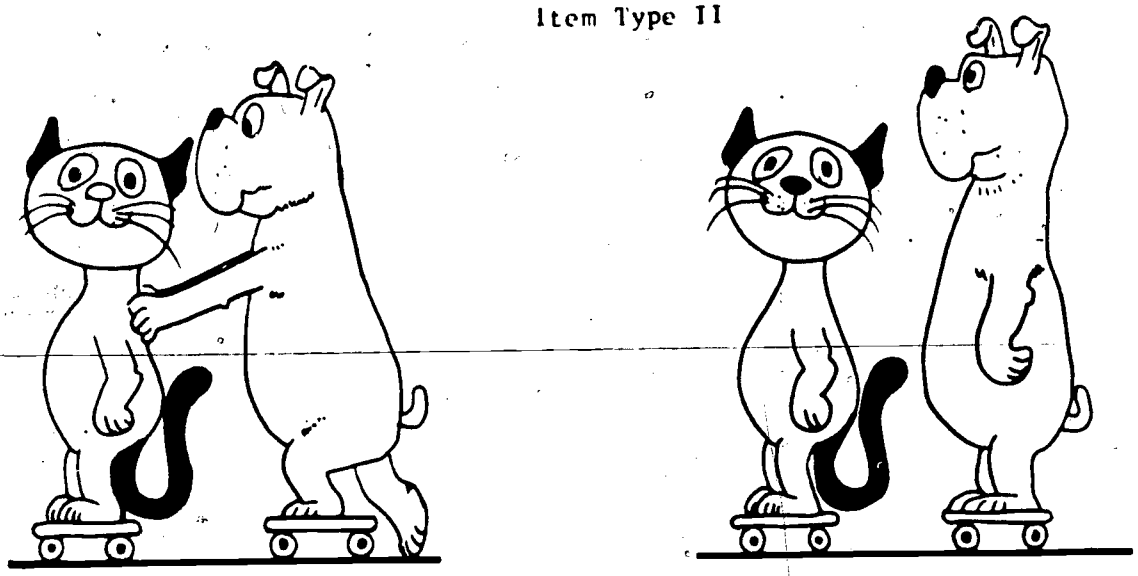
Pupil: "Este es el gato que 'stá persiguiendo el pájaro."

Interviewer: "Stá bien. Este es el gato que está persiguiendo al pájaro. Dímelo otra vez."

Pupil: "Este es el gato que 'stá persiguiendo al pájaro."

Figure 7 Two examples of assessment items (production of rel. clause types)
FROM: RUSSELL, WM. AND DAVID SNOW, SWRL TN-2-76-08, 1976

Item Type II



(a) (b)

Interviewer: ___ (Points to the cat in frame (a))
"Un perro está empujando a este gato."
___ (Points to the cat in frame (b))
"El perro no está empujando a este gato."
___ (Points again to one of the cats, say, the one in frame (b))
"¿Cuál gato es éste?"

Pupil: "Este es el gato qu'el perro no lo anda'mpujando."

Interviewer: "Muy bien. Dímelo otra vez: Este es el gato que el perro no está empujando."

Pupil: Este es el gato qu'el perro no lo anda'mpujando."

Each interview was preceded by a practice session, the purpose of which was to familiarize the pupil with the situations, characters and vocabulary items, and to teach and practice the task itself. Throughout, instructions were repeated, or the task reviewed, if hesitation on the pupil's part indicated that additional support would be helpful or if his response failed to give the interviewer confidence that the task had been understood. Our purpose was to make the task as clear as possible, both in order to enable the pupil to experience success in performing it, and to encourage an appropriate response with a relative clause, the form of such responses being what we intended to observe.

Figure 7 (continued)

CAUTIONARY AREA 1: DIALECT

It is easy to confuse the use of non-standard dialect forms with limited language proficiency. Suppose a child says in English,

"She ain't got none,"

or in Spanish,

"Tú vites como el troque andaba asina."

Because these utterances contain non-standard dialect forms, they are inappropriate in certain social circumstances. But this inappropriateness is not the primary concern of bilingual programs. A bilingual program must first consider the pupil's proficiency in a given language without regard to dialect. It is possible to be quite proficient in a language without being proficient in a particular dialect. The confusion between dialect appropriateness and language proficiency is encouraged by the common perception of non-standard forms as "errors" or "bad grammar." These forms may properly be viewed as errors of social appropriateness in particular situations, but they do not reflect a lack of general proficiency in the language.

CAUTIONARY AREA 2: INDIVIDUAL VARIATION

Another caution in judging proficiency has to do with individual variation in native proficiency. This situation is particularly difficult for the bilingual educator for reasons I will attempt to set forth. I bring up this difficult topic here because of its serious implications for the welfare of bilingual children.

The main points of this cautionary area are:

1. There is a basic aspect of language proficiency that does not vary much across mature native or near-native speakers of the language. It follows that all normal speakers have an intact native language. The observation that some bilingual speakers find themselves "between languages" is probably an illusion for reasons I will give.

2. Basic proficiency in a second language does, of course, vary from speaker to speaker up to the stage of acquisition where the second language is near-native, then basic proficiency behaves as in native speakers.

3. Basic proficiency may improve with instruction in a second language, and so it is taught in bilingual programs. In mainstream classrooms, a different kind of proficiency (which I call "virtuosity") is the subject of instruction.

4. For assessment, bilingual educators must understand which kind of proficiency they are evaluating.

The way native language proficiency varies from user to user is best understood by positing two kinds of native proficiency, "basic proficiency" and "virtuosity." Basic proficiency includes control over the sound system and syntactic patterns of the language, while virtuosity includes such things as skillful use of the language to convince and be convinced, explain and understand, teach and learn, entertain and be entertained, etc. "Basic proficiency" and "virtuosity" are my terms, not terms from the literature. In fact, this area of research is sufficiently unexplored at this time that the following comments should be taken merely as convictions based on years of research experience rather than as established facts.

Basic proficiency does not vary significantly from native speaker to native speaker at any given stage of language development. That is to say, all normal children pass through similar stages of basic proficiency development in their native language and eventually reach an adult level that is approximately the same as that of other speakers.

If basic proficiency is a characteristic of all native speakers, then it follows that al' normal bilingual children have full basic proficiency in at least one language. It is often contended that some bilinguals are "between languages" and somehow do not have a native language in the way that monolingual speakers do. I believe that this conclusion results from a confusion of basic proficiency with virtuosity. For instance, vocabulary development beyond the words needed for very basic communication is a matter of individual virtuosity and is not necessarily an indicator of basic proficiency.

It is important when assessing the language proficiencies of bilingual children to identify the languages in which they have basic proficiency, regardless of the virtuosity that they might have in these languages, because other proficiencies will be learned differently by individual children, depending on their potential for skillful use of language. We must first identify the languages in which the basis for this skillful use of language has been acquired. Then we can judge what more can be expected of a pupil according to his or her potential for virtuosity.

For assessment, one important implication of the distinction between basic proficiency and virtuosity is that there will be one of more basic language proficiencies for each child, and these must be identified.

For instruction, the virtuosity versus basic proficiency distinction has a further implication. Basic proficiency is typically acquired without instruction. Instruction may facilitate its acquisition in a second language by focusing experience on the new language. Beyond this, however, instruction will focus on virtuosity.

As a bilingual educator, you are much more concerned with basic proficiency than other educators. And it is more important for you to be aware of the distinction between basic proficiency and virtuosity. For instance, you may place a child in a mainstream classroom because his or her basic proficiency in English is native-like even though the child's evidence of virtuosity in English is not great. A reason for doing this would be that you judge that this child's particular potential for virtuosity has been realized at this stage of development. On the other hand, you may retain another child who has the same basic English proficiency in a bilingual classroom so as to allow the child's great potential for skillful use of English to develop before the pupil is placed in competition with native speakers. The latter decision would have to assume that the bilingual classroom in question offers rich experiences in English accompanied by opportunities to succeed academically in a native language.

It is even more important to your pupils than to pupils of the mainstream for language assessment to be skillfully performed. This is just one of the many ways in which your job is more difficult.

ASSESSING ENGLISH LITERACY SKILLS: WRITING

ANN HUMES
SWRL Educational Research and Development
Los Alamitos, California

NABE
Pre-Conference Workshop
Los Alamitos, California
April 19, 1980

ASSESSING ENGLISH LITERACY SKILLS: WRITING

A goal of transitional bilingual programs is to provide students with sufficient proficiency in English to function in the monolingual classroom. This paper discusses the steps involved in describing the English writing skills required in grades 1-6, the assessment of those skills, and some of the problems that were encountered in accomplishing the task.

SKILLS IDENTIFICATION

The first task involved identifying the critical writing skills, and the first problems were encountered when the literature was reviewed for help:

1. Although there currently is an intense interest in writing, the literature does not now provide adequate information on the specificity or range of critical writing skills. The lists of skills suggested by authorities are usually too brief and the skills themselves are too general, as these are, for example:

   > Uses accepted punctuation and capitalization.
   >
   > Uses accepted form and appropriate language in varying types of written communication (Petty, Petty, Newman, and Skeen, in Squire, 1977, pp. 89-90).

   Even more general is the frequently cited, "Writes effectively."

2. Sometimes a listed "skill" really consists of a skill plus a task or an item description. The following skill statement is typical:

   > Writes a description of a picture of a painting.

   "Writes a description" is the skill; the rest of the statement refers to the specific object to be described. However, the

same writing skill is often relisted as a separate skill that is differentiated only by that object to be described--for example, "Writes a description of a person."

3. Skill identification is also hampered by the literature's failure to designate the discourse-type subskills of a skill statement such as "Writes a description," that is, subskills like "Uses sensory terms," "Uses spatial ordering."
4. Finally, some skills proposed in the literature are more than too general; they are vague. This skill sampled from a state competency document (Georgia Department of Education, 1968) is typical:

> Understands that writing is a tool of communication (p. 77).

The Collection of Skills

Despite these problems, many skills were identified during the review of scholarly recommendations and competency lists. Each broad skill, such as "Writes a description," was analyzed to identify its subskills. The collection of skills and subskills was then compared with skills derived from the SWRL data available on language-arts textbook series, and any duplicated skills were deleted. To collect the data, page-by-page task analysis is performed on textbooks to determine the content of the instruction and the manner in which the content is practiced. Computer processing of this data produces lists of content by grade level. Consequently, a large data base describes actual textbook (and therefore classroom) instruction in detail. Page 2 of the handout is a copy of an actual computer printout of the textbook data.

The Critical Skills

The comprehensive list was then screened for essential skills according to three criteria: appearance on state or local competency lists, textbook emphasis, and the nature of a skill as either a prerequisite of skills at a higher grade level or as a subskill of another skill. Skills that met these criteria were retained on the list with the exception of some that were covered by the reading component (such as dictionary skills). The list was consequently reduced to approximately 180 different skills.

GRADE LEVELS

The next step in the project involved assigning appropriate grade levels to the collected skills. Again the literature posed problems:

1. Some language arts authorities do not assign grade levels to skills because they feel that sequencing is the local school district's prerogative.
2. Some authorities designate cuts in sequencing in blocks such as K-3 rather than in grades.
3. Other sources list skills for more than one grade level without noting the criteria that differentiate performance among grade levels. These grade 3, 4, and 5 skills are such an example:

   Grade 3: Writes short stories and/or poetry (p. 54).

   Grade 4: Writes stories and poetry (p. 68).

   Grade 5: Writes stories, plays, and poetry (p. 92).

   (Los Angeles City Schools, 1964)

4. Still other sources sequence skills objectives without presenting

any evidence to validate their ordering, and frequently their sequencing conflicts with textbook sequencing.

The chart on page 3 of the handout displays the grade-level variation that occurs between authorities' designations and textbook presentation. The chart is based on a study by Golub (1971), who analyzed differences between grade-level designations found in language arts books for teachers and the grade levels of instruction in students' textbooks. The chart shows the grade-level consensus found in four books written by eminent language arts scholars, the modal grade level at which four textbook series presented each skill, and the grade-level designations suggested by the SWRL textbook data. The SWRL grade level is the one at which most or usually all analyzed textbook series present (or have presented) the skill. This designation may or may not be the modal level; usually it is not.

Considerable grade-level variation is also found across state and local competency lists, as is evident in the chart on page 4 of the handout. This chart illustrates the variation across seven states. Information on these skills was compiled from representative state documents (Lawlor, 1979), which frequently place skills at levels higher than those at which the skills appear in textbook instruction.

Because the grade levels reported in the literature and by the competency lists are inconsistent and often do not parallel the actual level of textbook instruction, the grade levels of the writing skills derive from the SWRL textbook data. These data were also used because the most reliable data available on the classroom activities across the entire country are the textbooks students use. And when students must

function in the English classroom, the actual activities occurring there are the important criteria of success in that classroom. However, when an individual school district is developing its own list of critical skills, the grade levels of the skills, as well as the critical skills themselves, should be determined on the basis of the textbooks and curriculum of that particular district.

THE SKILLS FRAMEWORK

After the grade levels were specified, the skills were organized by categories into a framework of writing skills. The major categories of that framework are Handwriting, Spelling, Mechanics, Language, General Discourse, and Discourse Products. A Discourse Product is defined here as a whole composition, such as a story, an essay, a poem. Composing effective written products requires the use of many skills, or subskills, and these are covered in the General Discourse area. Page 5 of the handout shows a portion of the detailed framework--specific skills listed for subcategories of the General Discourse category.

The framework presents one way to organize writing skills. Other frameworks are possible; however, this one was satisfactory for our purposes. For identification purposes, category letters and individual skill numbers are combined to give each skill a code.

SKILL SPECIFICATIONS

After the skills were organized into a framework, skill specifications were written to delineate each skill and to ensure its appropriate

assessment. The skill specifications are based on content and task analysis and on textbook instruction. The relevant literature was also referenced in preparing the skill specifications. Samples of complete specifications for specific skills will be passed out later.

Skill Statement

The skill statement appears at the top of the page after its code number. In general, the skill statements are cast in terms used both by many language arts authorities and by competency lists. These skill statements generally must be more precisely defined by further explication before assessments can be devised.

Skill Descriptions

This explication is accomplished by the Skill Description. It notes factors involved in the use of each skill, including other inherent skills and prerequisite knowledge.

Assessments

The Skill Description is followed by the section on assessment. For most skills, three or more sample assessment items were written. However, for the writing samples that assess Discourse Products, only one item was written for each skill because these items are complex and because only one writing sample per skill should be included on an actual assessment instrument. Included in writing-sample specifications are the prompt, the scoring key, and scoring guidelines.

The items that were written are the most appropriate type for the skill to be assessed. Items that require a written response were

frequently provided because they are often the most appropriate assessment type. They were the only type included for Discourse Products because a student's ability to compose a complete written product can only be assessed through an actual composing task. When written-response items were suggested for skills other than Discourse Products, multiple-choice items were also included so that they can be used as substitutes when administering written-response items is not feasible. When multiple-choice items were either as appropriate as or more appropriate than written-response items, only multiple-choice items were prepared.

Item Descriptions

Following the sample assessment items is an item description that provides guidelines for preparing additional assessment items that are precise and appropriate.

PROBLEMS

Many problems that were encountered in wr ing the skill specifications may be encountered by others who wish to develop their own skill specifications based on their own local curriculum. These problems are in addition to the usual ones of avoiding sexism, ethnic stereotypes, and violence.

One problem was the discrepancy between the simplicity of the skill statement and the complexity of the actual skill. The skill statement frequently named what seemed to be a straightforward skill. However, writing the skill description required close analysis of all factors inherent in the skill and relevant to its assessment, and this close analysis often revealed that the skill was far more complex than its skill statement suggested. Some skills were exposed as so complex that they

should not be assessed. For example, many competency lists include for assessment a handwriting skill, "Writes sentences from dictation." However, content and task analysis revealed that this is no simple handwriting skill. Writing sentences from dictation involves many other skills, such as those of spelling, mechanics, and listening.

We found that some skills can only be assessed in terms of another skill. Distinguishing between proper and common nouns is one such skill. It is assessed by capitalization skills. Identifying most sentence types is covered by assessment of punctuation skills.

Sometimes a skill's complexity wasn't fully realized until assessment items and their descriptions were written. Then the skill description (and often even the items themselves) had to be completely rewritten.

Directions also posed a problem because a single inappropriate or omitted word may mislead students. Specific directions must be painstakingly written so students unquestionably understand what they are to do. Yet the directions must be devoid of technical words, when this restriction is possible.

Care must also be taken that each item assesses only one skill. For example, items assessing ability to construct coordinate subject-noun phrases must not include assessment of the skill of subject-verb agreement.

Written-response items and multiple-choice items pose different problems. Written-response items often appear to be simple to devise, but this simplicity is deceptive. Such items must be carefully developed or they will not elicit the desired response. If, for example, the skill being assessed is the transformation of a declarative sentence into a

question, the item must be devised so that the student does not simply rewrite the sentence in the same word order and add a question mark. Also, any component skill, or subskill, that should be assessed by the item must be explicitly prompted. If, for example, precise language is to be used in composing a written product, the item stimulus must tell students to use "exact words."

Written-response items are also difficult to score, so guidelines for rating responses must be included in the skill specifications. If such rating criteria cannot be devised, this inability suggests that there is a deeper problem with either the skill description or the assessment item.

On the other hand, multiple-choice items are easy to score, but they are usually even more exacting to devise than are the written-response items. Multiple-choice assessment items are particularly difficult to devise for writing because they should require a response that assesses writing ability rather than reading ability. This is an important factor in constructing writing assessment items, but it is frequently ignored.

The reading requirement of some multiple-choice items must also be considered. The amount of reading students must do to select a correct answer should be as limited as possible; however, this reading requirement sometimes cannot be lowered and still have the item assess the skill. Also the reading vocabulary used in the items must be limited to those words that are on controlled vocabulary lists for each grade level. Writing a good primary-level item with only those words available on primary reading lists is challenging.

The construction of distractor choices for multiple-choice items also poses development problems. Distractors used in the sample items and described in the item descriptions should be designed and written so that they serve diagnostic purposes, and this diagnostic information should be discussed in the item descriptions. Futhermore, distractor choices must be carefully screened so that they are appropriate for the grade level of the skill being assessed; for example, although the letter f is an appropriate distractor in items for spelling ph when it is assessed at grade 5, ph is not an appropriate distractor for f when it is assessed at grade 1.

Developing a language skills framework is a time consuming task. As should be evident by now, writing skill specifications is even more time-consuming. Yet such specifications must be written to determine what is to be done and to validate what has been done. The all-too-frequent approach consists of reading the skill statement and then whipping off a few items with a correct answer and several distractors without regard to isolating a single skill for assessment, to considering the mode being assessed, to guaranteeing the precision of the directions, or to serving diagnostic purposes with the use of appropriate distractors.

Writing skill specifications for this project required the full-time, eight-hour daily effort of four well trained professional staff members for approximately two months. However, this kind of effort is essential when the results of a project may be used in determining bilingual students' potential reclassification.

References

Georgia Department of Education. A design for an English curriculum. Atlanta: Georgia Department of Education, 1968.

Golub, L. S., and others. Present and proposed needs, objectives, and components of a language learning program. ERIC Reproduction Service Number ED 058 223, 1971.

Lawlor, J. State competencies for writing: Grades K-6 (Technical Note No. 3-79-07). Los Alamitos, California: SWRL Educational Research and Development, 1979.

Los Angeles City Schools. Instructional program grade charts. Publication No. 375, Los Angeles City Schools, 1964.

Squire, J. R. (Ed.). The teaching of English. The seventy-sixth yearbook of the National Society for the Study of Education, Part 1. Chicago: University of Chicago Press, 1977.

HANDOUTS FOR ANN HUMES PRESENTATION

SOUTHWEST REGIONAL LABORATORY
EDUCATIONAL RESEARCH AND DEVELOPMENT
PROFICIENCY VERIFICATION SYSTEM
SORT PROCEDURE FOR LANGUAGE ARTS
PAGE 32

PROGRAM: 4, GRADE: 6

SKILL: 62 PRONOUNS

PAGE:	PAGE CONTENTS:
12	INTERROGATIVE PRONOUN$ READ/STUDY$ INFORMATION$ ON WH- QUESTIONS
13	INTERROGATIVE PRONOUN$ READ/STUDY$ INFORMATION$ AS SUBJECTS
14	INTERROGATIVE PRONOUN$ READ/STUDY$ INFORMATION$ ON "WHAT" & "WHICH"
88	DEMONSTRATIVE$ READ/STUDY$ RULE(S)/DEFINITION(S)$ AS DETERMINER
89	DEMONSTRATIVE$ INDEFINITE FORM$ CHECK/IDENTIFY$ LIST$ AS DETERMINER
89	INDEFINITE FORM$ READ/STUDY$ RULE(S)/DEFINITION(S)$ AS DETERMINER
90	DEMONSTRATIVE$ INDEFINITE FORM$ WRITE$ SENTENCES FOR SPECIFIC INFORMATION/ITEMS$ AS DETERMINER
96	PERSONAL PRONOUN$ READ/STUDY$ FORM(S)
96	AS SUBSTITUTE$ READ/STUDY$ SENTENCES FOR SPECIFIC INFORMATION/ITEMS
96	PERSONAL PRONOUN$ AS SUBSTITUTE$ READ/STUDY$ RULE(S)/DEFINITION(S)
97	PERSONAL PRONOUN$ AS SUBSTITUTE$ CHECK/IDENTIFY$ WRITE$ REFERENT
97	PERSONAL PRONOUN$ AS SUBSTITUTE$ SUBSTITUTE$ FORM(S)$ IN SENTENCES
98	PERSONAL PRONOUN$ READ/STUDY$ RULE(S)/DEFINITION(S)$ OF FIRST, SECOND, THIRD PERSON
98	GENDER$ PERSONAL PRONOUN$ READ/STUDY$ RULE(S)/DEFINITION(S)
99	SINGULAR/PLURAL$ READ/STUDY$ RULE(S)/DEFINITION(S)
99	SINGULAR/PLURAL$ READ/STUDY$ FORM(S)
100	POSSESSIVE PRONOUN$ READ/STUDY$ RULE(S)/DEFINITION(S)$ INCLUDING FUNCTION AS DETERM
100	POSSESSIVE PRONOUN$ READ/STUDY$ INFORMATION
101	PERSONAL PRONOUN$ LABEL$ FORM(S)$ AS FIRST, SECOND, THIRD PERSON
101	SINGULAR/PLURAL$ LABEL$ FORM(S)
101	POSSESSIVE PRONOUN$ FILL IN$ CORRECT MEANING/EXAMPLE/FORM
102	NOMINATIVE$ OBJECTIVE$ READ/STUDY$ RULE(S)/DEFINITION(S)
102	NOMINATIVE$ OBJECTIVE$ READ/STUDY$ EXAMPLE(S)
102	NOMINATIVE$ OBJECTIVE$ READ/STUDY$ FORM(S)
103	NOMINATIVE$ READ/STUDY$ SENTENCES FOR SPECIFIC INFORMATION/ITEMS$ COMPOUND SUBJECTS
104	NOMINATIVE$ CHOOSE/DISCRIMINATE BETWEEN/AMONG$ CORRECT MEANING/EXAMPLE/FORM
104	NOMINATIVE$ READ/STUDY$ RULE(S)/DEFINITION(S)$ FOR COMPOUND SUBJECTS
105	OBJECTIVE$ READ/STUDY$ SENTENCES FOR SPECIFIC INFORMATION/ITEMS$ AFTER PREPOSITION
105	OBJECTIVE$ READ/STUDY$ RULE(S)/DEFINITION(S)$ FOR COMPOUND OBJECT OF PREPOSITION
106	NOMINATIVE$ OBJECTIVE$ WRITE$ SENTENCES FOR SPECIFIC INFORMATION/ITEMS
106	NOMINATIVE$ OBJECTIVE$ SAY/RECITE/ANSWER$ SENTENCE(S)/LINES$ ORALLY
106	NOMINATIVE$ OBJECTIVE$ FILL IN$ CORRECT MEANING/EXAMPLE/FORM
106	OBJECTIVE$ READ/STUDY$ RULE(S)/DEFINITION(S)$ FOR COMPOUND DIRECT OBJECTS
107	NOMINATIVE$ OBJECTIVE$ CHOOSE/DISCRIMINATE BETWEEN/AMONG$ CORRECT MEANING/EXAMPLE/F
107	NOMINATIVE$ READ/STUDY$ EXAMPLE(S)$ AFTER "BE"
107	NOMINATIVE$ READ/STUDY$ RULE(S)/DEFINITION(S)$ AFTER "BE"
111	* NOMINATIVE$ OBJECTIVE$ CHOOSE/DISCRIMINATE BETWEEN/AMONG$ CORRECT MEANING/EXAMPLE/F

NABE
Los Alamitos, California
April 19, 1980

SPECIFICATIONS FOR A MECHANICS SKILL

Skill: W4 Mp 7110

Skill Statement

Uses apostrophe in singular possessive forms.

Skill Description

The student uses apostrophes correctly in singular possessive forms. The student must identify the correct position of the apostrophe in the singular possessive (i.e., between the root word and the s; e.g., boy's). Singular possessives may be formed from singular proper or common nouns (e.g., John's/dog's) or from indefinite pronouns (e.g., someone's). However, indefinite pronouns are not included in assessment here because they are not as commonly used as nouns in textbook instruction at this grade level. Additionally, inanimate nouns are not assessed for possessives since they are less frequent and are not accepted by some authorities (e.g., not the cup's handle but the handle of the cup).

Implicit in this skill is the ability to distinguish between plural forms of nouns and singular possessives. However, students should not be expected to distinguish between singular and plural possessives at this grade level.

Knowledge of the technical terms apostrophe and singular possessive is not prerequisite to this skill.

Assessment

Sample Items

Directions: Which word is right?

1. _____ uncle is a sea captain.

   A. Dick's
   B. Dicks
   C. Dic'ks

2. Betsy has a pet bird. She wants to clean the _____ cage.

   A. bird's
   B. birds
   C. bir'ds

3. All the _____ went down the stairs.

   A. boy's
   B. boys
   C. bo'ys

Directions: Fill in the blank with the correct form of the underlined word. The form shows ownership.

4. Tony has a bicycle.

   _____ bicycle is red.

5. The bird has a nest.

   The _____ nest is in a tree.

6. My sister has a kitten.

   My _____ kitten drinks milk.

Item Description

The stimulus for the selected-response item type (items 1-3) is a sentence with a blank replacing a missing word. In items 1 and 2, the student selects the singular possessive form (choice A's) to fill in the blank. Distractors are forms with no apostrophe (choice B's) and forms in which the apostrophe precedes, by one letter, its correct position in the word (choice C's).

Item 3 is a distractor item for which the plural noun (choice B) is the correct response. Distractor items should be included in the assessment to give the student the opportunity to make discriminating responses.

The stimulus for the constructed-response item type (items 4-6) is a sentence with an underlined word. This word is to be used in forming the singular possessive. The student fills in the blank in the second sentence with the correct form.

Singular possessives selected for assessment can be proper or common nouns. However, proper nouns ending in s (e.g., James/Charles) should be avoided because there are two acceptable possessive forms (e.g., James's or James').

Providing sufficient sentence context is important in these items, particularly when common nouns are used. For example, in item 2, the sentence "Betsy has a pet bird" is included so that the student knows that the possessive form is singular. The sentence "She wants to clean the _____ cage" does not provide enough information to facilitate a correct choice since either bird's or birds' could be correct.

SPECIFICATIONS FOR A GENERAL DISCOURSE SKILL

Skill: W4 GDe 1120

Skill Statement

Elaborates paragraphs. Limits a paragraph to one main idea.

Skill Description

The student includes only one main idea in a paragraph; he/she is able to identify information that cannot be added to a paragraph because it does not relate to the main idea. This information may consist of irrelevant supporting statements or irrelevant supporting evidence for a supporting point; however, knowledge of this terminology is not prerequisite to this skill.

Assessment

Comments: This skill is best assessed by a writing sample. If this assessment is not feasible, the selected-response item may be used. Care must be taken that the item type used does not assess a reading rather than a writing skill as does one item type frequently included in assessment instruments for writing. This item type requires students to read a paragraph and then select a sentence that gives the main idea of the paragraph, thus assessing a receptive-language rather than a productive-language skill.

Sample Items

Directions: One sentence can be added to the paragraph because it fits the idea of the paragraph. Which sentence is it?

1. Football players must be strong to play in a game. They must be able to knock other players down. Players must be able to run fast because fast runners can carry the ball to make points in the game.

   A. Football players have to throw the ball hard.
   B. Football players wear shirts with numbers on them.
   C. Some football players go to school to learn to play ball.
   D. Basketball players run fast.

2. You must take good care of a pet bird. You must give it water to drink. Its cage must be kept clean so the bird will stay well. You can put paper in the bottom of the cage to make it easy to clean.

   A. The bird will get sick if it does not have water.
   B. A pet bird can sing sweetly.
   C. You can get a pet bird at a pet store.
   D. A pet rabbit may also live in a cage.

3. The trunks of elephants are strange, but useful. Elephants use their trunks to pick up things. Food is one thing that elephants pick up with their trunks. They also use their trunks to splash water over themselves.

   A. Elephants use their trunks to put food in their mouths.
   B. Elephants have funny feet, too.
   C. The ears of elephants have hair on them.
   D. Pigs also look strange.

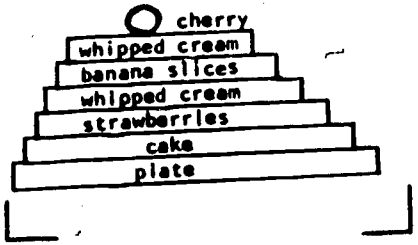
Item Description

The stimulus paragraph has an initial topic sentence that expresses the main idea. All other sentences pertain to the main idea stated in that topic sentence. At least two of these sentences should present supporting ideas and at least one sentence should state evidence for one of the supporting ideas. Longer paragraphs are desirable as examples of good writing, but would increase item difficulty. Main idea is most apparent in and pertinent to expository discourse. Therefore, the stimulus is written in expository discourse.

The student selects a sentence to add to a paragraph; the sentence fits the main idea. The correct answer should be either another supporting point or evidence for one of the supporting points in the stimulus. If the correct answer is evidence, it must pertain to any supporting points lacking supporting evidence in the stimulus unless all supporting points have such evidence.

Fruit and Whipped Cream Treat

Specifications for picture to be drawn:



Contents

1 yellow cake
1 cup of strawberries
1 cup of cut-up bananas
2 cups of whipped cream
1 cherry

Write at least one paragraph that explains to someone who has never fixed food how to put the Fruit and Whipped Cream Treat together.

- Start with the first thing you do.
- Tell about each part of the Treat that you put together.
- Tell about the parts in the order that you must put them together.
- Be sure to use exact words.
- Use words like next to join some of the sentences in your paragraphs.

Item Description

Using expository-writing skills, the student writes the composition requested by the stimulus. The stimulus must be a picture of a simple whole with its parts graphically displayed. The order in which the parts are connected should also be simply and graphically displayed. At this grade level, no working (i.e., moving) parts should be included in the part-whole stimulus object.

The series of instructions on including features helps ensure production of these features and facilitates the construction of an appropriate scoring key that evaluates these features.

Scoring Key

The skills array of the key's matrix consists of the features of good writing that are effected when students employ the content and form skills appropriate at this grade level for this writing task. The skills corresponding to each criterion are listed by framework number on the subsequent scoring guide. Performance ratings of good, average, or unacceptable are given on each criterion in accordance with the rating considerations that the scoring guide describes. Thus the scoring key can be used to elicit diagnostic information about individual component skills. It can also be used to determine a total writing score by assigning numerical equivalents to the good/average/unacceptable categories.

Scoring Criteria	Good	Average	Unacceptable
CONTENT:			
Includes all parts of the Treat.	___	___	___
Writes about the parts in the order in which they are put together.	___	___	___
Uses precise language.	___	___	___
Uses logical transitions.	___	___	___
Uses sentences that pertain to the main idea.	___	___	___
FORM:			
Uses correct grammar.	___	___	___
Uses good sentence structure.	___	___	___
Capitalizes and punctuates correctly.	___	___	___
Spells correctly.	___	___	___
Writes legibly, with appropriate margins and indentation.	___	___	___

Scoring Guide

The scoring guide describes the guidelines for determining what constitutes a good, average, or unacceptable score on each criterion in the scoring key.

W5 OPpr 4000 (continued)

The skills corresponding to each criterion are listed by framework number within the brackets that follow the criterion statement. The scoring guide references only those skills that (1) must be employed to produce the sample and (2) are designated as optimal assessment skills at the immediate and earlier grade levels. Scoring guidelines for one content characteristic and one form skill are exemplified below. Actual skill specifications include complete guidelines.

CONTENT EXAMPLE:

Includes all parts of the Treat (i.e., plate, cake, stawberries, whipped cream, banana slices, whipped cream, cherry). [Skill GOe 2313]

Good: All parts are included.

Average: Most parts are included.

Unacceptable: Many parts are missing.

FORM EXAMPLE:

Spells correctly. [Skills Sc 1000, 2000, 3000; Sv 1000, 2000, 3000, 4000; Sa 1000, 2000; Ssp 2000, 3000, 4000]

Good: Most words are correctly spelled.

Average: Several different words are misspelled.

Unacceptable: Many different words are misspelled.

Comment: The average and unacceptable ratings for the "Spells correctly" criterion refer to "different words" because many instances of misspelling the same word should be evaluated as one misspelled word.

PROBLEMS/SOLUTIONS

1. The inclusion of distractor type A requires that the student distinguish between singular and plural possessive forms at this grade level. However, we identified the use of the apostrophe in plural possessives as a fifth-grade skill. Consequently, distractor A was revised so that the apostrophe preceded its correct position in the word (e.g., bir'ds).

2. Distractor C is grammatical and plausible. Thus the assessment emphasis is on the correct interpretation of the picture stimulus, rather than recognition of word order. However, the items should not place undue emphasis on the student's ability to "read" pictures. Therefore, we decided that distractor type C should clearly be implausible or the picture should be changed.

3. The dialogue quotation is used as the stimulus in this item so that the exclamatory nature of the sentence can be suggested by the speaker tag. However, if the tag follows the quotation, the item actually assesses punctuation of dialogue quotations rather than terminal punctuation. Consequently, we specified that the speaker tag should precede the quotation in these items; e.g.:

   Jan shouted, "The barn is on fire__"

4. The problem here is that distractor C is inappropriate. This distractor forms the word bare, a homophone for the correct spelling. While the spelling of homophones is an important skill in and of itself, homophones should be assessed separately and should not be used as distractors in other spelling skills. Distractor C was revised to ere.

5. The correct response (choice D) requires additional context to establish its number. We decided to delete plural nouns that retain the same form as the singular (e.g., sheep, fish, deer) from this assessment.

6. As written, this item does not measure ability to discriminate rhyming words, but ability to discriminate final consonants. In order to match rhyming words, students must recognize that both the medial sound and the final sound are the same. Thus one distractor should have the same vowel sound as the target word but a different final consonant; the other distractor should have a different vowel sound, but the same final consonant. Distractor B was revised to fun.

7. This item assesses a reading rather than a writing skill--reading for the main idea. When students write, they don't construct a main idea after the paragraph. Rather, the writing problem is having a main idea and keeping to it.

Directions: One sentence can be added to the paragraph because it fits the idea of the paragraph. Which sentence is it?

Football players must be strong to play in a game. They must be able to knock other players down. Players must be able to run fast because fast runners can carry the ball to make points in the game.

A. Football players have to throw the ball hard.

B. Football players wear shirts with numbers on them.

C. Some football players go to school to learn to play ball.

D. Basketball players run fast.

8. This item type requires too much reading. The item was revised so that the student only had to read the paragraphs once, selecting the point at which a new paragraph should begin; e.g.:

Directions: Read this story part. Where should a new paragraph begin?

The bus returned to school in the afternoon. (A) Linda and Ricardo got off the bus and walked toward the classroom. (B) "Did you enjoy the trip to the zoo?" Ricardo asked. (C) Linda replied, "I sure did." (D) "My favorite part was the monkey house because some of those monkeys act just like people."

9. This item has two problems: (1) the directions are not specific enough, and (2) the stimulus is inappropriate as it is written. Formal and informal language are relative to audience. Thus the directions should specify the audience. An appropriate writing situation in which a sixth grader might be required to use formal language would be a letter to someone the student had never met.

The use of gang in this sentence is unclear. If the writer is referring to a group of juvenile delinquents, then gang may be a more formal and accurate description of the students than any of the answer choices. This item was revised:

Directions: The underlined word is too informal for most letters. Which word should you use if you wrote the sentence in a letter to someone that you have never met?

A new bunch of students will be coming to our school.

A. gang

B. group

C. crowd

INAPPROPRIATE ITEMS FOR ASSESSING WRITING SKILLS

1. Skill Statement

Uses apostrophe in singular possessive forms. [Grade 4]

Assessment

Sample Item

Directions: Which word is right?

Betsy has a pet bird. She wants to clean the _____ cage.

A. birds'

B. birds

(C.) bird's

2. Skill Statement

Recognizes correct word order: subject-verb-object. [Grade 2]

Assessment

Sample Item

Directions: Which is a good sentence that tells about the picture?

[picture: girl running after a boy]

(A.) Mary is chasing John.

B. Mary John is chasing.

C. John is chasing Mary.

3. Skill Statement

Uses exclamation point to end an exclamatory sentence. [Grade 4]

Assessment

Sample Item

Directions: Which is the best mark to use in the blank?

"The barn is on fire__" Jan shouted.

A. ,

(B.) !

C. ?

D. .

4. Skill Statement

Spells vowel-r pattern: /er/-are, /er/-ear. [Grade 4]

Assessment

Sample Item

Directions: Which letter or letters finish the word?

[picture: bear]

b_____

A. air

(B) ear

C. are

D. er

5. Skill Statement

Distinguishes noun forms: singular/plural. [Grade 3]

Assessment

Sample Items

Directions: Which word means more than one?

A. glass

B. sleep

C. house

(D) sheep

6. Skill Statement

Identifies words that rhyme. [Grade 2]

Assessment

Sample Item

Directions: Which word rhymes with the name of the picture?

[picture: nine]

A. fire

B. five

(C) fine

7. Skill Statement

Limits a paragraph to one main idea. [Grade 4]

Assessment

Sample Item

Directions: Which sentence tells the main idea of the paragraph?

Football players must be strong to play in a game. They must be able to knock other players down. Players must be able to run fast. Fast runners can carry the ball to make points in the game.

(A) It takes strength to play football.

B. Running is an important part of football.

C. Football is a fun game that everyone should play.

D. Scoring points is important.

8. Skill Statement

Uses dialogue for only one speaker in a paragraph. [Grade 5]

Assessment

Sample Item

Directions: Which conversation is written correctly?

(A) The bus returned to school in the afternoon. Linda and Ricardo got off the bus and walked toward the classroom. "Did you enjoy the trip to zoo?" Ricardo asked.
"I sure did," Linda replied. "My favorite part was the monkey house," she said, "because some of those monkeys act just like people."
Ricardo laughed and said, "Yes, you're right."

B. The bus returned to school in the afternoon. Linda and Ricardo got off the bus and walked toward the classroom. "Did you enjoy the trip to the zoo?" Ricardo asked. "I sure did," Linda replied. "My favorite part was the monkey house," she said, "because some of those monkeys act just like people." Ricardo laughed and said, "Yes, you're right."

C. The bus returned to school in the afternoon. Linda and Ricardo got off the bus and walked toward the classroom. "Did you enjoy the trip to the zoo?" Ricardo asked.

"I sure did," Linda replied. "My favorite part was the monkey house," she said, "because some of those monkeys act just like people." Ricardo laughed and said, "Yes, you're right."

9. Skill Statement

Uses appropriate formal/informal language. [Grade 6]

Assessment

Sample Item

Directions: The underlined word is too informal. Which word should you use?

A new gang of students will be coming to our school.

A. bunch

(B) group

C. crowd

"READING ASSESSMENT AND THE BILINGUAL TEACHER"

NABE Pre-Conference Workshop

Laila Fiege-Kollmann

April 19, 1980

INTRODUCTION

Bilingual children and children whose mother tongue differs from English (NES/LES) usually take part in the reading instruction intended for monolingual English speaking children. We know that the NES/LES or bilingual children are learning to function in English and that they are acquiring reading skills at the same time. Yet, we often tend to forget this dual acquisition process, and expect these children to attend to reading instruction in the way typical monolingual children do. Not only are the bilingual or English language learners taught to read by monolingual reading methods; their progress in reading is assessed with the standard reading tests mandated for use in the public schools today.

"Bilingual and NES/LES children do not receive high scores on reading assessment." This is a comment we often hear and maybe even take as a fact of life without asking some hard questions about assessment in general, the instruments used in a given case in particular, the demographic parameters on the populations under study, and so on. Reading assessment in general is presently undergoing scrutiny. For years, standardized achievement tests were used to measure progress in reading. Recently, there have been several studies to indicate that the skills tested on many standardized achievement instruments are on the periphery of the skills emphasized and practiced in instruction (Buchanan & Milazzo, 1980; Berliner & Rosenshine, 1976; Armbruster et al., 1977). Presently, reading assessment research is focusing on the link between instruction and assessment, i.e., in a normal classroom setting, progress in reading should be measured by instruments which reflect the skills children have had the opportunity to learn through instruction.

At SWRL for many years now we have been doing research on the nature of reading instruction as is evidenced in the contents of commonly used reading series and reading surveys. I will discuss very minimally the design and development of reading proficiency survey instruments, and

then I will spend some time explaining some recent reading proficiency data gathered from monolingual, bilingual, and NES/LES children.

Design and Development of Survey Instruments

One of the foremost aspects in the assessment of reading skills is the definition of the scope, sequence, and emphasis of reading skills within instructional practices. At SWRL we accomplished this by a lengthy process of coding instructional activities, grade-by-grade, and page-by-page. This computer-analyzed information shows the introduction, emphasis and maintenance of reading skills across grade levels, and it is being used as a basis for designing large scale information systems measuring reading skill proficiency. Such surveys give a summary description of what has been accomplished in a school year and provide important indicators of what the student can be asked to do in the following year.

I will discuss survey data from two information systems, developed in co-operation with two different school districts. The first set of data (Tables 1 and 2) will be used to look at the relationship of age and reading performance; the second set of data will be used to discuss children's language background and reading performance.

Age

It is often said that the child from a different language background is older than the monolingual child at a given grade level. In looking at third graders in a large urban school district, we found that 80% of the monolinguals are 9 years old, and 14% are 10 years old; for bilinguals the percentage differed about 10%: 68% were 9 years old, and 22% were 10 years old. This difference is less than one normally would have expected.

Table 1 presents data on a fifth grade survey tailormade to the needs of an urban school district. The survey taps skills that the district considered important for the students to know before leaving

the elementary school system. The table indicates the breakdown of average percent scores by skill area and age group. As can be seen from the results, the scores go down as the grade levels progress. This decrease in scores is regular and systematic except in one case. In the comprehension skill area 12 year old children perform worse than 13 year old children. From other scores in Table 1 one would have expected the opposite to be true. I can not explain this discrepancy without further analysis of the data.

Table 1

Breakdown of average percent scores by skill area and age group; 5th grade; N=2856

Skill area		Age 10	11	12	13
Word Meaning		88	82	64	46
Sentence Completion		92	86	69	52
Verb, Noun Agreement		91	85	66	60
Comprehension		79	72	52	60
Table of Contents, Index		92	87	80	67
Alphabetization, Glossary	Study Skills	87	80	66	65
Overall		87	81	65	58

The main conclusion to be drawn from Table 1 is the one already stated: as the age level of the students at a particular grade level increases the scores tend to decrease. If children from a different language background are older than their monolingual classmates, their scores also tend to be lower than the scores of their younger classmates. Table 1 can also be used to show the difference in skill area scores between age groups. Comprehension skills are hardest for the 10-year-old students whereas word meaning and sentence completion

skills are more difficult for the 13-year-old students. For example, less than half of the 13-year-old students answered each word meaning item correctly while the performance score was 60% for comprehension skills. For the 10-year-old students word meaning scores (88%) were about 10% higher than the comprehension scores (79%). Comprehension skills represent global skills or skill aggregates which depend on the students' ability to mesh their own knowledge with the information found in text. It may be that the older students can more effectively use the contextual cues available in longer text and bypass same of the problems they have in direct word meaning tasks. On the other hand, the younger students may benefit from the shorter and more direct tasks, patterned after exercises found in their workbooks.

The younger students do about 20% better in comprehension skills than older ones. Yet, this difference is much less than it is for word meaning, sentence completion, and verb/noun agreement skills where the difference in scores ranges from 30 to 40%. It may be that the older students can take advantage of the general knowledge they have in comprehension skills which depend more on global skills and skill integration than the other skill areas, which include skills found solely in reading instruction.

The 10-year-old students performed well in the study skills surveyed, and the 13-year-old students received their highest performance scores in this skill area. This is what one would have liked to expect since the study skills surveyed consisted of skills which children practice in reading but which are used also when reading in other subject matter areas such as social science and health. They are organizational skills that require less reading than what is demanded in other skill areas.

Language Proficiency

Table 2 shows the breakdown of average percent scores by skill area and language proficiency for the same survey which was discussed

in the previous section. The school district categorized the students as Fluent-English-Speakers, Some-English-Speakers, and No-English-Speakers. The No-English-Speaker label is misleading. I do not believe that there are any children in the public school system who do not speak and understand at least minimal English, and so I have labeled the groups Fluent English, Functional English, and Minimal English. The scores are as is to be expected. The less English one knows, the lower the scores.

Table 2

Breakdown of average percent scores by skill area and language proficiency; 5th grade, N=2856

Skill	Language Proficiency		
	Fluent English	Functional English	Minimal English
Word Meaning	86	65	49
Sentence Completion	91	77	52
Verb, Noun Agreement	90	70	52
Comprehension	77	61	47
Table of Contents, Index	91	81	64
Alphabetization, Glossary	85	75	53
Overall	87	70	52

The results presented in Table 2 offer other findings as well. The average scores of Functional English speakers are about 10 to 20% lower than the scores of Fluent English speakers. The smallest differences can be found in the study skills, comprehension and sentence completion skills. These students tend to do less well with the word meaning and

verb/noun agreement skills. These skills require lots of exposure to the language to be learned combined with practice, and they often are stumbling blocks for second language learners. In study skills which depend more on organizing than extensive reading, these readers do relatively well.

In another study, the reading proficiency of 3rd and 6th grade students of an urban school district was surveyed. A brief description of the content of each survey is given in Tables 3 and 4. Table 5 presents the cumulative distribution of the results in 3rd grade; Table 6 lists the 6th grade results. In this particular study, we had a representative sample of the school district as a whole, and the results of this sample at 60% level are given in the right hand columns of Tables 5 and 6. We also had a representative sample of Hispanic students in the district sample, and these students were classified by the district as follows:

English speakers
Bilinguals
Limited English speakers, in English reading programs
Limited English speakers, not in English reading programs
Non-English speakers

Table 3

Content of the Grade Three Survey

Skill Area	Number of Items
Decoding (consonant digraphs; variant c, g; irregular vowel patterns)	11
Structural Analysis (plurals; contractions; abbreviations; verb inflections; syllabication)	16
Vocabulary (sight words; definitions; word types)	14
Comprehension (story detail; sequence; classification; prediction)	12
Location/Study (parts of a book; alphabetical order; directions)	10
Overall survey length =	63 items

Table 4

Content of the Grade Six Survey

Skill Area	Number of Items
Decoding (vowel patterns; consonant digraphs)	15
Structural Analysis (quantity prefixes)	4
Vocabulary (sight words; definitions; multiple meaning words; antonyms, synonyms, figurative language	14
Comprehension (main idea; sequence; cause/effect; classify information; quotation marks; conclusions; relevant/irrelevant information)	20
Location and Study Skills (reference books; guide words; charts; graphs; maps; diagrams)	21
Total Survey Length =	74 items

Table 5

Performance Summary on the Grade 3
Reading Survey
Cumulative Distribution

Student Classification	Skill Area	Cumulative Distribution 90%	80%	70%	60%	60%
	Decoding					83
District Overall (N=1773)		55	71	80	85	
English (N=200)		55	70	77	82	
Bilingual (N=245)		24	32	44	60	
Ltd. English, Reading (N=112)		18	21	31	36	
Ltd. English, Not Reading (N=61)		9	(9)	18	(18)	
Non-English (N=34)						
	Structural Analysis					76
District Overall		45	55	73	78	
English		42	57	70	77	
Bilingual		13	21	37	44	
Ltd. English, Reading		14	16	28	33	
Ltd. English, Not Reading		6	9	(9)	(9)	
Non-English						
	Vocabulary					69
District Overall		46	60	64	71	
English		28	48	55	68	
Bilingual		6	10	15	27	
Ltd. English, Reading		8	16	18	22	
Ltd. English, Not Reading		9	(9)	(9)	(9)	
Non-English						
	Comprehension					75
District Overall		47	58	73	77	
English		43	61	72	78	
Bilingual		13	18	36	41	
Ltd. English, Reading		12	14	24	26	
Ltd. English, Not Reading		3	6	9	(9)	
Non-English						
	Location/Study					69
District Overall		33	48	63	72	
English		20	49	64	73	
Bilingual		11	23	32	49	
Ltd. English, Reading		12	17	19	29	
Ltd. English, Not Reading		3	(3)	6	12	
Non-English						

Table 6

Performance Summary
on the Grade 6 Reading Survey
Cumulative Distribution

Student Classification	Skill Area	Cumulative Distribution 90%	80%	70%	60%	60%
	Decoding					84
District Overall (N=1800)		52	66	70	86	
English (N=198)		44	58	67	84	
Bilingual (N=270)		7	13	16	53	
Ltd. English, Reading (N=30)						
	Structural Analysis					69
District Overall		23	(23)	60	(60)	
English		23	(23)	60	(60)	
Bilingual				50	(50)	
Ltd. English, Reading						
	Vocabulary					71
District Overall		13	20	35	62	
English		13	20	34	64	
Bilingual				7	17	
Ltd. English, Reading						
	Comprehension					70
District Overall		17	33	45	60	
English		17	30	41	56	
Bilingual			3	10	20	
Ltd. English, Reading						
	Location/Study					71
District Overall		21	34	47	60	
English		20	36	49	58	
Bilingual		3	13	19	22	
Ltd. English, Reading						

In looking at Table 5, one has to conclude that some students in the Non-English and Limited-English groups are clearly mislabeled. If 9 students out of 34 in this Non-English group can score at 90% level in the word meaning skill, they have to be able to read and understand English. It seems that our classification procedures need redefining and reformation.

The cumulative distribution scores show how well the students did at 90%, 80%, 70%, and 60% levels. For example, 55% of the bilingual third graders answered 90% of the survey items correctly in the decoding skill area, and 77% performed at 70% or higher level in the same skill area (the percentage at the 70% level is larger than at the 90% level because it includes all students who performed at the 90% and 80% levels as well).

In comparing the 3rd grade survey results between the district overall and the language groups (Table 5), it becomes clear than the English speaking Hispanic children and the bilingual children perform as well as the district does as a whole. This is true for all skill areas.

English speaking Hispanic and bilingual children in the 6th grade, (Table 6) perform as well as the district as a whole in the decoding skill area , but in all other skill areas their scores are about 10% lower than the performance of the district as a whole. Thus, in the 3rd grade, these children are performing at the overall district level whereas the performance in critical skill areas drops as the grade levels progress. There may be several reasons for this drop: more, older immigrant children, more mobility, more absenteeism at the higher grade levels, reading materials that are out of synchronization with the maturity level of the older students, and so on. For example, earlier in this paper I discussed the relationship between age and reading proficiency. Although the Hispanic English-speaking and the bilingual children performed as well as the district overall, they did not perform as well as the monolingual Anglo children did. On the average, the scores of the Anglo children were about 15% better in the 3rd grade in

each skill area except vocabulary where the difference was about 20%. Among the 6th graders the difference was about 10%, except in the decoding skill area where there was no difference. The gap in the scores between the Anglo children and the Hispanic bilingual/English speaking children seems to be narrowing down by the 6th grade, which is encouraging.

Discussion

I have presented here some comparative performance data on children from different language backgrounds. The survey instruments used were designed to comply with specific school district needs. The survey items were developed to represent the kinds of items children see in their instructional materials. Although the survey instruments have been developed for the English speaking population, they can give us pertinent information about the relationship between the language background and reading proficiency. The data bring out more research questions than they answer: student, age, degree of proficiency, relationship of what is surveyed and what is actually taught, and so on.

The cut-and-dry classification into different language groups results in situations where children are grouped wrong. Presently there are no instruments or interview techniques on which the practitioner could rely in making more accurate decisions about student placement.

It is sometimes assumed that monolingual English readers typically receive high scores and that these readers perform equally well in all skill areas. As these data show, the performance patterns among skill areas are mixed; they tend to vary from skill area to skill area. Readers usually do best in decoding skills whereas performance levels drop in vocabulary and comprehension skills (Fiege-Kollmann, 1979, 1980). For instance, in the 6th grade district sample (Table 6), 84% of the children scored 60% or higher in the decoding skill area, but only 70% scored at that level in the comprehension skill area.

The performance in all skill areas, and especially in structural analysis and study skills, is dependent on whether or not the children have received instruction in the specific skills within the skill area. The children may learn the meaning of a word outside the instructional setting, but if they are not taught to interpret maps or diagrams or to understand the meaning of a quantity prefix, they can not perform well on items measuring these types of skills.

It is usually the case that the skill area which is easiest for the monolingual English speaking readers is also the easiest for other language groups. Thus, all groups performed best in the decoding skill area whereas performance level drops for vocabulary or comprehension skill areas. In these studies, there was no one set of reading skills in which the basic differentiation between different language groups changed in any appreciable way.

The discussion of reading proficiency in terms of achievement in various skill areas demonstrates the need for such an approach in reading assessment. The comprehensive nature of reading instruction is reflected best in a skill profile, rather than in a single performance score. Skill area scores will give a truer picture of individual achievement by pinpointing the skill areas where a student is progressing well and where more instruction is needed. Furthermore, the skill area information will give the teacher or practitioner a basis for moving the students on to a higher level of instruction within one skill area even though they may require additional help in some other area. If decoding scores are high and study skills low, the student should be able to move on to more advanced decoding while in study skills more instruction might be necessary.

Another important issue critical to reading proficiency assessment is the match between what is taught and what is measured. If the match is good, the teachers at least know that the students have had the opportunity to learn the skills actually measured.

Often the skills included in a survey instrument are specified by a district or state continuum, rather than being based on reading instruction and curriculum materials. If the survey instrument is designed to follow continuum objectives, the match between these objectives and instructional practices becomes critical; should the match be poor, the likelihood of low performance scores is high.

These issues are of particular importance when the language background of the students is different from the mainstream instruction. These students are extremely dependent on school learning since it is less likely for incidental learning to take place outside the reading classroom.

In looking at reading performance, it is important to know (1) the instrument, its content and purpose; (2) the curriculum, so that the one being assessed does not become the victim between curriculum and assessment; and above all, (3) the child. No instrument will assess some aspects of the child's skill proficiency as accurately as the teacher who interacts with him/her on a regular basis. The assessment instrument should be used to confirm teacher expectations especially when used with children form different language background.

References

Armbruster, B. B., Stevens, R. & Rosenshine, B. Analyzing Content Coverage and Emphasis: A Study of Three Curricula and Two Tests (Technical Report 26). University of Illinois at Urbana Champaign: Center for the Study of Reading, 1977.

Berliner, D. C. & Rosenshine, B. The Acquisition of Knowledge in the Classroom. Beginning Teacher Evaluation Study (Technical Reports Series). San Francisco, California: Far West Laboratory, 1976.

Buchanan, A. & Milazzo, P. Equating Instructional Accomplishment Inventories and Standardized Achievement Tests. Los Alamitos, California: SWRL Educational Research and Development, 1980.

Fiege-Kollmann, L. A Profile of Reading Instruction (Technical Note 3-79-22). Los Alamitos, California: SWRL Educational Research and Development, 1979.

Fiege-Kollmann, L. General District Analysis: Survey of Essential Skills in Los Angeles Unified School District, Reading Component 1979 (Technical Note 2-80-08). Los Alamitos, California: SWRL Educational Research and Development, 1980.

SOME CONSIDERATIONS IN CONSTRUCTING AND ADMINISTERING LANGUAGE PROFICIENCY TESTS

NABE WORKSHOP PRESENTATION
April 19, 1980

Bonita Ford
National Center for Bilingual Research

SOME CONSIDERATIONS IN CONSTRUCTION AND ADMINISTERING LANGUAGE PROFICIENCY TESTS

NABE Workshop Presentation

Bonita Ford

Introduction

In this presentation I intend to consider some of the problems which one is likely to encounter in constructing language proficiency tests, concentrating mainly on those having to do with Spanish testing. I will also consider some of the problems which may arise in administering these kinds of tests, and I will suggest some strategies for dealing with these problems.

What is a language proficiency test?

A language proficiency test is a tool which is used to determine the degree to which a child can understand, speak, read, and write in a given language--say, English or Spanish. This is not to be confused with tests which have been constructed to determine language dominance (Does this person know one language better than the other?) and which involve testing in two languages at the same time.

Why is it not advisable to translate language proficiency tests?

Very often the need arises to test language proficiency in a language for which there are no tests. The solution to this problem is either to construct a test for that language or to translate (adapt) an existing test from one language to the other. In the majority of cases people choose the second solution for a number of reasons--ease, speed, economy, etc. This practice, however, can result in a test which is not valid in the second language, that is, it does not test the same things that it did in the original nor does it test what it appears to.

Each language has its own particular types of constructions, some of which are mastered at an earlier age and some of which are not mastered until later because they are more difficult. These structures may or may not be the same in both languages. In translating a language proficiency test one might end up testing for a very easy type of construction in one language instead of a difficult one, as was originally intended, or one might miss testing structures which are important in the second language. Consider, for example, the underlined word in the following sentences:

John is a student.
John is happy.
John is hungry.

An English language proficiency test would be assessing a person's ability to produce the appropriate form of the verb "to be" and all three sentences would require the same form: "is." The same sentences translated into Spanish would require three different verbs:

Juan es estudiante.
Juan está contento.
Juan tiene hambre.

Likewise, a simple adjective in English, such as in "John is tall," Mary is tall," John and Mary are "tall," will require three different forms of the adjective because of gender and number agreement in Spanish: Juan es alto, María es alta, Juan y María son altos.

Construction A Language Proficiency Test

The purpose of a language proficiency test is to determine whether a student can understand a language and whether he/she can communicate in that language. We may also be interested in finding out whether the student can not only communicate orally, but can also read and write in that language. A fremework such as the following can be used to guide us in constructing a language proficiency test:

Comprehension	Production
• Listening	• Speaking
• Reading	• Writing

Within each of these categories we might want to test different areas, such as grammar, vocabulary, decoding, spelling, etc. We could then go about constructing different tasks to assess each of these processes separately, while still keeping in mind that they are all interrelated. Based on my experience in constructing language proficiency tests, I will point out some problems that might be encountered and recommendations for avoiding those problems.

Some Problems One Might Encounter in Constructing a Language Proficiency Test

1. Instructions. The instructions should never be more complicated that the task itself. Care must be taken to ensure that the vocabulary used in stating the instructions is as simple as possible and never beyond the student's appropriate reading level. Instructions must be written in very simple language and examples should be provided, where possible.

2. Pictures and Drawings. Pictures or drawings are a very easy way to test a child's vocabulary and they are often used in testing comprehension, where the child is asked to point to the appropriate picture after a word is given in the target language. The problem often encountered here is that it is not always easy to find or draw a picture of what you want to test for, and this may constrain your choice of items. For example, the vocabulary item to be tested in the following pair of drawings was "ceja"--eyebrow.

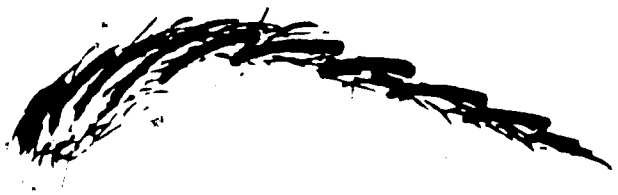


Figure 1: "ceja"--eyebrow.

Unfortunately, most of the children to whom this drawing was shown could not come up with the correct word. The problem was not that they didn't know the word "ceja," but that the drawing did not look like an eyebrow at all. When other people were asked what they thought the drawing represented, they said "a piece of yarn" or "a caterpillar." The drawing in Figure 2 yielded better results.

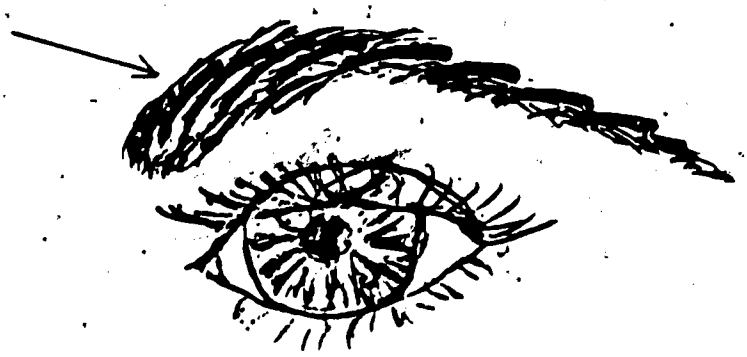


Figure 2

3. Item Validity. Does the item really test what you think it does? Once you have determined what you want to test on a particular item, the next step is to construct a question that will test it. You make up a question and you are convinced that people who know the answer will respond correctly, and that whoever doesn't respond correctly doesn't know the correct answer. It's quite possible, however, that even though you were guided by the best of intentions, the question you constructed did not really test what you thought it did.

In attempting to test the phrase "dejar entrar al perro"--to let the dog in-- we had to go through three different pairs of drawings before we came up with a good set of responses (Figures 3, 4 and 5). The object was to get the student to correctly identify the picture that corresponded to the sentence, "The girl let the dog in the house." The two drawings in Figure 3 were confusing because Figure 3a could be interpreted as "The girl wants the dog to come out," or "The girl wants to go in the house." Figure 3b could be "The dog wants to go in the house." Neither drawing clearly elicits the correct response, "The girl let the dog in the house." Likewise, the pair of drawings in Figure 4. Figure 5 shows the pair of drawings which produced the best results. Because there is a logical relationship between the two pictures--the girl sees that the dog wants to come in the house, then she opens the door and lets the dog in--students are more likely to choose the correct drawing.

Figure 3

a.

b.

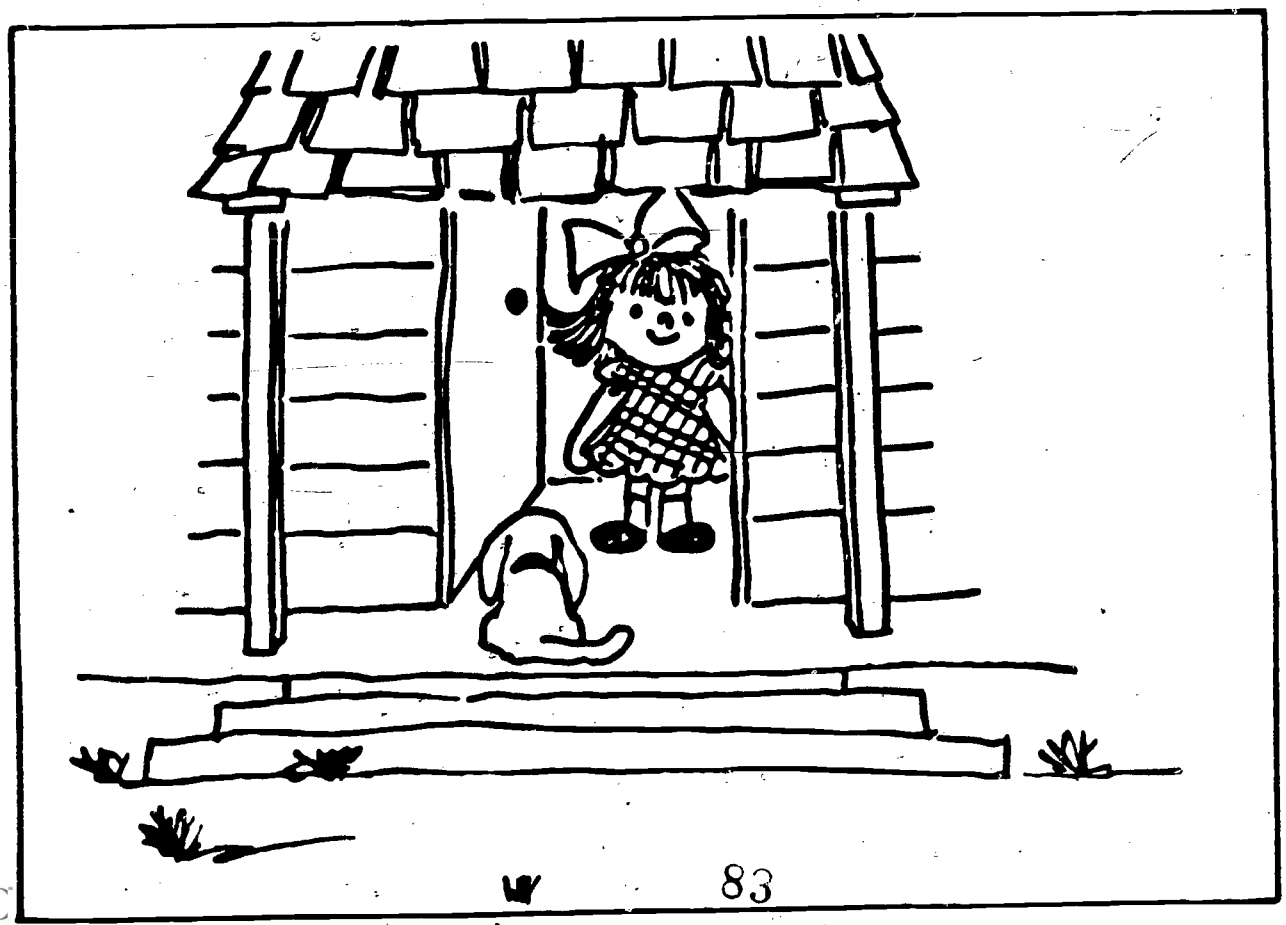
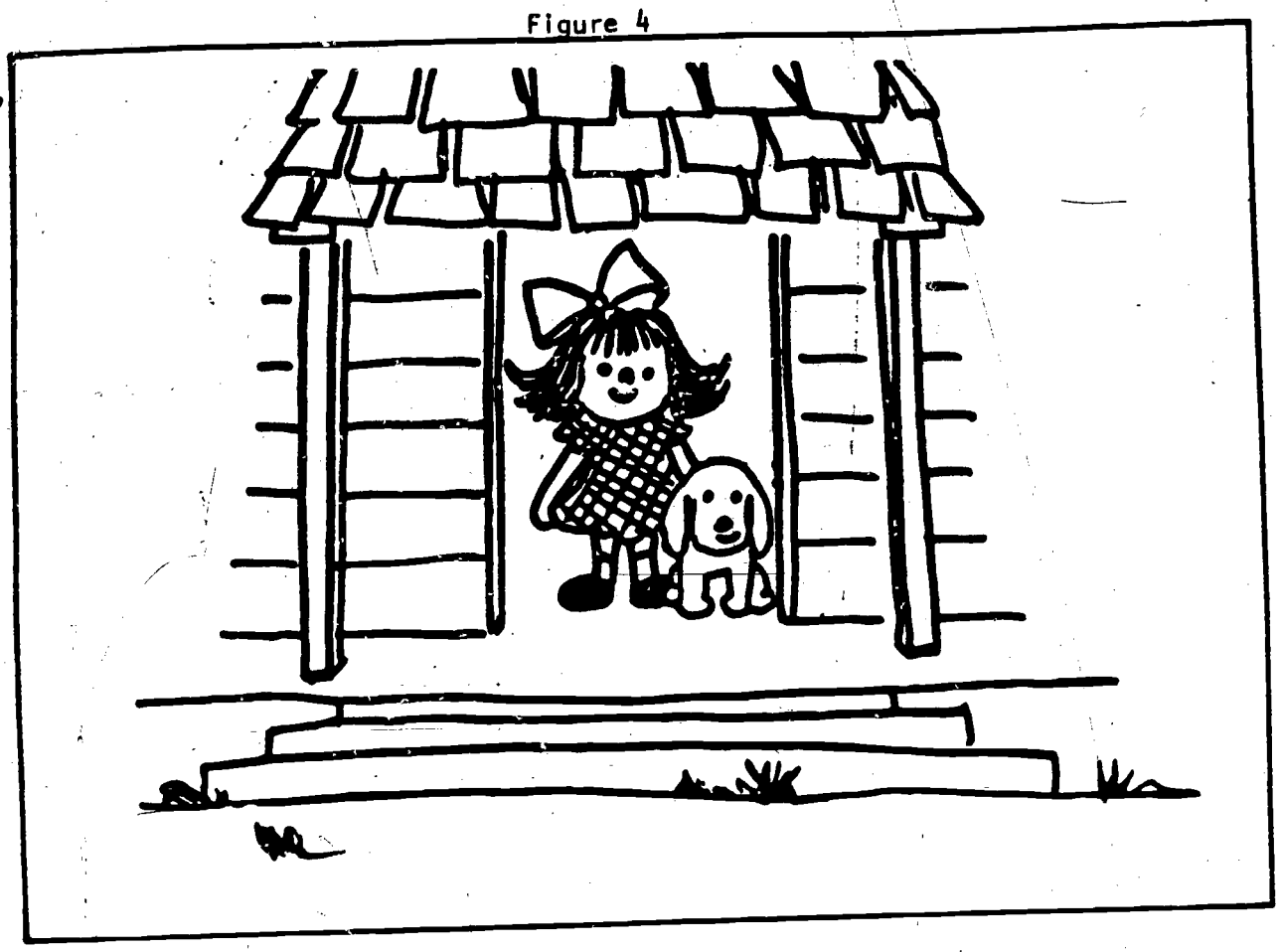


Figure 4

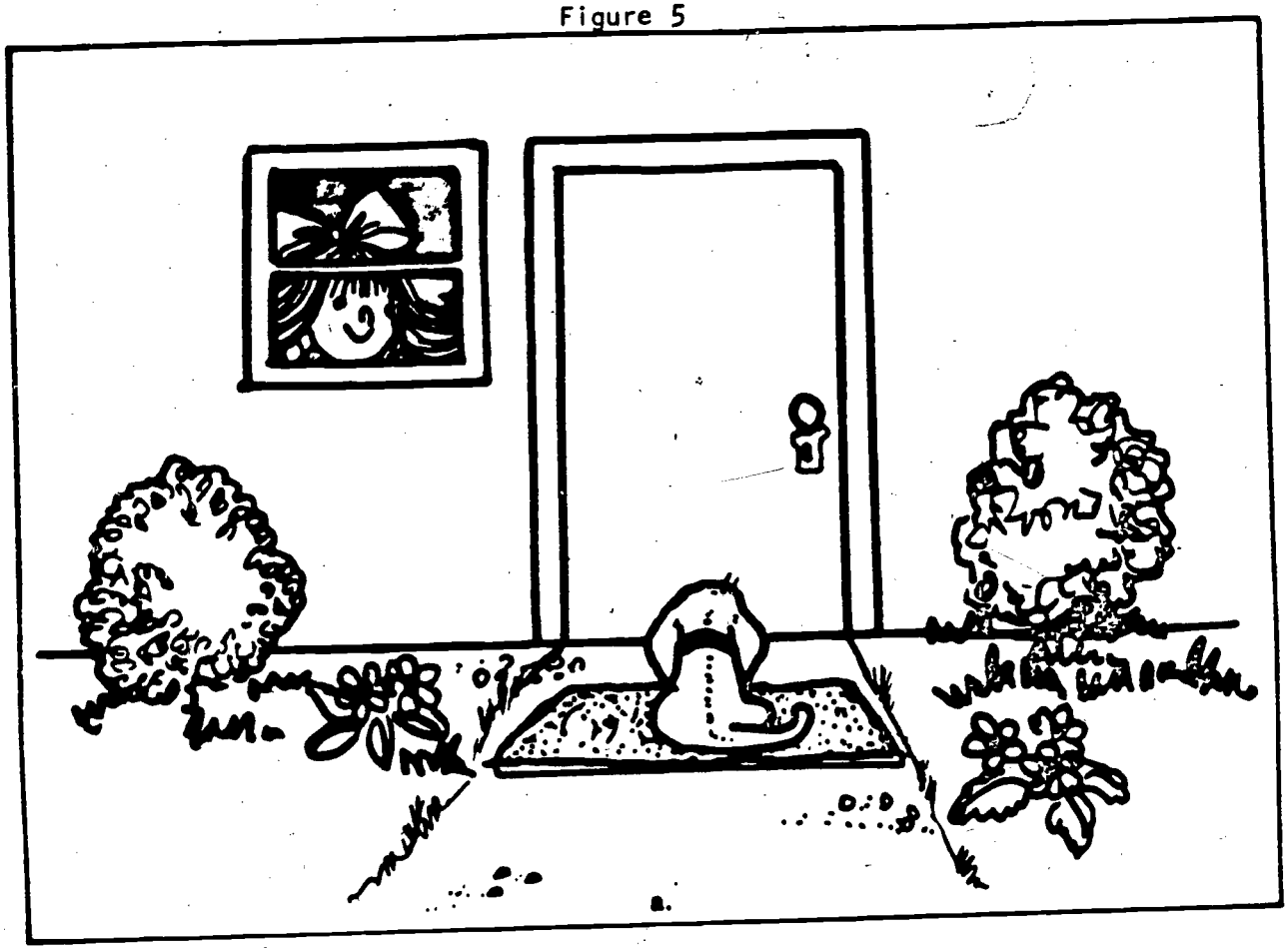
(7)

a

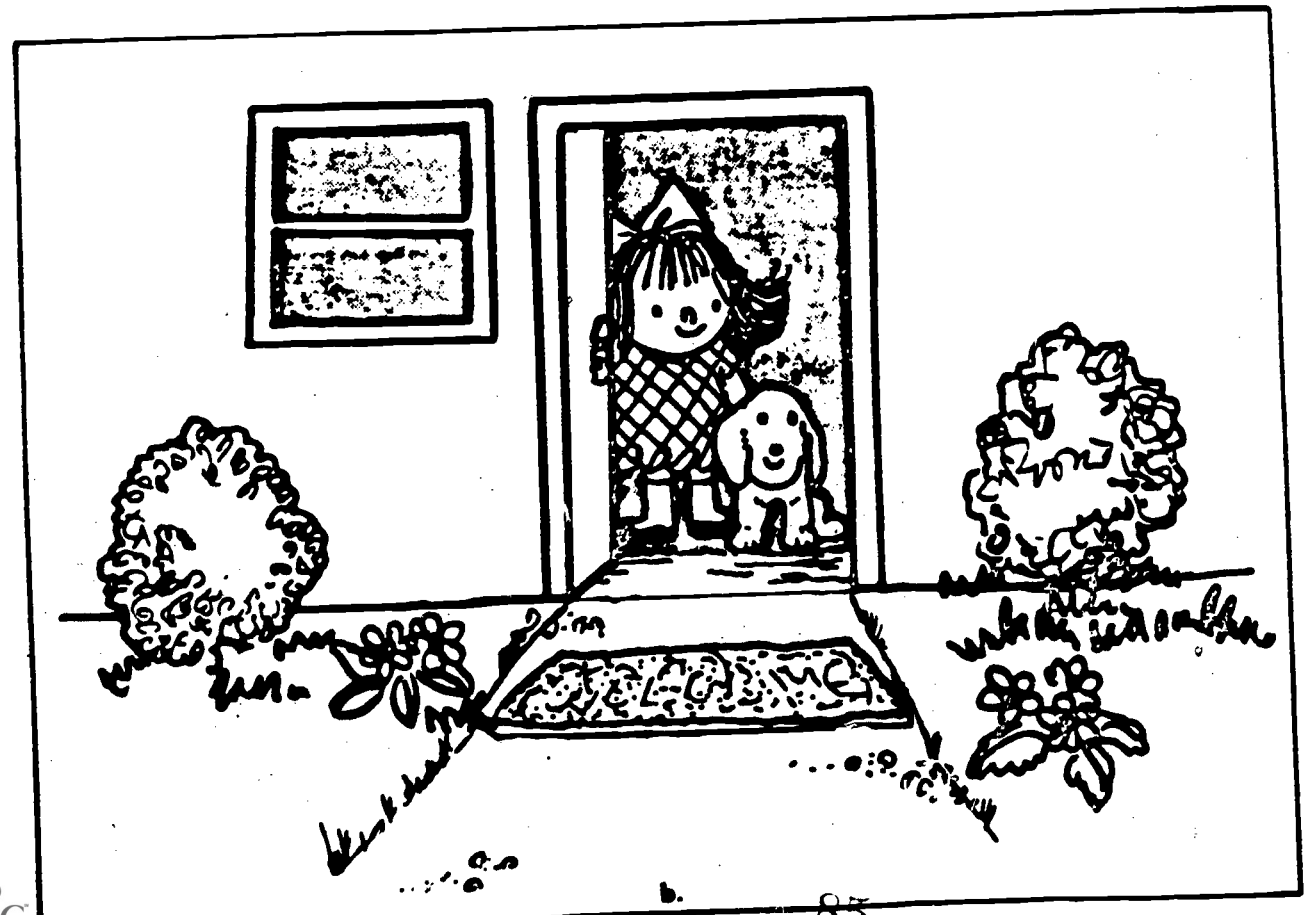


b

Figure 5



a.



b.

The recommendation I would make in order to avoid the kinds of problems pointed out in this example and in the preceding one is to test out the items with other adults as well as with children in order to find out whether you really are testing what you think you are.

4. Production Tasks in General. The trouble with trying to get children to produce what they know is that, in most cases, comprehension is being tested as well. It is very difficult to separate comprehension from production. For example, if you tell a child, "Write down the names of the objects you see in this room," the child may very well know the names of several objects in the room but has not understood the directive.

   Oral production tasks bring up another set of problems. How do you get a child to speak? If you are dealing with a shy child, if the child has not understood the task, or if the task appears silly or senseless to the child, the child may not respond at all.

To sum up the recommendations for avoiding the problems discussed above, in constructing a test one should:

1. Make tasks as simple as possible.
2. Use simple language and simple instructions.
3. Avoid using words which may be culturally biased--you may be testing a child who speaks a different dialect and who may not understand the word you are using.
4. Use aids whenever possible, such as pictures, drawings, tape-recorded sounds, etc.

Some Problems one Might Encounter in Administering a Language Proficiency Test

1. Instructions. Unless the instructions have been simply and clearly stated a child may not understand what is expected of him/her. Take the time to make sure the child understands the task to be performed.
2. Reticence. A child may hesitate in responding or may simply refuse to answer. Children from different cultures have different ways of interacting with teachers and adults in general. Don't assume that because a child does poorly on a task, that the child has language problems; the reason may not be language-related at all.

One very common type of task for testing sound discrimination is to show a child a picture of an abject (e.g., a ball) and ask, "Is this a pall"? If the child can hear the difference between "p" and "b," the correct response would be "No," or "No, it's a ball."

Do we assume that a child who answers "yes" has a language problem? Not necessarily. Maybe the child won't tell the tester when something is wrong because to do so would be considered rude in his or her culture. Maybe children aren't supposed to correct their elders. It is obvious that, in any testing situation, one must be very sensitive to these kinds of cultural differences.

3. Scoring. If you are not sure whether a response is correct or not, don't count it wrong. The child may have grown up in a different environment, and may know many things which you don't. A child, for example, who has grown up on a ranch probably knows much more about animals than you do--different names for different types of animals, different types of feed for each animal, etc. Furthermore, many words and expressions have different meanings in different cultures. People who have had contact with the Mexican culture know that a "taco" is a type of food made with a fried tortilla and stuffed with meat, lettuce, and tomatoes. In Argentina, however, and in other parts of South America, a "taco" is the heel of a shoe; in Spain, "taco" can mean "a bad word, an oath." When in doubt, suspend judgment until you can verify that the response is definitely incorrect.

Summary

When constructing items for a language proficiency test be aware of the problems involved. Don't assume that because you think you are testing for something in particular, you necessarily are doing so. Take care to make written instructions as simple and as clear as possible. Be aware of possible dialect differences and avoid using words that might have different meanings from those intended.

When administering language proficiency tests, be sure that the child understands the instructions and feels at ease. Don't count things

wrong if you are not certain that they are wrong; suspend judgement, if necessary, until you can verify the reporse. And lastly, remember that people have different accents and different ways of saying things, depending on the area in which they grew up and the dialect which they speak.

ADAPTATION OF ENGLISH PROFICIENCY INSTRUMENTS FOR KOREAN

NABE Pre-Conference Workshop

Kenneth K. Kim

April 19, 1980

ADAPTATION OF ENGLISH PROFICIENCY INSTRUMENTS FOR KOREAN

Kenneth K. Kim
NABE Preconference at NCBR, 4/19/80

According to California state Assembly Bills 1329 and 65, NES/LES students should be assessed in their primary language proficiency, because the students' primary language is to be used as the language of instruction in school.

Many test developers have produced matching English and Spanish language assessment instruments with the same rationale and similar test format. Frequently such matching versions have merits because the procedures of administration, the scoring, and the interpretation of the test results are the same or similar. One might also think that it is relatively easy to compare the results of the two tests to determine language dominance because of the similarity in test contents and methods. Perhaps all these considerations have prompted the trend to adapt existing English or Spanish proficiency tests for use in other languages.

However, adaptation of the existing English or Spanish tests for non-Indoeuropean languages is not as simple as it might be thought because of the differences in the linguistic and orthographic systems of the languages involved. The purpose of this paper is to discuss problems arising from attempts to adapt some existing language assessment instruments for use in assessing proficiency in Asian languages. It is not the concern of the presentation to assess the validity of the original versions of the instruments being discussed. The concern is rather whether the original versions can be adapted without seriously

distorting the rationale, the objectives, and the validity appropriately credited to the original instruments. Although the discussions in this paper will focus on Korean, the same kinds of problems are expected to exist in other Asian languages and also in many non-Indoeuropean languages.

The four tests to be discussed are BINL (Basic Inventory of Natural Language), BSM (Bilingual Syntax Measure), LAB (Language Assessment Battery), and LAS (Language Assessment Scales), all of which are approved by the State of California. The four tests are distinct from each other in a number of significant ways: (1) in the coverage of skills being tested; (2) method of response elicitation; (3) scoring; and (4) theoretical ground for the method and coverage. A brief description of the four tests is presented in Table 1. Since the types of problems are different from test to test due to the differences in contents and method of testing, each test will be discussed separately .

BINL (Basic Inventory of Natural Language)

This test uses pictures to prompt the student to give an oral response. This method can be equally effective for any language. The major problem in adapting BINL for other languages is mainly in the analysis of the students' speech samples for the purpose of scoring. The BINL scoring system refers to linguistic units such as word, phrase, clause, etc., as indicated on the sample BINL Scores Sheet in Table 2. The fluency score in column C on the left hand side is the number of

Table 1

Test Description

Tests	BINL (Basic Inventory of Natural Language)	BSM (Bilingual Syntax Measure)	LAB (Language Assessment Battery)	LAS (Language Assessment Scales)
Rationale & Emphasis	Linguistic complexity of oral samples collected in naturalistic speech settings best reflects children's language proficiency.	Acquisition of morphological and syntactic items best reflects degree of language proficiency.	Language proficiency should be tested in four skill areas listening, speaking, reading, and writing.	Language proficiency test should include items from all linguistic skills categories including discourse.
Skills & Methods	Picture-elicited oral production.	Elicit specific morphological and syntactic forms through structured conversation using cartoon type pictures.	Listening/speaking: your name, how old, body part naming, common object naming, sentence-picture matching. Reading: picture-word matching. Writing: picture-word matching, grammar, writing mechanics.	Phoneme discrimination: minimal pairs (same-different). Vocabulary: picture naming. Phoneme production: word/sentence repetition Listening comprehension: point to correct picture of given sentence. Story retelling: after hearing recorded story retell it or answer probe questions.
Levels	K-adult	Level I: K-2 Level II: 3-12	Level I: K-2 Level II: 3-6 Level III: 7-12	Level I: K-5 Level II: 6-12
Testing Time	5 min./child	10 min./child	Level I: 5-10 min./child Level II & III: 41 min.	20 min./child
Scoring	Counting linguistic units, such as words, phrases, clauses. Assign different weights to different units. Computation of counts and points to obtain proficiency score.	Grammatical analysis of response. Count correct responses.	Count correct answers.	Count correct answers.. Teacher's subjective rating of children's discourse production. Computation of conversed scores.

Table 2

Sample BINL Scores Sheet

INDIVIDUAL ORAL SCORES SHEET (IOS)

BASIC INVENTORY OF NATURAL LANGUAGE (BINL)

Student's Name Elena Hernandez

Grade 1 Teacher Villa

Language English

[X] PRE TEST 9-24-76 DATE [ ] POST TEST DATE

I. Simplified Scoring

	A. Complete Sentences	B. Partial Sentences	C. Fluency	Samples	Grammar Analysis Columns: Partial Sentence (1)	Independent Clause (10)	Modifiers (10)	Phrase (20)	Dependent Clause (30)	Sub-Total
1		✓	1	boats	1					1
2		✓	2	red boats	1		10			11
3		✓	3	with two boys	1		10	20		31
4	✓		3	because he does		10			30	40
5		✓	4	boys to play ball	1			20		21
6	✓		3	I see a theca		10	10			20
7	✓		5	The ball -- (me - uh) I like it		10	10			20
8	✓		3	He brings it		10				10
9	✓		3	You want to		10		20		30
10	✓		3	(and) I always go		10	10			20
	6	4	30	Totals						204

A B C

Total Index of Language Ability

II. Syntactic Analysis (Hand Scoring)

30 Total number of words: (Total of Column C)

10 Total number of samples: (Enter sum of ✓ marks in Columns A & B)

3.0 Average Sentence Length: (Divide total number of words by total number of sentences)

20.4 Average Level of Complexity: (Divide Index of Language Ability (above right) by Total number of sentences)

RAPID SCORING INSTRUCTIONS (Machine Scoring)

1. Transcribe the language samples, one sample per line
2. Number the samples 1 - 10 (minimum)
3. List the number of words per sample on the same line in Column C (Fluency)
4. Group the pages by language, class, school and forward to CHECpoint Systems for machine scoring

NUMBER OF:

PARTIAL SENTENCE	4
INDEPENDENT CLAUSES	6
MODIFIERS	5
PHRASES	3
DEPENDENT CLAUSES	1

CHECpoint Systems

words in the sample transcribed in the same row. This fluency score along with the counts of other units is used to compute the proficiency level score.

In English and Spanish, the words are simply orthographic units except for a few contracted forms. Thus, any teacher who can read and write these languages can count the number of words in the samples. Such an easy orthographic clue is not apparent in Korean. In Korean, an orthographic syllable, which is also a phonological syllable in most cases, is a cluster of a few letters. Each orthographic syllable in a word containing more than one syllable can be readily distinguished from other syllables by a short but recognizable vertical space between them, and individual words and phrases in a sentence are separated from other adjacent words or phrases by a slightly wider space between them as illustrated below.

(1) a. 밥, 국, 그리고 생선. (4 words)

b. Rice, soup, and fish. (4 words)

One might notice the parallelism in the Korean and English examples above: that is, words are those units that are separated from others by a clearly recognizable space. However, the following example shows that the situation is not so simple.

(2) a. 존이 그 사과를 먹었다. (4 words)
John-S the apple-O ate.

b. John ate the apple. (4 words)

The two nouns in the above Korean example, that is, 'John' and 'apple,' are immediately followed by the subject and object case marking particles

as indicated by S and O respectively. Should these particles be counted despite the lack of orthographic marking? The following example seems to suggest the answer.

(3) a. 잡 히 시 었 겠 다.
a b c d e f

b. '(Someone superior to the speaker) must have been caught!'
c e d a f b

Although all the seven syllables in (3)a are written without any intervening space of the kind shown in (1) and (2), the string is highly complex as illustrated in (3)b. This string is composed of (a) a verb root /잡/, (b) passive voice infix /히/, (c) honorific infix /시/, (d) past tense infix /었/, (e) assumptive infix /겠/, and (f) exclamatory sentence ending in the low level of speech /다/. Most of these items would have to be expressed by independent words or phrases in English as illustrated in (3)b. The example clearly shows that the number of orthographic spaces is not a measure of the grammatical complexity of the string, nor does it reflect the language proficiency of the speaker. For these reasons, the root, the infixes, and the ending should be counted separately.

Likewise, the noun particles in Korean as in (2) frequently play a role comparable to word order in English which determines the grammatical function of the words. Since the functions of nouns are explicitly indicated by the particles in Korean, the order, for example, between a subject and an object can be freely interchanged without affecting the meaning or grammaticality of the sentence. Considering

the important function of the particles, it seems to be only fair to count them as separate items.

The difference in the units of counting may require redefinition of other larger linguistic units such as phrase and clause. The difference in the units is also likely to cause different ratios between the various units being counted (e.g., number of words or phrases within a sentence). All these, in turn, may require different interpretation of the score on Korean tests.

In principle, procedures for analysis of the speech samples and for the computation of the scores can be developed. But whether the length of a sentence or the number of phrases and clauses within a sentence is an appropriate measure of language proficiency in Korean is another matter. Without further studies on the relationship between language proficiency and scores computed on the basis of the number of various linguistic units in Korean, a Korean version of BINL may be merely a mental exercise, the singnificance of which is seriously in doubt.

BSM (Bilingual Syntax Measure)

Some second language acquisition researchers share a hypothesis that there are universal cognitive mechanisms that govern the language acquisition processes. Dulay and Burt (1974) and Bailey, Madden, and Krashen (1974) tested the hypothesis by examining the acquisition order of so-called English "functors" in learners of English as a second language. Functors are the function words that have only a minor role in conveying the meaning of a sentence, such as inflections of nouns

(John - John's); verbs (jump, jumps, jumped), articles, auxiliaries (is going), copulas (He is), and prepositions. These functors can be easily elicited independently of the topic of conversation, and it is easy to determine whether or not they are used correctly. The studies revealed that the acquisition order of the English functors was largely the same regardless of the different first language background of the learners, including the native speakers of English.

The BSM is based on the research findings about the acquisition of functors. If a student has acquired only the easiest functors, he or she is assigned a low level of proficiency. If the student has acquired functors higher in the hierarchical order, he or she is assigned an appropriately higher level.

Now, what about a BSM for Korean language which, as we have seen, has a linguistic system drastically different from English or Spanish? Following are some factors that determine the acquisition order or complexity hierarchy among the functors.

(4) Factors affecting the acquisition hierarchy of the functors:

a. Concept complexity: e.g., simple vs. complex tenses
simple indicative vs. complex aspects

b. Regularity in morphological or syntactic variation:
e.g., regular vs. irregular past tense

c. Frequency of usage: e.g., progressive tense, pronouns, articles, prepositions

All these factors vary from language to language. Table 3 shows English and Spanish functors used in BSM along with the information about whether there is an equivalent grammatical item in Korean for these

Table 3

Structures Tested by BSM and Korean Equivalents

Structures	English	Spanish	Korean
Word order	yes	yes	yes
Pronoun cases	yes	yes	yes (case particles)
Progressive	yes	yes	yes
Copula	yes	yes	yes
Short/long plural	yes	no	no
Auxiliary (be)	yes	yes (estar)	no
Articles	yes	yes	? (usage different)
Regular/ irreg. Past	yes	---	yes
Present indicative	yes	yes	no
Possessive	yes	yes	? (identical with the case particles above)
Conditional aux.	yes	---	no
Perfect aux.	yes	---	no
Present subjunctive	---	yes	no
Past subjunctive	---	yes	no
Past participle	yes	---	no
Infinitive	---	yes	? (usages different)
Adjective gender	no	yes	no
Conjunction que	---	yes	yes

Yes: There is an equivalent grammatical form that can be tested.
No: There is no equivalent.
?: There is an equivalent but it may not be tested for the indicated reasons.
---: Not tested.

functors, and if there is, whether acquisition of the functor can be used as a measure of proficiency in Korean. As shown in this table, there do not seem to be enough grammatical structures that can be used to assess proficiency in Korean. The conclusion is then that the adaptation of BSM for Korean language requires a complete redesign of the test based on the acquisition hierarchy among independently selected Korean functors.

LAB (Language Assessment Battery)

The LAB includes three sets of tests for three different grade levels, each of which includes four subtests for listening, speaking, reading, and writing. Level I questions are mostly general questions testing comprehension of simple sentences, naming of common objects, letter and word recognition, and writing alphabet letters, words, and simple sentences. No specific target grammatical structures are tested. Many questions thus can be adapted to Korean without affecting the objectives of the original test questions.

Levels II and III contain, in addition to general questions like those in Level I, questions designed to test specific sound contrast and grammatical items.

Some problems that are likely to be encountered in the development of a Korean version of the test are as follows:

(1) There are three different levels of tests. What are the vocabulary items appropriate to each level? Information is needed about the concept difficulty, and reading and spelling difficulty of words that children of different ages can cope with. The test developer

should develop a Korean lexicon for this purpose or at least be familiar with the principles to be considered in the selection of words.

(2) Which sounds are easy or difficult to discriminate and which sounds are easy or difficult to produce? To answer this question, one should be thoroughly familiar with the methods of contrastive analysis and error analysis. Except for those items testing specific English grammatical structures, most items can be easily adapted to Korean due to the generality of the objectives of the items.

LAS (Language Assessment Scales)

The LAS assesses proficiency in five subskills; (1) discrimination of speech sounds, (2) naming common objects, (3) target phoneme production in words and sentences, (4) comprehension which is tested by matching a sentence played by a tape-recorder with the appropriate picture depicting the meaning of the sentence, and (5) listening to a recorded story and retelling the story.

The test has only two pages and, except for the story retelling task, the procedures of administering and scoring the test are simple and straightforward. If the concept and structural complexity of the words and sentences presented in the test can be controlled in an appropriate manner, the test can be relatively easily adapted to any language without changing the format or type of contents being tested. The criteria and procedures for the selection of the sound pairs, words, and sentence structures in the original test are not stated clearly anywhere. Nevertheless, they are easy to infer because of the conventional test format and the straightforwardness of the test objectives. In short, this test seems to be the easiest to adapt.

To conclude, the first two tests, that is, BINL and BSM, require further research on the relationship between language proficiency and the various linguistic units in Korean. Any imitation version without research evidence would seriously lack validity. On the other hand, the remaining two, that is, LAB and LAS, could be adapted with little difficulty due to the common conventional testing method and the greater generality of the test items.

At any rate, should there be any pressing need for an instrument to assess a new language and should the decision be made to adapt an existing instrument for the new language, it is very important that each of the original test items be analyzed and evaluated in terms of the specific objective of the test item to determine the feasibility of the adaptation. Furthermore, every effort should be made to incorporate linguistic, psychometric and cultural considerations from people with appropriate training and background.

WORKSHOP EVALUATION

In order to evaluate the workshop, a questionnaire was distributed after the final presentation. Of the seventeen participants, fourteen completed the questionnaire. Six questions were asked, four of them forced-choice and two open-ended. The questions were designed to seek information about workshop effectiveness and usefulness and to provide opportunity for the participants to suggest ways to improve the workshop. The evaluation form is reproduced on the following pages and an analysis to each question follows.

NCBR/SWRL
NABE Workshop
April 19, 1980

WORKSHOP EVALUATION

We need your frank and constructive feedback to be able to plan other Workshops. Please take a few minutes to fill out this form. Thank you.

1. This Workshop helped me by:

_____synthesizing and organizing ideas
_____reminding me of ideas I have neglected
_____clarifying some of my ideas
_____confirming my ideas and techniques
_____changing my ideas
_____redefining some ideas I had to make them more valuable
_____giving me new ideas, i.e., more approaches than I was aware of before
_____providing increased understanding of new approaches
_____giving me ideas I can put to use
_____motivating me to find out more
_____motivating me to try out some ideas
_____other_____

2. In terms of practicality, the ideas discussed were:

_____very practical (easy to use and in tune with schools)
_____practical (able to be used with a minimum of adjustment)
_____somewhat practical (able to be used with change and considerable effort)
_____impractical (would require too much change and effort to be rewarding)

3. Do you feel this was a good use of your time?

_____definitely yes
_____yes
_____somewhat
_____no
_____definitely not

4. Here are some ways to improve this Workshop:

(Brainstorming)

(Sharing ideas)

(Presentations and demonstrations)

5. Overall I would rate the Workshop

_____excellent
_____very good
_____good
_____fair
_____poor

6. Comments:

Question #1

The participants were asked to evaluate the usefulness of the workshop by checking the statements which described the ways in which the workshop was of benefit to them. Two of the participants checked all of the options in this question. Of the remaining participants, seven responded that the workshop was useful in "synthesizing and organizing ideas." The next most frequent response (six) was in "clarifying some of my ideas" and in "providing increased understanding of new approaches." Participants also responded highly on "reminding me of ideas I had neglected"; "redefining some ideas I had to make them more valuable"; and on "motivating me to fine out more" (five each). Of the participants who did not check all the options, four expressed that the workshop was helpful in "confirming my ideas and techniques" and in "giving me ideas I can put to use." The options which received the lowest responses were "changing my ideas"; "giving me new ideas, i.e., more approaches than I was aware of before"; and "motivating me to try out some ideas" (three each). None of the participants wrote in the space provided for comments. Participant responses and the response totals are presented in Table 1.

Table 1

Ways Workshop was Useful

Participant	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Total
Synthesizing and organizing ideas	x	x	x	x	x					x	x		x	x	9
Reminding me of ideas I had neglected	x	x	x		x			x		x			x	x	7
Clarifying some of my ideas	x	x	x		x		x			x			x	x	8
Confirming my ideas and techniques	x	x	x		x									x	6
Changing my ideas	x	x												x	3
Redefining some ideas I had	x	x		x		x	x		x				x		7
Giving me new ideas	x	x									x				3
Providing increased understanding	x	x	x					x	x	x		x		x	8
Giving me ideas I can put to use	x	x						x		x		x		x	6
Motivating me to find out more	x	x	x						x		x		x	x	7
Motivating me to try out some ideas	x	x											x		3
Other															0

Question #2

In this forced-choice question, participants were asked to indicate if the ideas discussed were very practical, practical, somewhat practical, or impractical. Six out of the fourteen responded that the ideas were very practical (easy to use and in tune with schools). Four rated the ideas discussed as somewhat practical (able to be used with change and considerable effort), and three responded that the ideas were practical (able to be used with a minimum of adjustment). One participant did not respond to this question at all. Table 2 gives the responses and the totals of each response.

Table 2

Responses To Rating Question

Participants	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Total
Very Practical	x	x	x		x	x						x			6
Practical								x					x	x	3
Somewhat Practical							x		x	x	x				4
Impractical															0
No Response				x											1

Question #3

In this forced-choice question participants were asked whether they felt that the workshop was a good use of their time or not. Seven of the fourteen participants responded with a "definitely yes" that the workshop was a good use of their time. Five more participants responded "yes," and the rest responded "somewhat." None of the participants indicated that the workshop was not a good use of their time. Table 3 gives the responses by participant and the totals.

Table 3

Workshop Was Good Use Of Time

Participant	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Total
Definitely "Yes"	x		x			x				x	x	x		x	7
Yes		x		x	x			x	x						5
Somewhat							x						x		2
No															0
Definitely "No"															0

Question #4

This question gave the participants an opportunity to suggest ways to improve the workshop. Nine participants chose to respond to this open-ended question. The comments ranged from "No need to change format" to suggestions for improving the Workshop by changing the format, to something such as "Panel discussion format." Other suggestions recommended more participation by the participants, such as actual test administration or test construction exercises. The responses are quoted verbatim below:

- No need to change format
- More participation by participants
- More background into research
- Not enough time
- Provide more actual examples; more absolute solutions
- Panel discussion format
- Working through a problem. Experience actual development of simple project
- Discuss strengths and weaknesses of specific tests regarding criteria discussed
- Should be in various languages
- Provide participants opportunity to give the tests
- Valid and reliable assessment instruments
- Presentations on English literacy could have drawn clearer discussion and conclusions regarding bilingual pupil assessment

Question #5

In this forced-choice question participants were asked to select an overall rating for the Workshop. Five participants rated the Workshop "Good." Four participants rated it as "excellent" and another four as "Very good." A rating of "Fair" was given by one participant. The participant responses to the question are given in Table 4.

Table 4

Workshop Rating

Participants	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Total
Excellent	x									x	x			x	4
Very good		x	x		x	x									4
Good				x			x	x	x			x			5
Fair													x		1
Poor															0

Question #6

Finally, participants were given space to make additional comments about the Workshop. Seven participants responded. Each response is given below verbatim:

- It was well-planned and most helpful.
- Investigation should be done to find out what kind of language assessment tools are being used in Mexico and the rest of Latin America.
- Very good. Pasé muy buena experiencia. Espero que sigamos teniendo estos workshops.
- Workshop was helpful to me to the extent that I will be scrutinizing the assessment instruments I use with my children. Presenters should have given us a list of test instruments they consider better or most appropriate for language proficiency determination.
- It appears that bilingual testing is still in its infancy.
- All presentat-ons were well prepared, clear, short, and to the point. Thank you for the excellent service!
- Research is of course necessary, however, should lead to the development of practical assessment instruments. These instruments are what is now needed to accurately assess NES/LES students.