DOCUMENT RESUME

ED 349 815 FL 020 634

AUTHOR	Damico, Jack S.
TITLE	Performance Assessment of Language Minority Students.
PUB DATE	Aug 92
NOTE	50p.; In: Focus on Evaluation and Measurement. Volumes 1 and 2. Proceedings of the National Research Symposium on Limited English Proficient Student Issues (2nd, Washington, DC, September 4-6, 1991); see FL 020 630.
PUB TYPE	Viewpoints (Opinion/Position Papers, Essays, etc.) (120)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Bilingual Education Programs; Elementary Secondary Education; Language Proficiency; Language Research; *Limited English Speaking; Researchers; *Student Evaluation; Test Construction; *Testing; Test Selection

ABSTRACT

Performance assessment of language minority students is a complex process that requires the application of theoretically defensible procedures that are carefully designed and systematically implemented. Due to the differences between language minority students in the schools and those English-as-a-Second-Language/English-as-a-foreign-language students typically studied by language testing researchers, performance assessment in the schools must involve utilization of procedures that are more authentic, more functional, more descriptive, and more individualized than those typically recommended by second language testing researchers. This paper proposes a descriptive approach to performance assessment that is theoretically defensible and psychometrically sufficient. The characteristics necessary for successful performance assessment, the assessment process, and actual assessment techniques are discussed. Responses to the paper by J. Michael O'Malley and Cecilia Naverette are appended. (VWL)

Performance Assessment of Language Minority Students

Jack S. Damico
University of Southwestern Louisiana

Introduction

Performance assessment of language minority students is a complex process that requires the application of theoretically defensible procedures that are carefully designed and systematically implemented. Due to the differences between language minority students in the schools and those ESL/EFL students typically studied by language testing researchers, performance assessment in the schools must involve the utilization of procedures that are more authentic, more functional, more descriptive, and more individualized than those typically recommended by second language testing researchers. This paper proposes a descriptive approach to performance assessment that is theoretically defensible and psychometrically sufficient. The characteristics necessary for successful performance assessment, the assessment process, and actual assessment techniques are discussed.

> "To me it seems to be generally desirable in instructional contexts to focus tests diagnostically only against a contextual backdrop where attention is directed toward comprehending or producing meaningful sequences of elements in the language" (Oller, 1983:354).

When I first read this passage in a working manuscript of John Oller's 1983 chapter, "A consensus for the eighties?", it had a galvanizing effect on me. This suggestion of a "pragmatic" approach to assessment was the final impetus for me to shift my theoretical stance and my practices involving performance assessment of language minority students in the schools. I recognized that although I was very interested in the excellent work going on in second language testing research, my research involving language minority students in the schools required something different.

While there are a number of purposes for evaluation and assessment of language minority students in the schools (Henning, 1987), those purposes most relevant to my concerns revolved around the student and the student's ability. Using Oller's conceptual writings as guidance, my work has focused on ways to provide a rich description of the individual student's communicative performance and that

student's underlying language proficiency. To accomplish these objectives, the assessment procedures designed and implemented for these students have to be more **authentic**, more **functional**, more **descriptive**, and more **individualized** than those procedures recommended in the second language testing literature. This is due to the differences between the language minority students my research targeted and the ESL/EFL students discussed by many other language testing researchers.

The Differences Between The Populations

The second language testing literature typically focuses on students who are enrolled in ESL or EFL classes. These students are tested to determine placement in second language classrooms or to determine their progress in these classes. With much of this research, the students are usually older than elementary level and/or there is an assumption that these students have normal language-learning abilities. The situation facing the typical language minority student in our public schools -- particularly at the elementary and mid-school level -- is very different.

Unlike the majority of students discussed in the second language testing research literature, the language minority students usually targeted for evaluation in the public schools are located in environments that are not conducive to language diversity. They are frequently the only students in their classrooms who are non-English speakers. Additionally, their teachers are unlikely to have knowledge of their first language or even of the process of second language acquisition. As a result, normal acquisitional phenomena may be viewed as an indication of language-learning problems (Hamayan & Damico, 1991). Within this environment there may even be prejudice toward students who are speakers of other languages (Ogbu, 1978).

Second, unlike older ESL/EFL students, the language minority students are typically compared in their routine academic and conversational performance with the mainstream students rather than other language minority students. In such cases, they naturally perform more poorly since they don't have the same proficiency in English and, because they are not being tested and compared in the ESL or EFL classroom with their peers, their performances are always suspect when they perform below the mainstream expectations (Cummins, 1984; Ortiz & Wilkinson, 1987).

Third, many of these language minority students come from what Ogbu (1978) has termed the "caste minority" group. That is, a

group of individuals that may or may not have been born in this country but who are usually regarded by the mainstream and dominant population as being inferior. As a result, these individuals are **perceived** as being less-intelligent, less-motivated, and less-able to match the mainstream students in a range of activities. These biased perceptions give rise to lowered expectations and frequently result in a disempowerment of these students that is manifested not only in their academic performances but in the ways that they perform and are evaluated during assessment (Cummins, 1986; 1989; Mercer, 1984; Ogbu, 1978).

The fourth difference between these students only heightens the perceptions created by the first three differences: these students are usually individuals that do not have well-documented first language proficiency. Rather than having demonstrated their first language through performances that enabled them to enter into a second language learning context (e.g., an EFL class), many language minority students are kindergartners or first graders who are still acquiring their first language. Consequently, little is known about their native language proficiency. Even when these students are older, they are frequently recent immigrants that have few academic records from their home countries that could document their performances (Cloud, 1991). As a result, there is no assumption of sufficient first language proficiency or normal language-learning capacity. Rather, given the first three conditions, these students may be suspected of poor language proficiency in both their first and second languages. That is, they may be suspected of exhibiting a language-learning impairment.

The final difference is a natural consequence of this suspicion and it makes the situation for many language minority students most desperate: the purposes of assessment are frequently different. Unlike the ESL/EFL students who are assessed for placement in classes to supplement their normal language-learning proficiency with the addition of a second language, language minority students may be assessed to determine remedial placements or placement within special education programs. A harsh reality in our public schools is that many language minority students are mis-diagnosed and enrolled in special education programs or remedial tracks that reduce their academic potential as normal language-learners (Cummins, 1984; Fradd, 1987; Oakes, 1985; Ortiz & Yates, 1983).

As a result of these differences, performance assessment for many language minority students requires a different focus. Not only must the usual testing purposes be accomplished (O'Malley, 1989), the evaluator must also be able to address specific questions regarding an individual's underlying language proficiency and learning potential. In the remainder of this paper, a descriptive approach with a pragmatic focus that has been effective in the performance assessment of language minority students will be detailed. Although

this descriptive approach is aimed at actual evaluation and diagnosis of individual students for selection and placement purposes, it can also be utilized in program evaluation for formative and summative evaluation purposes (Navarrete, Wilde, Nelson, Martinez, & Hargett, 1990).

Descriptive Performance Assessment

To adequately evaluate language minority students in the schools, performance assessment practices must be consistent with the currently accepted theoretical construct of language proficiency (Bachman, 1990a; Oller, 1989; Oller & Damico, 1991) and they must be carefully designed to meet the numerous assessment requirements within the public school environment. The theoretical grounding will help ensure strong validity indices while the design characteristics will aid in the construction of assessment procedures that possess sufficient reliability and educational utility.

The Construct of Language Proficiency

Since 1961 when John B. Carroll raised concerns regarding the artificialness of the "discrete-point" testing methodology, numerous second language researchers have advocated assessment and evaluation procedures that stress the interrelatedness of language as a psychological construct. Calling for the development of "integrative" (Briere, 1973; Oller, 1972; Spolsky, 1973; Upshur, 1971), "pragmatic" (Oller, 1979; 1983), "edumetric" (Cziko, 1983; Hudson & Lynch, 1984), "communicative" (Bachman, 1990a; Canale, 1987; 1988; Olstain & Blum-Kulka, 1985; Shohamy, 1991; Van Lier, 1989; Wesche, 1987) and "informal" (Brindley, 1986; Navarrete, et al., 1990) assessment procedures, these researchers utilized the most defensible constructs of language proficiency available to them to justify their test design. Descriptive performance assessment is also based on the most defensible construct of language proficiency available.

Currently, language proficiency is viewed as a componentially complex psychological construct with a powerful synergistic quality that enables language or communicative ability to act as a coherent and integrated totality when it is manifested in performance (Bachman, 1990b; Carroll, 1983; Oller, 1983; 1989). While there are several models or frameworks that are consistent with this construct (e.g., Bachman, 1990a; Canale & Swain, 1980; Cummins, 1984; Shohamy, 1988), the hierarchical model of language proficiency proposed by Oller (1983; 1989; Oller & Damico, 1991), seems most appropriate and is utilized for the design of this descriptive approach to performance assessment.

In this hierarchical model, language proficiency is recognized as a multicomponential and generative semiotic system that functions in an integrated fashion in most communicative contexts. For practical purposes, language proficiency exist when the individual components (e.g., syntax, morphology, phonology, lexicon) function as an integrated whole. Further, in keeping with the synergistic perspective, this integrated whole is unpredicted by the behavior of the individual components when they are described separately (Fuller, 1982). This is because these separable components are more apparent than real; they are essentially terminological distinctions created in the mind of the linguist or evaluator for ease of discussion and analysis. As noted previously, language is a generative system that exists for the transmission and coding of meaning and these components are aspects of this process. However, they are not divisible and discrete in their functioning; they function holistically. Consequently, when observable aspects of these components are isolated in artificial tasks during assessment, the tasks are not assessing language or communication but some splinter skill quite different from true language proficiency.

Another facet of this synergistic quality of the hierarchical model is that language proficiency is not viewed as an autonomous semiotic system. It is an integrated system that is intimately influenced by other semiotic and cognitive systems and by extraneous variables. Consequently, performance is highly influenced by factors like memory, perception, culture, motivation, fatigue, experience, anxiety, and learning. For effective and valid assessment, therefore, it is important that language and communicative behavior be assessed in natural contexts.

As discussed elsewhere (Damico, 1991a; Oller & Damico, 1991), reliance on Oller's hierarchical model results in a number of advantages when trying to account for data and concepts reported in the second language literature. Such discussion, however, is beyond the scope of this paper. For our current purposes, this model helps in the design of an effective descriptive performance assessment system.

The Design Characteristics of Descriptive Assessment

Based on the hierarchical model regarding the construct of language proficiency and on the specific purposes for performance assessment previously discussed, there are several design characteristics necessary to the development of effective procedures for descriptive performance assessment of language minority students (Damico, 1991a; Damico, Secord & Wiig, 1991). These characteristics concern the authenticity of the data collected and analyzed, the functionality of the behaviors evaluated, a sufficiently rich description of language

proficiency to accomplish assessment objectives, the necessity for a focus on each individual being assessed, and the assurance of psychometric veracity (See Table 1).

Table 1
Design Characteristics Necessary for Descriptive Assessment

Authenticity of the Collected Data
- Linguistic Realism
- Ecological Validity

Focus on Functionality
- Effectiveness of Meaning Transmission
- Fluency of Meaning Transmission
- Appropriateness of Meaning Transmission

Rich Description of Language Proficiency
- Descriptive Analysis
- Explanatory Analysis

Emphasis on the Individual

Assurance of Psychometric Veracity
- Reliability
- Validity
- Educational and Programmatic Utility

Authenticity

The first characteristic necessary to effective descriptive assessment relates directly to the synergistic quality of language proficiency. Since language proficiency is synergistic in terms of its internalized structure, its relationship with other semiotic and cognitive abilities, and its interaction with external variables, it is not possible to assess language and communication apart from the influence of intrinsic cognitive factors and extrinsic contextual features. Consequently, assessment must be structured to observe language during actual communicative activities within real contexts. The language and communication behaviors assessed must be authentic (Damico, 1991a; Oller, 1979; Seliger, 1982; Shohamy & Reves, 1982).

For our purposes, authenticity means that the methods used in assessment focus on data that possess linguistic realism and ecological validity (see Table 1). Linguistic realism requires that assessment procedures treat linguistic behavior as a complex and synergistic phenomenon that exists primarily for the transmission and interpretation of meaning (Crystal, 1987; Oller, 1979; 1983; 1989; Shuy,

1981). The task of assessment, therefore, is to collect data that are meaning-based and integrative rather than data that attempt to fragment language or communication into discrete points or components. Consequently, the data of interest in assessment should be actual utterances and other meaningful "chunks" of linguistic behavior that serve to transmit an idea or intention from a speaker or writer to a listener or reader (Damico, Secord, & Wiig, 1991). In this regard, there should be a focus on the analysis of discourse.

Ecological validity is another aspect of authenticity that must be accommodated during assessment. Numerous researchers have demonstrated that the behaviors manifested during isolated and artificial testing procedures are unlike the linguistic and communicative behaviors noted in real language usage situations (Carroll, 1961; Cummins, 1984; Douglas & Selinker, 1985; Shohamy, 1983). Rather than trying to isolate the assessment process from contextual influence, therefore, assessment should be accomplished in naturalistic settings where true communicative performance is occurring and is influenced by contextual factors. Such practices will enable the evaluator to discern the effects of contextual variables on the student's communicative performance and will enable assessment to remain consistent with the emphasis on relativism in behavior analysis (Kagan, 1967; Oller, 1979).

Oller's (1979) work regarding the development of pragmatic assessment procedures is consistent with this focus on authenticity. By incorporating the work of several other language testing researchers, Oller suggested that language testing adhere to three "pragmatic criteria." The first, that data be meaning-based, required that data be collected from tasks that were motivated by a desire to transmit meaning or achieve comprehensibility during meaning transmission. The second, that data be contextually-embedded, required that the data under scrutiny be produced in a contextually rich environment. The third criterion, that data be temporally-constrained, required that tasks used to collect the data fit into the normal temporal envelope of communicative interaction. Taken together, these three pragmatic constraints act to ensure linguistic realism and ecological validity.

Functionality

The second characteristic necessary for effective descriptive assessment is concerned with how performance is evaluated or, put in more operational terms, what should be measured. In the second language testing literature, there have been a number of attempts to identify the various components of language proficiency and then design tools or procedures to measure these components (Bachman, 1990a; Bachman & Palmer, 1982; Canale, 1983; 1987; Harley, Cummins, Swain, & Allen, 1990; Swain, 1985). This research has

been controversial and no clear consensus has emerged regarding what components to measure and how to measure them. As discussed by Oller (1979; 1989; Oller & Damico, 1991), this is probably due to the powerful integrative trait manifested by language proficiency during performance. Division into components should not occur during data collection or preliminary analysis. Such strategies strip the synergy inherent in language behavior. Rather, there should be an initial focus on language proficiency and communicative performance from a functional perspective. (Attention to more traditional linguistic componential perspectives should be reserved for the stage referred to later as "explanatory analysis").

The focus on functionality suggests that instead of testing a student's knowledge of discrete points of superficial language structure -- or even the student's ability to effectively demonstrate separate componential knowledge of strategic or grammatical competence -- in order to indicate potential language or communicative difficulty, the evaluator asks the question, "How successful is this student as a communicator"? This question of success is based on how well the student functions on three criteria: the effectiveness of meaning transmission, the fluency of meaning transmission, and the appropriateness of meaning transmission (see Table 1).

The Effectiveness of Meaning Transmission. This criterion relates to the primary goal of communication: the formulation, comprehension, and transmission of meaning. Since language is a semiotic system that exists to achieve an understanding of what occurs in the world and since some aspect of this understanding is formulated into communication to relate that understanding to others, how well the individual handles this message (either as a speaker or hearer) is directly relevant to that person's success. The key element, of course, is the message and achieving comprehensibility so that the message is transmitted. Using a functional focus, if the meaning is transmitted -- regardless of how that transmission is achieved -- then communication is accomplished and the individual is effective.

The Fluency of Meaning Transmission. From a functional performance perspective, however, success is more than just getting the meaning across. As stressed by Carroll (1961), the fluency of the interaction must also be considered since successful communicators must be able to formulate, transmit, or comprehend the message within the temporal constraints of communicative interaction. If a student's communicative attempt is delayed, then the flow of communication is affected and this will result in a devaluation of the individual's rating as a communicator. Additionally, a successful communicator can also repair an initial interaction if meaning transmission is not successful. As a speaker, can the student reformulate the message so that it is better comprehended by others? As a listener, can the student successfully ask for clarification or effectively

utilize contextual cues if the initial message is incomprehensible? This ability to seek clarification is another facet of the fluency of meaning transmission.

The Appropriateness of Meaning Transmission. The third criterion for success as a functional communicator is to accomplish the first two objectives in a manner appropriate to the contextual constraints in which the student is immersed at the time of the interaction. Realistically, language and communication are significantly influenced by the expectations that members of a linguistic community share regarding their communicative norms. The attitudes that individuals form and the opportunities afforded to individuals in that community are frequently dependent on those expectations. When addressing the needs of language minority students, this criterion is very important. On numerous occasions, language minority students are poorly evaluated not because of an inability to transmit meaning, but to do so in a culturally appropriate manner (Cummins, 1984; Hamayan & Damico, 1991; Iglesias, 1985).

These three criteria enable the descriptive assessment process to focus on the functional dimension of communicative ability in a manner that transcends the need to divide language proficiency into a variety of skills, modules or components at the time of assessment. Consequently, the synergistic quality of language so important within communicative settings is preserved. Additionally, this initial focus on functionality enables the evaluator to answer real and pragmatic questions about the minority language student's ability to function in the second (and the first) language context as a successful communicator. Actual techniques that can be utilized to accomplish this functional focus will be discussed below.

Descriptiveness

The third essential characteristic of descriptive assessment involves the purposes of evaluation and the types of analyses performed to achieve these purposes. As previously discussed, many language minority students are not only assessed for selection and placement in bilingual programs or for limited proficiency instruction in English. Many of these students are also assessed to determine their underlying language-learning proficiency for placement in special education or remedial programs. It is important, therefore, that descriptive assessment focus on two objectives. First, in order to meet the needs of regular bilingual education programs, governmental funding requirements, and legal regulations (O'Malley, 1989), assessment should provide a detailed description of the individual's communicative performance in English. Second, this descriptive information must then be used to comment on the student's underlying language proficiency. That is, the first objective of descriptive assessment is to determine how successful the student is as a commu-

nicator in English and then, if the student is not successful, to determine the reasons for this lack of success. To accomplish these objectives, the descriptive process must utilize a **bi-level analysis paradigm** that incorporates an initial descriptive analysis of communicative performance with a detailed explanatory analysis of language proficiency (Damico, 1991a) (see Table 1).

Descriptive Analysis. At this descriptive level of analysis, the evaluator typically uses actual descriptive assessment procedures to observe the student's communicative English performance in the contexts and modalities of interest and to determine whether or not the student is communicatively successful. This determination is made by asking the question presented in the previous section on functionality, "How successful a communicator is the student in the context and modality of interest"? This question is answered by one of two strategies. In the first strategy, the descriptive procedure is designed to focus directly on observable behaviors that have been found to be necessary for successful communication in the targeted context and modality. These behaviors are usually selected while designing the descriptive assessment tool through the application of criterion-based, communication-based, or curriculum-based procedures (Cziko, 1983; Hudson & Lynch, 1984; Marston & Magnusson, 1987; Nelson, 1989; Tucker, 1985).

The second strategy used to answer the functionality question at the descriptive level of analysis focuses on potential problematic behaviors. In this strategy, descriptive procedures are designed to detect behaviors that are believed to be valid indices of communicative difficulty. Specific behaviors are identified as indicating when a student is experiencing problems during the communicative interaction and these behaviors are targeted for coding during assessment (Damico, Oller & Storey, 1983; Goodman & Goodman, 1977; Mattes, 1985).

Regardless of the strategy used, this descriptive level of analysis provides a determination of the student's success as a communicator in the targeted English contexts. It is at this level that the primary assessment objectives discussed by Henning (1987) and O'Malley (1989) for limited English proficient (LEP) student identification, placement, and program evaluation are accomplished. By using the actual descriptive analysis procedures, data are provided to document the student's strengths and weaknesses in English, the student's overall success as an English language user, and even the student's progress over time (when pre/post analyses are conducted). Many of the descriptive analysis procedures provide objective scores to rank individual students for placement or reclassification purposes or, if such scores are not available, scoring adaptations like those suggested by Navarrete et al. (1990) may be applied.

It should be noted that during this descriptive level of analysis, communicative performance should be analyzed in different contexts and in different modalities to ensure a rich and wide-ranging coverage of the minority student's success in various language usage manifestations. As discussed by Cummins (1984), there are various dimensions of language proficiency that interact to produce performance distinctions in bilingual (and monolingual) language users (e.g., the CALP and BICS distinction). Similar recommendations for the assessment of language performance in different modalities and contexts have been advocated by others (Canale, 1983; Damico, 1991a; Luria, 1981; Oller, 1979). A discussion of several descriptive analysis procedures is provided below.

If there are no difficulties noted after a descriptive analysis is performed or if there are no attempts to refer the language minority student for further (special education or remedial) testing, then the student is considered for appropriate placement in a bilingual program, a program for English instruction, or in the mainstream classroom. These placements are dependent on how successful the student is as an English language user in the contexts and modalities of interest. This decision meets the general student assessment purposes discussed by Henning (1987) and O'Malley (1989). If there are difficulties indicated at this level of analysis and if there are concerns regarding the student's language-learning proficiency, then the second level of analysis is necessary.

Explanatory Analysis. This analysis seeks to determine the causal factors for the communicative difficulties noted in the descriptive analysis. At this analytic level, the examiner notes the absence of the indices for success (the first strategy above) or the presence of the problematic behaviors (the second strategy above) and seeks to determine why these behaviors occurred.

The explanatory level of analysis typically does not involve additional data collection or assessment procedures. Rather, this level involves a deeper interpretation of the data collected at the descriptive level. The evaluator attempts to explain what aspects of the context, the student's social/cultural experience, or the individual's cognitive abilities or linguistic proficiency can account for the described difficulties. At this level, the evaluator attempts to determine the adequacy of the student's underlying language proficiency or comment on the effectiveness of the student's deeper semiotic capacities (see Oller & Damico, 1991). To accomplish this analysis, a number of strategies can be utilized (Damico, 1991a; Goodman & Goodman, 1977). One effective way to structure this analysis for language minority students is with a set of questions that may be systematically applied to explanatory analysis. A modification of Damico's list (1991a) will be discussed below.

The descriptiveness characteristic enables this performance assessment approach to accomplish the various purposes of assessment while maintaining a functional perspective with authentic language and communicative data. In order to fully adhere to each of these three design characteristics, however, the fourth characteristic of descriptive performance assessment is required. That is, the assessment must be individualized.

Individualized Assessment

Descriptive performance assessment requires individualized assessment. Unlike a number of the discrete point language tests and even some of the more integrative testing procedures, the descriptive process emphasizes individualized observation and analysis. This characteristic is particularly important when determining the need for an explanatory analysis. While the description of the student's communicative performance from an authentic and functional perspective is difficult enough, the analysis of the student's underlying language proficiency based on results of the descriptive analysis procedures is virtually impossible unless conducted on one student at a time. In order to richly describe the complex phenomenon of language proficiency for the purposes previously discussed, time and effort are required. Given the importance of the placement decisions, however, such individualization should not seem excessive. Quality language education of language minority students should require nothing less.

Psychometric Veracity

The final essential characteristic of descriptive performance assessment is psychometric veracity. Similar to a general definition of construct validity (Cronbach, 1971; Messick, 1980), this characteristic reflects the interaction of the other four characteristics once they are carefully implemented. This concept embraces the idea that the tests and procedures used during assessment must be genuine and effective measures of language proficiency and communicative performance. Consequently, veracity requires strong psychometric qualities of reliability, validity, and educational or programmatic utility. In order to exhibit veracity, the assessment procedures must focus on authentic data and must target specific behaviors to use as indices of language proficiency and communicative performance. The evaluator must know what behaviors indicate successful or unsuccessful communication and these behaviors must be able to reflect on the student's underlying language proficiency. The identification of the indices during test development involves several steps (see Table 2).

Table 2
The Steps Required for Determination of Valid Behavioral Indices

Step 1. Select the targeted contexts and modalities.

Step 2. Identify functional behaviors required for meaning transmission.
- Strategy One: Identify behaviors that are indices of successful communicative performance.
- Strategy Two: Identify behaviors that are indices of communicative difficulty.

Step 3. Determine the Reliability of the selected behaviors.
- Temporal reliability
- Interexaminer reliability

Step 4. Determine the Validity of the selected behaviors.

Step 5. Determine the Educational and Programmatic Utility of the selected behaviors and the assessment procedure.

First, the test designer or the evaluator must determine the actual contexts and communicative modalities that will be targeted by the descriptive procedure. This is essential since the data targeted and the student's manifestations of language and communicative ability will differ across contexts and modalities. For example, the language needed to be successful during a writing lesson in the classroom is different from the language needed for a conversation at a friend's home after school. The proficiency required and the communicative behaviors manifested are quite different (Cummins, 1984). To identify valid indices, therefore, the context and modality of interest must be selected.

Once the target manifestations are selected, then behaviors that have a functional role in the transmission of meaning in that manifestation should be identified. This identification is the second step and the two strategies that may be used to accomplish it have been previously discussed. The third step involves the psychometric concept of reliability. Reliability is necessary to ensure that the specific behaviors used as indices for language proficiency are consistent and stable enough in their occurrence (temporally reliable) to be considered as true indices of a stable underlying language proficiency. Additionally, these behaviors must also be easy to observe and code by individuals trained to use the procedures. If different evaluators cannot easily code these behaviors and agree on their occurrences (interexaminer reliability), then the potential of these behaviors as effective indices is greatly diminished. If these behaviors are demon-

strably reliable over time and across examiners, however, then the next step can be taken with these behaviors.

A determination of the validity of these behaviors as indices is the next step. Simply put, how well do these reliable behaviors actually reflect on the language minority student's underlying language proficiency? If a descriptive procedure using these behaviors enables an examiner to make accurate predictions about an individual's language-based performance over time and outside of the assessment situation, then validity is demonstrated and the descriptive procedure is useful (Bachman, 1990a; Cronbach, 1971; Oller, 1979). Without some indication of a procedure's validity and an index's role in that validity, however, the effectiveness of the procedure is reduced.

The last step in establishing psychometric veracity involves determining the educational or programmatic utility of the procedures. Regardless of how reliable and valid a procedure (and its indices) might be, it must be relatively easy to learn and apply in the educational setting or instructional program and it must reflect the instructional goals and objectives of the setting or program. If a procedure requires an inordinate amount of time or equipment to implement or if the data obtained is inconsequential to the setting or program, then it is unlikely that the procedure will be embraced by any evaluator constrained for time. Descriptive performance assessment procedures must reflect the realities and limitations of the school systems that employ the evaluators.

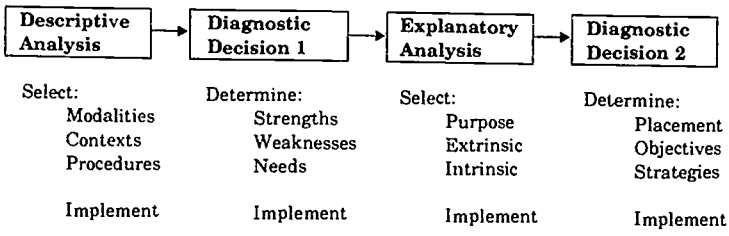
These five characteristics of descriptive performance assessment act to ensure that the processes and procedures used to evaluate language minority students are effective. To implement performance assessment, however, it is not enough to merely describe the characteristics of the descriptive approach. A description of the actual process of assessment and some of the procedures useful for performance assessment is also necessary.

The Descriptive Assessment Process

As previously discussed, an evaluator using the descriptive approach conducts the assessment process differently from other assessment approaches. Remaining consistent with the authenticity characteristic, communication is assessed as it functions holistically in its various manifestations and within naturalistic contexts. The process for a complete assessment involves four sequential stages. First, the evaluator conducts a descriptive analysis. Based on the findings of this analysis, the second stage involves making the first set of diagnostic decisions. If assessment is still warranted after this stage, the next stage of assessment involves the utilization of an ex-

planatory analysis. Finally, based on the results of this deeper level of analysis, the second set of diagnostic decisions is made. Figure 1 provides a flowchart description of this descriptive assessment process.

Figure 1
A flowchart description of the
Descriptive Performance Assessment Process



When conducting a **descriptive analysis**, the evaluator asks the question, "In the present domain of interest, how successful is this student as a communicator?" To conduct this stage of the process, the evaluator determines the modalities of language use and the observational contexts that should be assessed. The evaluator then chooses communicative assessment procedures that will allow the description of communication as its functions in whatever modalities and contexts are appropriate to the objectives of the program or setting. These tend to vary according to the individual programs. For example, Damico (1991a) has specified three manifestations of language use that are important to his assessment objectives. In other situations, however, only one manifestation of language use might be important. Regardless of whether several modalities (e.g., writing, speaking, reading) or several contexts (e.g., conversation with friends, job interview, classroom discussion group, lesson recitation) are evaluated, assessment procedures that describe authentic communication from a functional perspective are required for evaluation. These assessment procedures should typically focus on all aspects of communicative effectiveness (language and speech) together and allow for a determination of communicative success based on the three criteria of effectiveness, fluency, and appropriateness of meaning transmission.

Once the stage of descriptive analysis is completed, the evaluator makes the **first** set of **diagnostic decisions**. At this time, the evaluator uses the data collected during the previous stage to determine whether or not there are any communicative difficulties in the targeted language manifestations. If no difficulties are noted, then the assessment process is completed and the evaluator can describe the individual's strengths as evidence of communicative success and

strong language proficiency. If difficulties are noted in one or more of the language manifestations, however, then the evaluator has several tasks.

First, from a functional perspective, the evaluator describes the student's individual strengths and weaknesses. This information will assist in identification and placement into supportive educational programs. Next, the evaluator determines the actual needs of the student at this stage of the assessment process. If there is evidence of strong first language ability or if there is no desire to refer the student on for further (special education) assessment, then the actual placement decisions (e.g., bilingual classroom, ESL instruction) can be made for this student and the evaluation is completed. However, if there is no evidence of normal first language ability, if there are indications of potential language-learning difficulties at the deeper level of language proficiency or semiotic capacity, or if others in the educational settings request further evaluation, then the next stage of the descriptive assessment process occurs.

This stage of the descriptive process involves the second level of analysis described under the bi-level analysis paradigm mentioned previously -- **explanatory analysis**. At this stage of assessment, the evaluator examines the difficulties noted during the descriptive analysis stage and attempts to determine why the individual had these particular difficulties in the manifestations under scrutiny. Initially, extraneous variables are examined as potential explanations for these problematic behaviors (e.g., second language acquisition phenomena, contextual complexity, listener reactions, significant cultural differences). If no extraneous explanatory factors are noted, then a more systematic analysis of the actual linguistic data is initiated. Based on this analysis, the evaluator determines the underlying causes of the problematic behaviors identified during the descriptive analysis stage and can determine if a true intrinsic language-learning disorder exists.

Finally, at the last stage of the process, the **second** set of **diagnostic decisions** are made. Based on the results of the explanatory analysis and the opportunities available to the student, appropriate placement recommendations are made. If the explanatory analysis demonstrates difficulties due to extraneous variables and not due to the student's underlying language proficiency or deeper semiotic capacity, then the recommendations will be for various types of support systems or programs that will benefit the student's acquisition of English and academic material within regular educational formats. A number of authors have discussed various pedagogical strategies along these lines (Chamot & O'Malley, 1987; Cochran, 1989; Crandall, Spanos, Christian, Simich-Dudgeon, & Willetts, 1988; Fradd, 1987; Hamayan & Perlman, 1990; O'Malley & Chamot, 1990). If the difficulties appear to be due to more intrinsic cognitive, linguis-

tic, or semiotic factors, then appropriate placement and remediation in special education may be warranted (Cummins, 1984; Damico & Nye, 1990; Willig & Ortiz, 1991).

In summary, the assessment process utilizes authentic and functional procedures to analyze an individual's communication in a descriptive and pragmatic fashion through the application of a four stage process centering around a bi-level analysis paradigm. This process is consistent with the hierarchical and synergistic model of language proficiency proposed by Oller and may be used to fulfill the purposes of assessment for language minority students.

The Descriptive Analysis Procedures

To remain consistent with the descriptive assessment approach, the assessment tools and procedures utilized to describe language and communication (i.e., procedures implemented during the descriptive analysis stage), must be able to analyze language use in authentic situations from a functional perspective. Over the past 15 years, a number of procedures and tools have been developed that fit these requirements. In general, these procedures can be conveniently organized according to their different data collection formats and the primary behavioral manifestations that these procedures target. In keeping with Cummins (1984), Canale (1984), and Damico (1991a), these tools and procedures will be organized into four major data collection formats (probes, behavioral sampling, rating scales and protocols, and direct and on-line observation) and two general language usage manifestations (conversational and academic). While these divisions are too general to provide a rigorous classification system, they will permit sufficient organization for this discussion.

Probe Procedures

The probe format is the most widely used of the four data collection strategies. It has been used in the design of norm-referenced tests and many integrative and descriptive procedures. There are numerous variations within this category. For example, there are picture elicitation probes, question and answer probes, elicited imitation probes, interactive computer probes, direct translation probes, and role-playing activities to name a few. Probes are structured tasks or activities that elicit a specific language behavior from the individual being assessed. With probes, the evaluator may anticipate the type of response that will be elicited from the student. This is because the task performed, whether discrete-point, integrative, or pragmatic, is carefully designed to elicit a specific behavior.

For conversational purposes, there are a number of probe activities that have been suggested. Brinton and Fujiki (1991), for example, have structured an assessment activity used to probe a

student's ability to revise utterances, maintain topics, and ask relevant questions during conversation. Olshtain and Blum-Kulka (1985) have suggested several elicitation techniques that focus on language usage, while Deyes (1984) and Jonz (1987) have adapted cloze procedures for more discourse or conversational assessment. In working with adolescents, Brown, Anderson, Shillicock & Yule (1984) have suggested a number of task-based activities that tap the student's ability to use transactional language during interactions.

Academically, probes activities are widely used. The current trend toward criterion-referenced assessment in second language testing (Cloud, 1991; Cziko, 1983; Hudson & Lynch, 1984) and some applications of curriculum-based assessment (Marston & Magnusson, 1987; Tucker, 1985) are dependent on probe activities. Simon (1989) has developed a comprehensive analysis of classroom communicative abilities needed to transition from elementary to secondary school that employs several probe strategies in addition to other formats. Her work has been found to be quite successful in commenting on the success of language minority students to function in the mainstream classroom environment. When reviewed, all achievement tests and many locally-constructed measures of the academic performance of language minority students are discovered to be designed as probes (Brindley, 1986; Cloud, 1991; O'Malley, 1989) and for academic purposes, this format is very effective.

A final academic probe procedure that warrants discussion is the cloze test. As discussed by Oller (1979), and others (Hamayan, Kwait, & Perlman, 1985; Jonz, 1987) cloze procedures accurately reflect on the student's underlying language proficiency and are usually highly correlated with academic performance in English (Oller & Perkins, 1978; 1980). For example, Laesch and van Kleeck (1987) demonstrated significant correlations between their cloze procedure and the California Test of Basic Skills. The cloze procedure was effective in measuring the language needed in academic tasks and it discriminated between subjects with varying degrees of proficiency.

Behavioral Sampling

The second assessment format involves behavioral sampling procedures. Within this popular strategy, the student being assessed completes some required task and this performance is audio-recorded (or video-recorded) and transcribed or the performance is collected in some other way. After data collection, the behavioral sample is analyzed. This format has been extensively applied over the past 25 years and there are numerous procedures available for both conversational assessment and for evaluation of many academic activities.

Conversationally, a number of functional language sampling procedures are available. Loban (1976) emphasizes dimensions of clar-

ity of expression, fluency, command of lexical expression, and comprehension. Blank and Franklin (1980) have developed an appropriateness scale to analyze spontaneous language samples and Adams and Bishop (1990) focus on exchange structures, clarity, and cohesion. The recent work of both Van Lier (1989) and Wesche (1990) show potential as functional descriptive assessment procedures for language minority students. Emphasizing the role of organizational structure, predictability, and speaker/listener rights, these researchers are attempting to focus on discourse skills that are most relevant to second language users in their second languages.

One language sample procedure that is effective in determining conversational success is Clinical Discourse Analysis (Damico, 1985a). This procedure employs a set of 17 problematic behaviors and the theoretical framework of H.P. Grice's Cooperative Principle (1975) to provide a descriptive evaluation that focuses specifically on the effectiveness, fluency, and appropriateness criteria mentioned previously. Listed below are the 17 targeted behaviors as classified within the Gricean framework:

Quantity Category
- Failure to provide significant information to the listener
- Using nonspecific vocabulary
- Informational redundancy
- Need for repetition

Quality Category
- Message inaccuracy

Relation Category
- Poor topic maintenance
- Inappropriate response
- Failure to ask relevant questions
- Situational inappropriateness
- Inappropriate speech style

Manner Category
- Linguistic non-fluency
- Revision behaviors
- Delays before responding
- Failure to structure discourse
- Turn-taking difficulty
- Gaze inefficiency
- Inappropriate intonational contour

Research has indicated that these behaviors are effective in identifying students with communicative difficulty (Damico, 1985b; 1991a; Damico & Oller, 1980; Damico, Oller, & Storey; 1984).

There are a number of effective academically-related behavior sampling procedures available for performance assessment. For example, there are many excellent narrative analyses that can be used for assessment purposes. Applebee (1978) provides a narrative organizational analysis based on developmental stages involving the production of coherent text that can be adapted for assessment while Westby (1991; Westby, Van Dongen & Maggart, 1989), Garnett (1986), Hedberg and Stoel-Gammon (1986), and Roth (1986) have demonstrated the effectiveness of several complex story grammar and narrative analyses for evaluative purposes. While it is realized that narrative development and organization are highly culturally-dependent (Heath, 1986), these behavioral analyses provide important data regarding English task expectations.

Curriculum-based assessment from a subject-based perspective (Deno, 1985; Marston & Magnusson, 1987; Tucker, 1985) and reading miscue analysis (Goodman & Goodman, 1977) frequently involve behavioral sampling in order to accomplish the actual goals of the assessment. Recently, both types of procedures have been advocated as promising informal assessments in bilingual education (Navarrete et al., 1990). Two other promising approaches along these lines are Nelson's "Curriculum-based language assessment" (1989; 1991) and Creaghead's "Classroom Script Analysis" (1991). Both of these procedures use behavior analysis to determine whether or not the student has the communicative strategies (Nelson) or the interactive scripts (Creaghead) essential to effective functioning in a classroom setting.

Finally, when targeting academically-related assessment, Portfolio Assessment must be considered. This behavioral sampling procedure is currently receiving much attention in education. Arising for evaluation purposes from the literacy and language arts fields (Flood & Lapp, 1989; Jongsma, 1989; Mathews, 1990; Valencia, 1990; Wolf, 1989), this procedure is somewhat different from many behavioral sampling procedures in that a primary "evaluation" of the artifacts placed in the student's portfolio involves generalized comparisons rather than detailed analyses. Still, this procedure is very effective in documenting the student's current performance level and his/her progress over time. If portfolio assessment is used with care and if specific evaluative procedures and processes are meshed with the current concept, then this procedure should be very effective in the academic evaluation of language minority students (Moya & O'Malley, 1990).

Rating Scales and Protocols

The third format for data collection involves rating scales and protocols. This format enables the examiner to observe the student as a communicator in the context of interest and then rate or de-

scribe that student according to a set of reliable and valid indices of communication. After the observation, the examiner completes a rating scale or protocol when the student being assessed is no longer present. Two frequent variations of this format are checklists and interview questions. Typically, procedures within this format have some sort of evaluation (e.g., numerical scale, age range, semantic differential, forced judgment of appropriate/inappropriate) for each behavior on the scale or protocol.

For conversational purposes, a number of rating scales have been developed. Damico and Oller (1985) created a functional language screening instrument, Spotting Language Problems, that is an effective rating for screening school-age individuals for communication difficulties, while Mattes (1985; Mattes & Omark, 1984) and Cheng (1987) have designed several protocols involving both verbal and nonverbal behaviors which are helpful in the descriptive assessment of Spanish and Asian LEP students. A widely known descriptive protocol, the Pragmatic Protocol (Prutting and Kirchner, 1983; 1987), focuses on a large number of language usage behaviors and requires that the evaluator rate the student's ability to use these behaviors appropriately or inappropriately. Of course, a number of the well-established evaluation procedures in ESL and EFL make use of rating scales as a basis for their evaluations (e.g., ACTFL and FSI oral interview) and more are being developed. While modifications may make these procedures more relevant, many of these procedures are currently not appropriate to the needs of the students targeted in this paper.

Academically, there are rating scales, protocols, checklists, and interview questionnaires that focus on the functional needs of the student in the classroom. For example, Ortiz (1988) has offered a questionnaire consisting of 25 questions that revolve around the evaluation of the student's educational context while Cloud (1991) has provided several questionnaires to describe home background from an academic perspective, classroom environment, and previous educational experience. In terms of checklists, O'Malley (1989) provides an "Interpersonal and Academic Skills Checklist" that focuses on 30 skills important for cognitive academic language proficiency and a "Literacy Development Checklist" to guide the evaluator in functional assessment of language minority students. In related applied linguistic fields, Nelson (1985), Creaghead and Tattershall (1985), and Larson and McKinley (1987) have also provided checklists that may be beneficial while the work of Archer and Edward (1982) and Bassett, Whittington and Staton-Spicer (1978) can be adapted for assessment within this format. Although this work was not developed originally for language minority students, these tools have been successful for our assessment purposes.

Direct and On-line Observation

The final division of assessment procedures employs direct and on-line observation. Although effective applications of this data collection approach are still relatively rare in language assessment, the approach holds promise. This format involves the direct observation of a student's communicative interaction and the real-time and immediate coding of the communicative behaviors observed. Consequently, these procedures are able to provide detailed and objective data on the speaker's performance rather than just a final judgment of sufficient/insufficient or appropriate/inappropriate communicative performance.

Two observational systems will be detailed. Both are applicable for conversational and academic evaluation. The first, Social Interactive Coding System (SICS), was designed by Rice, Sell, and Hadley (1990) to describe the speaker's verbal interactive status in conjunction with the setting, the conversational partner, and the activities in which the speaker is engaged. This tool requires a 20-minute observational period during which the evaluator observes and codes free play for 5 minutes and then takes a 5 minute break to fill in any codes which might have been missed. This "5 minutes on, 5 minutes off" format is followed for four consecutive cycles until the 20 minutes of direct observation is accomplished. This procedure was designed for use in a bilingual preschool setting but can be modified for other age groups.

The second direct observational procedure is Systematic Observation of Communicative Interaction (SOCI) (Damico, 1985b; 1991b). This tool was designed to employ a balanced set of low inference and high inference items to achieve a reliable coding of illocutionary acts, verbal and nonverbal problematic behaviors, and a determination of the appropriateness of the student's communicative interaction. Once trained to identify and code the behaviors, the evaluator observes the student for 12 minutes and codes the interactions observed each 10 seconds. This yields 72 coded cells of data per observation. The evaluator observes the student from four to seven times and this allows for sufficient data to make representative descriptions of behavior. This tool has very high reliability and validity indices (Damico, 1985b).

The Explanatory Analysis Procedure

As previously discussed, explanatory analysis involves a deeper analysis of the data collected during the descriptive analysis stage to find how/why the student exhibits the communicative difficulties documented. To answer this question, the evaluator must determine

whether the problematic behaviors are due to factors extrinsic to the student or due to intrinsic difficulties at the student's deeper level of language proficiency or semiotic capacity. The true language-learning disabled student will have intrinsic explanatory factors.

While there are several ways to conduct explanatory analysis, this paper will briefly discuss the procedure reported by Damico (1991a). According to this procedure, analysis proceeds with the evaluator asking a series of questions that enable a systematic review of those variables that might have contributed to the communicative difficulties in English. Since detailed discussion is reported elsewhere (Damico, 1991a), only the questions will be provided.

In analyzing the communicative difficulties revealed during descriptive analysis, the evaluator should apply two general sets of questions. First, regarding **extrinsic explanatory factors**:

1. Are there any overt variables that immediately explain the communicative difficulties in English? Among the potential considerations:

    - Are the documented problematic behaviors occurring at a frequency level that would be considered within normal limits or in random variation?
    - Were there any procedural mistakes in the descriptive analysis phase which accounts for the problematic behaviors?
    - Is there an indication of extreme test anxiety during the observational assessment in one context but not in subsequent ones?
    - Is there significant performance inconsistency between different contexts within the targeted manifestation?
    - Is there significant performance inconsistency between different input or output modalities?

2. Is there evidence that the problematic behaviors noted in the second language can be explained according to normal second language acquisition or dialectal phenomena?

3. Is there any evidence that the problematic behaviors noted in the second language can be explained according to cross-cultural interference or related cultural phenomena?

4. Are the communicative difficulties due to a documented lack of proficiency only in the second language but not in the first?

    - Is there documented evidence of normal first language proficiency?
    - Has the student received sufficient exposure to the second language to predict better current performance?

- Does the student exhibit the same types of problematic behaviors in the first language as in the second?

5. Is there any evidence that the problematic behaviors noted in the second language can be explained according to any bias effect that was in operation before, during, or after the assessment?

    - Is the student in a subtractive bilingual environment?
    - Is the student a member of a disempowered community?
    - Are negative or lowered expectations for this student held by the student, the student's family, or the educational staff?
    - Were specific indications of bias evident in the referral, administrative, scoring, or interpretative phases of the evaluation?

If there are no extrinsic explanations for the data obtained during the descriptive analysis phase of the assessment process, then there must be a greater suspicion that the targeted student does have an intrinsic impairment. If this is the case, then the student should exhibit some underlying linguistic systematicity in both languages that can account for the majority of the behaviors noted in the descriptive analysis stage. If the communicative difficulty cannot be accounted for by asking the first five questions, then the final question aimed at **intrinsic explanatory factors** should be conducted.

6. Is there any underlying linguistic systematicity to the problematic behaviors which were noted during the descriptive analysis phase? This can be determined by completion of the following steps:

    - Ensure that no overt factors account for the problematic behaviors (first five questions),
    - Isolate the turns or utterances which contain the problematic behaviors,
    - Perform a systematic linguistic analysis on these data points looking for consistency in the appearance of problematic behaviors.

This last step means taking the utterances or productions that contained the problematic difficulties and performing a co-occurring structure analysis (Muma, 1978; Damico, 1991a). This will determine if the appearance of the problematic behaviors systematically co-occurs with an increase in linguistic complexity. There are several systematic analyses which have been found to be very effective when conducting this type of analysis of the problematic behaviors. For example, to systematically analyze from a grammatical perspective, the work of Crystal (1979; 1982) and his syntactic, phonological, and prosodic profiles are very effective as is the work of Miller and

Chapman (1983). For effective semantic analyses, Crystal's PRISM (1982), Blank, Rose, and Berlin's (1978) four levels of perceptual-language distancing, and Kamhi and Johnston's propositional complexity analysis (1991) are all practical. Other effective analysis systems that take different perspectives have been described by Halliday and Hasan (1976) and Brown and Yule (1983).

If the evaluator follows the sequence of the questions for explanatory analysis, many of the students will not need a detailed co-occurring linguistic analysis. Their problematic behaviors will be explained by extrinsic variables. If this descriptive approach is implemented, language minority students will stand less chance of being mis-identified as language impaired when they only exhibit language or cultural differences and the second set of diagnostic decisions can be appropriately determined.

Conclusion

Performance assessment of language minority students must actually target and evaluate true linguistic performance. For several reasons, this is not an easy task. First, we are still too far from a sufficient understanding of language as a semiotic and behavioral phenomenon and from a sufficient understanding of measurement theory and practice to design the ideal assessment processes and procedures. Second, linguistic and communicative assessment is a complicated process that requires effort and expertise on the part of both the test designers and the test users. Good language assessment requires the services of well-trained applied linguists and behavior analysts. Third, our assessment efforts are directed to the group of students in the schools that can least afford poor application and implementation. For many of these students, tests serve as gates and evaluators as gatekeepers to prevent them from achieving their learning potential.

Given our obligations toward language minority students, however, we must attempt to do the best that we can at the present time. We must use our expertise in a proactive manner to design assessment procedures that allow us to meet the needs of our language minority students at the same time we meet the needs of our school systems and programs. These students deserve no less. Given our current knowledge base, it is possible to conduct performance assessment in an effective and efficient manner. To do so, there must be a focus on theoretically defensible procedures and processes that generate authentic and pragmatic results. By adopting a descriptive approach to assessment that utilizes a hierarchial and synergistic construct of language proficiency, this paper has provided some reasonable suggestions and options. While some aspects of this specific pro-

cess and the discussed procedures may not fit the needs of many evaluators, descriptive performance assessment as an approach allows a focus on authentic behaviors from a functional perspective with enough descriptive power to supply answers to the assessment questions of interest in the schools. By implementing a descriptive performance assessment approach, we can serve as agents of our school systems and as advocates for the language minority students that we serve and care about. As professionals, we deserve no less.

References

Adams, C. & Bishop, D.V.M. (1990). Conversational characteristics of children with semantic-pragmatic disorder. I: Exchange structure, turntaking, repairs, and cohesion. British Journal of Disorders of Communication, 24, 211-239.

Applebee, A.N. (1978). The child's concept of story. Chicago: University of Chicago.

Archer, P. and Edward, J.R. (1982). Predicting school achievement from data on pupils obtained from teachers: Towards a screening device for disadvantaged. Journal of Educational Psychology, 74, 761-770.

Bachman, L.F. (1990a). Fundamental considerations in language testing. Oxford: Oxford University Press.

Bachman, L.F. (1990b). Constructing measures and measuring constructs. In B. Harley, J. Cummins, M. Swain and & P. Allen (Eds.), The development of second language proficiency (pp. 26-38), Cambridge: Cambridge University Press.

Bachman, L.F. & Palmer, A.S. (1982). The construct validation of some components of language proficiency. TESOL Quarterly, 16, 449-465.

Bassett, R.E., Whittington, N., and Staton-Spicer, A. (1978). The basics in speaking and listening for high school graduates: What should be assessed? Communication Education, 27, 300-307.

Blank, M. and Franklin, E. (1980). Dialogue with preschoolers: A cognitively-based system of assessment. Applied Psycholinguistics, 1, 127-150.

Briere, E. (1973). Cross-cultural biases in language testing. In J.W. Oller and J.C. Richards (Eds.). Focus on the learner: Pragmatic perspectives for the language teacher (pp. 214-230). Rowley, MA: Newbury House.

Brindley, G. (1986). The assessment of second language proficiency: Issues and approaches. South Australia: National Curriculum Resource Center.

Brinton, B. & Fujiki, M. (1991). Setting the context for conversational language sampling. Best Practices in School Speech Language Pathology, 2.

Brown, G., Anderson, A., Shillicock, R., and Yule, G. (1984). Teaching talk: Strategies for production and assessment. Cambridge: Cambridge University Press.

Brown, G. and Yule, G. (1983). Discourse Analysis. Cambridge: Cambridge University Press.

Canale, M. (1983). From communicative language pedagogy. In J.C. Richards and R.W. Schmidt (Eds.), Language and communication (pp. 2-28) London: Longman.

Canale, M. (1984). A communicative approach to language proficiency assessment in a minority setting. In C. Rivera Ed.), Communicative competence approaches to language proficiency assessment: Research and application (pp. 107-122). Avon, England: Multilingual Matters.

Canale, M. (1987). Language assessment: The method is the message. In D.Tannen and J.E. Alatis (Eds.), The interdependence of theory, data, and application. (Georgetown University Roundtable) (pp. 249-262). Washington, DC: Georgetown University Press

Canale, M. (1988). The measurement of communicative competence. In R.B. Kaplan et al. (Eds.), Annual review of applied linguistics, VIII, (pp. 67-84) New York: Cambridge University Press.

Canale, M. & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. Applied Linguistics, 1, 1-47.

Carroll, J.B. (1961). Fundamental considerations in testing for English proficiency of foreign students. In Testing the English proficiency of foreign students (pp. 31-40). Washington, DC: Center for Applied Linguistics.

Carroll, J.B. (1983). Psychometric theory and language testing. In J.W. Oller, Jr. (Ed.), Issues in language testing research (pp. 80-107). Rowley, MA: Newbury House.

Chamot, A.U. & O'Malley, J.M. (1987). The Cognitive Academic Language Learning Approach (CALLA): A bridge to the mainstream. TESOL Quarterly, 21, 227-249.

Cheng, L.L. (1987b). Assessing Asian language performance. Rockville, MD: Aspen.

Cloud, N. (1991). Educational assessment. In E.V. Hamayan & J.S. Damico (Eds.), Limiting bias in the assessment of bilingual students (pp. 220-245). Austin, TX: PRO-ED.

Cochran, C. (1989). Strategies for involving LEP students in the all-English-medium classroom: A cooperative learning approach. Washington, DC: NCBE.

Crandall, J., Spanos, G., Christian, D., Simich-Dudgeon, C., & Willetts, K. (1988). Combining language and content instruction for language minority students. Wheaton, MD: NCBE.

Creaghead, N.A. (1991). What does this student know about school? Analysis of classroom scripts for interaction. Best Practices in School Speech Language Pathology, 2.

Creaghead, N.A. and Tattershall, S.S. (1985). Observation and assessment of classroom pragmatic skills. In C.S. Simon (Ed.) Communication skills and classroom success: Assessment of language-learning disabled students (pp. 105-134). San Diego: College-Hill Press.

Cronbach, L.J. (1971) Test validation. In R. Thorndike (Ed.), Educational measurement (pp. 443-507). Washington, DC: American Council on Education.

Crystal, D. (1979). Working with LARSP. New York: Elsevier.

Crystal, D. (1982). Profiling Linguistic Disability. London: Edward Arnold.

Crystal, D. (1987). Towards a 'bucket' theory of language disability: taking account of interaction between linguistic levels. Clinical Linguistics and Phonetics, 1, 7-22.

Cummins, J.P. (1983). Language proficiency and academic achievement. In J.W. Oller, Jr. (Ed.). Issues in language testing research (pp. 108-130). Rowley, MA: Newbury House.

Cummins, J.P. (1984). Bilingualism and special education: Issues in assessment and pedagogy. San Diego: College Hill Press

Cummins, J.P. (1986). Empowering minority students: A framework for intervention. Harvard Educational Review, 56, 18-36.

Cummins, J.P. (1989). Empowering minority students. Sacramento, CA: CABE.

Cziko, G.A. (1983). Psychometric and edumetric approaches to language testing. In J.W. Oller, Jr. (Ed.). Issues in language testing research (pp. 289-307). Rowley, MA: Newbury House.

Damico, J.S. (1985a). Clinical Discourse Analysis: A functional language assessment technique. In C.S. Simon (Ed.), Communication skills and classroom success: Assessment of language-learning disabled students (pp. 165-204) San Diego: College Hill Press.

Damico, J.S. (1985b). The effectiveness of direct observation as a language assessment technique. Unpublished doctoral dissertation. University of New Mexico. Albuquerque, NM

Damico, J.S. (1991a). Descriptive assessment of communicative ability in LEP students. In E.V. Hamayan & J.S. Damico (Eds.), Limiting bias in the assessment of bilingual students (pp. 157-218). Austin, TX: PRO-ED.

Damico, J.S. (1991b). Systematic Observation of Communicative Interaction: A valid and practical descriptive assessment technique. Best Practices in School Speech-Language Pathology, 2.

Damico, J.S. & Nye, C. (1990). Collaborative issues in multicultural populations. Best Practices in School Speech-Language Pathology, 1, 127-140.

Damico, J.S. and Oller, J.W., Jr. (1980). Pragmatic versus morphological/syntactic criteria for language referrals. Language, Speech, and Hearing Services in Schools, 9, 85-94.

Damico, J.S. and Oller, J.W., Jr. (1985). Spotting Language Problems. San Diego: Los Amigos Research Associates.

Damico, J.S., Oller, J.W., Jr., and Storey, M.E. (1983). The diagnosis of language disorders in bilingual children: Pragmatic and Surface-oriented criteria. Journal of Speech and Hearing Disorders, 48, 385-394.

Damico, J.S., Secord. W.A., & Wiig, E.H. (1991). Descriptive language assessment at school: Characteristics and design. Best Practices in School Speech Language Pathology, 2.

Deno, S.L. (1985). Curriculum-based measurement: The emerging alternative. Exceptional Children, 52, 219-232.

Deyes,T. (1984). Towards an authentic "discourse" cloze. Applied Linguistics, 5, 128-137.

Douglas, D., and Selinker, L. (1985). Principles for language tests within the 'discourse domains' theory of interlanguage. Language Testing, 2, 205-226.

Flood, J. & Lapp, D. (1989). Reporting reading progress: A comparison portfolio for parents. The Reading Teacher, 42, 508-514.

Fradd, S.H. (1987) The changing focus of bilingual education. In S.H. Fradd & W.J. Tikinoff (Eds.), Bilingual education and bilingual special education: A guide for administrators (pp. 45-74). Austin, TX: PRO-ED.

Fuller, B. (1982). Synergetics. New York: MacMillan.

Garnett, K. (1986). Telling tales: Narratives and learning- disabled children. Topics in Language Disorders, 6, 44-56.

Goodman, K. & Goodman, Y. (1977). Learning about the psycholinguistic processes by analyzing oral reading. Harvard Educational Review, 47, 317-333.

Grice, H.P. (1975). Logic and conversation. In P. Cole and J. Morgan (Eds) Syntax and Semantics: Speech Acts, Vol. 3 (pp. 41-59). New York: Academic Press.

Halliday, M.A.K. and Hasan, R. (1976). Cohesion in English. London: Longman.

Hamayan, E.V. & Damico, J.S. (1991) Developing and using a second language. In E.V. Hamayan & J.S. Damico (Eds.), Limiting bias in the assessment of bilingual students (pp. 40-76). Austin, TX: PRO-ED.

Hamayan, E., Kwiat, J., and Perlman, R. (1985). Assessment of language minority students: A handbook for educators. Arlington Heights, IL: Illinois Resource Center.

Hamayan, E.V. & Perlman, R. (1990). Helping language minority students after they exit from bilingual/ESL programs. Washington, DC: NCBE.

Harley, B., Cummins, J., Swain, M. & Allen, P. (1990). The nature of language proficiency. In B. Harley, J. Cummins, M. Swain & P. Allen (Eds.), The development of second language proficiency (pp. 7-25), Cambridge: Cambridge University Press.

Heath, S.B. (1986). Taking a cross-cultural look at narratives. Topics in Language Disorders, 7, 84-94.

Hedberg, N.L. and Stoel-Gammon, C. (1986). Narrative analysis: Clinical procedures. Topics in Language Disorders, 7, 58-69.

Henning, G. (1987). A guide to language testing: Development-Evaluation-research. Rowley, MA: Newbury House.

Iglesias, A. (1985). Cultural conflict in the classroom: The communicatively different child. In D.N. Ripich and F.M. Spinelli (Eds.), School discourse problems (pp. 79-96). San Diego, CA: College-Hill Press.

Jongsma, K.S. (1989). Portfolio assessment. The Reading Teacher, 43, 264-265.

Jonz, J. (1987). Textual cohesion and second-language comprehension. Language Learning, 37, 409-438.

Kagan, J. (1967). On the need for relativism. American Psychologist, 22, 131-142.

Kamhi, A.G. & Johnston, J.R. (1991). Semantic assessment: Determining propositional complexity. Best Practices in School Speech Language Pathology, 2.

Laesch, K.B. & van Kleeck, A. (1987). The cloze test as an alternative measure of language proficiency of children considered for exit from Bilingual Education. Language Learning, 37, 171-189.

Larson, V.L. and McKinley, N.L. (1987). Communication assessment and intervention strategies for adolescents. Eau Claire, WI: Thinking Publications.

Loban, W. (1976). Language development: K-12. Urbana, IL: National Council of Teachers of English.

Luria, A..R. (1981). Language and cognition. New York: J. Wiley and sons.

Marston, D. & Magnusson, D. (1987). Curriculum-based measurement: An introduction. Minneapolis,MN: Minneapolis Public Schools.

Mathews, J.K. (1990). From computer management to portfolio assessments. The Reading Teacher, 43, 420-421.

Mattes, L.J. (1985). Spanish Language Assessment Procedures: A communication skills inventory. San Diego: Los Amigos Research Associates.

Mattes, L.J. and Omark, D.R. (1984). Speech and language assessment for the bilingual handicapped. San Diego, CA: College-Hill Press.

Mercer, J.R. (1984). What is a racially and culturally nondiscriminatory test? A sociological and pluralistic perspective. In C.R. Reynolds and R.T. Brown (Eds.), Perspectives on bias in mental testing (pp. 293-356). New York: Plenum Press.

Messick, S. (1980). Test validity and the ethics of assessment. American Psychologist, 35, 1012-1027.

Miller, J. and Chapman, R. (1983). SALT: Systematic analysis of language transcripts, User's manual. Madison: University of Wisconsin.

Moya, S.S. and O'Malley, J.M. (1990, April) A portfolio assessment model of ESL. A paper presented at International TESOL Convention, San Francisco, CA.

Muma, J.R. (1978). The language handbook. Englewood Cliffs: Prentice-Hall.

Navarrete, C., Wilde, J., Nelson, C., Martinez, R. & Hargett, G. (1990). Informal assessment in educational evaluation: Implications for bilingual education programs. Washington, DC: NCBE.

Nelson, N.W. (1985). Teacher talk and child listening — fostering a better match. In C.S. Simon (Ed.), Communication skills and classroom success: Assessment of language-learning disabled students (pp. 65-104). San Diego, CA: College-Hill Press.

Nelson. N.W. (1989). Curriculum-based language assessment and intervention. Language, Speech, and Hearing Services in Schools, 20, 170-184.

Nelson, N.W. (1991). Targets of curriculum-based language assessment. Best Practices in School Speech-Language Pathology, 2.

Oakes, J. (1985). Keeping track: How schools structure inequality. New Haven, CN: Yale University Press.

Ogbu, J. (1978). Minority education and caste: The American system in cross-cultural perspective. New York: Academic Press.

Oller, J.W., Jr. (1972). Scoring methods and difficulty levels for cloze tests of proficiency in English as a second language. Modern Language Journal, 56, 151-158.

Oller, J. W., Jr. (1979). Language tests at school. London: Longman.

Oller, J.W., Jr. (1983). A consensus for the 80's? In J.W. Oller, Jr. (Ed.), Issues in second language testing research (pp. 351-356). Rowley, MA: Newbury House.

Oller, J.W., Jr. (1989). Conclusions toward a rational pragmatism. In J.W. Oller, Jr. (Ed.), Language and Experience (223-250). Lanham, MD: University Press of America.

Oller, J.W., Jr. & Damico, J.S. (1991). Theoretical considerations in the assessment of LEP students. In E.V. Hamayan & J.S. Damico (Eds.), Limiting bias in the assessment of bilingual students (pp. 77-110). Austin, TX: PRO-ED.

Oller, J.W., Jr. & Perkins, K. (Eds.), (1978). Language in education: Testing the tests. Rowley, MA: Newbury House.

Oller, J.W., Jr. & Perkins, K. (Eds.), (1980). Research in language testing. Rowley, MA: Newbury House.

Olshtain, E., and Blum-Kulka, S. (1985). Cross cultural pragmatics and the testing of communicative competence. Language Testing, 2, 16-30.

O'Malley, J.M. (1989). Language proficiency testing with limited English-proficient students. Georgetown University Round Table on Languages and Linguistics (pp. 235-244). Washington, D.C.: Georgetown University Press.

O'Malley, J.M. & Chamot, A.U. (1990) Learning strategies in second language acquisition. New York: Cambridge University Press.

Ortiz, A.A. (1988). Evaluating educational contexts in which language minority students are served. Bilingual Special Education Newsletter, 7, 1-4, 7.

Ortiz, A.A. and Wilkinson, C.Y. (1987). Limited English proficient and English proficient Hispanic students with communication disorders: Characteristics at initial assessment and at reevaluation. Austin TX. The University of Texas, Handicapped Minority Research Institute on Language Proficiency.

Ortiz, A.A. and Yates, J.R. (1983). Incidence of exceptionality among Hispanics: Implications for manpower planning. NABE Journal, 7, 41-54.

Prutting, C.A. and Kirchner, D.M. (1983). Applied pragmatics. In T.M. Gallagher and C.A. Prutting (Eds.), Pragmatic assessment and intervention issues in language (pp. 29-64). San Diego: College-Hill Press.

Prutting, C.A. and Kirchner, D.M. (1987). A clinical appraisal of the pragmatic aspects of language. Journal of Speech and Hearing Disorders, 52, 105-119.

Rice, M.L., Sell, M.A., and Hadley, P.A. (1990). The Social Interactive Coding System (SICS): An on-line, clinically relevant descriptive tool. Language, Speech, and Hearing Services in Schools, 21, 2-14.

Roth, F.P. (1986). Oral narrative abilities of learning-disabled students. Topics in Language Disorders, 7, 21-30.

Seliger, H.W. (1982). Testing authentic language: the problem of meaning. Language Testing, 2, 60-73.

Shohamy, E. (1983). The stability of the oral proficiency trait on the oral interview speaking test. Language Learning, 33, 527-540.

Shohamy, E. (1988). A proposed framework for testing the oral language of second/foreign language learners. Studies in Second Language Acquisition, 10, 165-180.

Shohamy, E. (1991). Discourse analysis in language testing. Annual Review of Applied Linguistics, 11, 115-131.

Shohamy, E. and Reves, T. (1982). Authentic language tests: where from and where to? Language Testing, 2, 48-59.

Shuy, R. (1981). A holistic view of language. Research in the teaching of English, 15, 101-111.

Simon, C.S. (1989). Classroom Communication Screening Procedure for Early Adolescents: A handbook for assessment and intervention. Tempe, AZ: Communi-Cog Publications.

Spolsky, B. (1973). What does it mean to know a language? Or how do you get someone to perform his competence? In J.W. Oller & J.C. Richards (Eds.) Focus on the learner (pp. 164-176). Rowley, MA: Newbury House.

Swain, M. (1985). Large-scale communicative language testing: A case study. In Y.P. Lee, et al. (Eds.), New directions in language testing (pp. 35-46). Oxford: Pergamon Press.

Tucker, J.A. (1985). Curriculum-based assessment: An introduction. Exceptional Children, 52, 199-204.

Upshur, J.A. (1973). Productive communication testing: progress report. In J.W. Oller & J.C. Richards (Eds.) Focus on the learner (pp. 177-183). Rowley, MA: Newbury House.

Valencia, S. (1990). A portfolio approach to classroom reading assessment: The whys, whats, and hows. The Reading Teacher, 43. 338-340.

Van Lier, L. (1989). Reeling, writhing, drawling, stretching and fainting in coils: Oral proficiency interviews as conversations. TESOL Quarterly, 23, 480-508.

Wesche, M. (1987). Second language performance testing: The Ontario Test of ESL as an example. Language Testing, 4, 28-47.

Westby, C.E. (1991). Narrative analysis. Best Practices in School Speech Language Pathology, 2.

Westby, C.E., Van Dongen, R., and Maggart, Z. (1989). Assessing narrative competence. Seminars in Speech and Language, 10, 63-76.

Willig, A.C. & Ortiz, A.A. (1991). The non-biased Individualized Educational Program: Linking assessment to instruction. In E.V. Hamayan & J.S. Damico (Eds.), Limiting bias in the assessment of bilingual students (pp. 281-302). Austin, TX: PRO-ED.

Wolf, D.P. (1989). Portfolio assessment: Sampling student work. Educational Leadership, 46 (7), 35-39.

Response to Jack Damico's Presentation

J. Michael O'Malley
Georgetown University

There is a growing national interest in alternative assessment as a means of determining the knowledge and skills of students in our schools. This interest stems in part from dissatisfaction with standardized tests but also originates in theoretical arguments about how children learn and what they learn. The interest is reflected in the number of professional articles on alternative assessment, in topics addressed at conferences and in workshops, in statewide testing policies, and in the national debate on the format for a national achievement test. This national interest is compatible with and has acted to advance the concerns expressed for years about standardized tests by educators of language minority students.

The Damico paper presented at this symposium, "Performance Assessment of Language Minority Students," picks up on the interests of bilingual and English as a second language (ESL) educators for appropriate uses of assessment and suggests a variety of alternative assessment procedures for identifying and placing students acquiring English. The rationale for performance or alternative assessment, according to Damico, lies in the unique characteristics of students acquiring English in schools and in theoretical arguments concerning the hierarchical and integrative nature of language proficiency in communicative contexts. Damico describes the essential features of what he refers to as "descriptive assessment," sketches out the process for conducting the assessment, and offers a classification scheme with supportive examples to illustrate varied forms of performance assessment. Additionally, he offers an approach for both descriptive and explanatory analysis of the data from performance assessment that is intended to provide a comprehensive picture of student language proficiency in education.

Damico largely succeeds in what he sets out to do although there are some minor issues that I would differ with at various points in his paper. My principal concerns are what he did not cover under the general rubric of performance assessment, some aspects of which may be more in need of attention than the topics he raises. The best way to illustrate these concerns is to retrace some of the same ground Damico covers but from a different perspective, thereby building a foundation for the areas that I think need further discussion. While discussing these topics, I will present a rationale for alternative assessment, a definition of "academic language proficiency," and draw out the implications of these for alternative assessment in schools. Following that analysis, I will return to Damico's paper for some further comments.

Rationale for Alternative Assessment

The growing interest in alternative assessment among language minority educators has been marked by an increasing number of requests for related workshops at the Georgetown University Evaluation Assistance Center (EAC)-East in 1990 and 1991. The topics covered in these workshops have included various forms of alternative assessment and portfolio development and have been presented throughout the entire Eastern region, including Puerto Rico. Educators participating in these workshops have commented on the utility of alternative assessment for classroom applications.

The rationale for practitioner requests for information on alternative assessment lies in part in dissatisfaction with standardized tests but also stems from specific instructional needs that have not been addressed in assessment. Educators are looking for assessment that will meet multiple purposes. They are looking for assessment that can be used for identification and placement, as Damico indicates, but they are also looking for assessment that will provide a continuous record of student growth. Educators need to know how students are progressing so that they can adapt instruction to student needs, communicate indicators of progress to the student or to parents, and develop a plan for assisting the student to handle academic content in English.

The need to maintain a continuous record of student progress is an important difference from the purposes of assessment that are described by Damico. Having a continuous record of student progress requires that assessment take place periodically throughout the school year and must fit within limited time constraints when teachers have other planning and instructional responsibilities to meet. Not all of the procedures suggested by Damico meet these time constraints, and some do not seem suited to maintaining a record of student progress.

Educators who have requested EAC-East workshops are also looking for assessment that reflects multiple perspectives on student language proficiency so that they can balance one form of information against another in analyzing student performance. They are especially interested in expanding on the limited perspective permitted from the use of standardized tests with language minority students, since so often the students receive low scores due to factors that are unrelated to their actual knowledge and skills. The problem with having multiple perspectives on language proficiency is that the information needs to be integrated in a systematic way.

The integration and interpretation of information from multiple assessment need considerably more attention than Damico had the

opportunity to give it. What is required is a clear focus on the purposes of the assessment, the educational goals and objectives the instruments being used are designed to assess, and a procedure for interpreting each type of assessment in relation to these objectives and to each other. I will address the interpretation issue later in commenting on applications in schools.

Another benefit educators participating in these workshops hope to gain is a perspective on conducting assessment that is authentic. They are looking for assessment procedures that reflect actual tasks that students work on in classrooms rather than the relatively isolated tasks performed in responding to multiple choice tests. As Damico notes, alternative assessment can provide information that is authentic in that it reflects actual "communicative activities within real contexts" (p. 11). Damico defines authenticity in terms of linguistic realism and ecological validity. In linguistic realism, linguistic behavior is treated as a "complex and synergistic phenomenon that exists primarily for the transmission and interpretation of meaning" (p. 11). In ecological validity, communication occurs in a naturalistic setting "where true communicative performance is occurring and is influenced by contextual factors" (p. 12). The emphasis here is on the use of language for communication, a point that Damico emphasizes repeatedly, as when he notes that the interest in assessment should be on the question, "'How successful is this student as a communicator'" (p. 14).

One difficulty with Damico's approach to authenticity is that he primarily uses a linguistic rather than an academic base for language proficiency. Although he alludes to Cummins' (1984) distinction between basic interpersonal communicative skills (BICS) and cognitive academic language proficiency (CALP), he does not integrate the distinction into his definition of language proficiency or into the discussion of assessment instruments he describes. Furthermore, although the underlying theory on which Damico's paper is based (Oller & Damico, 1991) discusses academic language proficiency, the linguistic origin of the theory does not lead to specific recommendations for assessment of academic language skills.

A comprehensive assessment of the performance of language minority students in school will not be complete without a more precise view of the demands inherent in using academic language. Cummins' definition of academic language skills in terms of two orthogonal continua -- one focusing on the cognitive complexity of the task, and the other on the degree of contextualized support for meaning -- has served the field well in a variety of ways. Nevertheless, the definition is incomplete because it fails to describe the nature of the cognitive activity that makes academic tasks complex. The design of assessment for academic language skills needs more precision than is afforded by this general outline.

Definition of Academic Language Proficiency

My colleague Anna Chamot and I have been working for the past six years on a content based instructional method in English as a second Language that we have referred to as the Cognitive Academic Language Learning Approach (CALLA) (Chamot & O'Malley, 1987). Two distinctive features of CALLA are its incorporation of learning strategies into instructional activities and the inclusion of academic language from the content areas into ESL objectives and tasks. While relying on Cummins' definition of CALP, we realized early on that the definition had limitations precisely because the nature of the academic task requirements that lead to cognitive complexity were incompletely specified. My description of academic language in what follows is drawn from our CALLA Handbook (Chamot & O'Malley, forthcoming) and from our earlier book developing the theoretical and research base for CALLA (O'Malley & Chamot, 1990).

Academic language can be defined in terms of the vocabulary and conventions specific to any content area but, more importantly, can be understood most clearly in terms of the language functions needed for authentic academic content. Academic language functions are essential tasks that language users must be able to perform in the different content areas, and they are what makes the task simple or complex. These functions often differ from social interactive language functions. For example, one social language function is greeting or addressing another person. Sub-categories of greeting are greeting a peer, a superior, or a subordinate, and making the greeting either formal or informal. On the other hand, academic language may involve using language functions such as identifying and describing content information, explaining a process, analyzing and synthesizing concepts, justifying opinions, or evaluating knowledge.

In many classrooms academic language tends to be unidirectional: the teacher and textbook impart information and students demonstrate their comprehension by answering oral and written questions. But academic language can also be interactive. Teachers and students can discuss new concepts, share analyses, and argue about values in both teacher-student and student-student interactions. Since academic language functions such as describing and explaining can also occur with basic interpersonal interactions, it is the specific academic context that makes these functions apply to academic language proficiency.

Language functions needed in academic content include informing, classifying, comparing, justifying, persuading, synthesizing, and evaluating, as represented in Table 1. Most of these functions are required -- or should be required -- in all content areas, including mathematics, history, science, and literature. To accomplish these

functions successfully with academic content requires the use of both lower and higher order thinking skills. Lower order thinking skills and less cognitively complex tasks might include recalling facts, identifying vocabulary, and giving definitions. Higher order skills involve using language to analyze, synthesize, and evaluate. There is obviously a close connection between the difficulty of the academic language task and higher order thinking.

Table 1
Academic Language Functions

Language Function	Student Uses Language to:	Examples
Seek Information	explore the environment or acquire information	Use who, what, when, where, and how to collect information
Inform	report, explain, or describe information or procedures	Retell story or content-related information in own words, tell main ideas, summarize
Analyze	separation of whole into parts	Tells parts or features of object or idea
Compare	analyze similarities and differences in objects or ideas	Indicate similarities and differences in important parts or features of objects or ideas, outline/diagram/web, indicate how A contrasts/compares with B
Classify	sort objects or ideas into groups and give reasons	Show how A is an example of B, how A is related to B, or how A and B go together but not C and D
Predict	predict implications	Predict implications from actions or from stated text
Hypothesize	hypothesize consequences	Generate hypotheses to suggest consequences from antecedents
Justify	give reasons for an action, a decision, or a point of view	Tell why A is important, why you selected A, or why you believe A
Persuade	convince another person of a point of view	Show at least two pieces of evidence or arguments in support of a position
Solve Problems	determine solution	Given stated problem, reach solution
Synthesize	combine ideas to form a new whole	Put A together with B to make C, predict or infer C from A and B, suggest a solution for a problem
Evaluate	assess the worth of an object, opinion, or decision	Select or name criteria to evaluate, prioritize a list and explain, evaluate an object or proposition, indicate reasons for agreeing or disagreeing

It is because of this close interrelationship that academic language skills can best be identified by describing both the language functions and the level of thinking skills needed to perform a specific task. The student's level of proficiency is then described in terms of both the functions and the thinking skills that are employed on the task. Furthermore, the functions and thinking skills demanded by a task prescribe the complexity of the language structures and the number of independent concepts that must be integrated in performing the task. Thus, the linguistic aspects of the task are prescribed by the content. This is important, because it leads to the conclusion that the evaluator needs to analyze the academic content requirements in order to understand the language requirements of instruction. In other words, assessment of academic language does not begin with an analysis of language or with language theory, but with an analysis of the academic objectives and content requirements. This is quite different from the approach advocated by Damico.

The theory underlying the approach to assessment I am suggesting originates in cognitive psychology. A substantial body of theory has emerged describing the mental processes learners use in performing complex tasks and how these processes influence learning (e.g., Anderson, 1985; Gagné, 1985; Garner, 1987; Jones & Idol, 1990). One of the conclusions from this research is that individuals use active mental processes while learning, including the learning that occurs in second language acquisition (O'Malley & Chamot, 1990). At least a portion of these mental processes entail a higher order understanding of the requirements for learning on any particular activity and an examination of prior knowledge that will assist in the new learning. One of the other conclusions is that an important component of new learning is the domain-specific knowledge that individuals bring to the task, suggesting that an analysis of the specific content demands in any domain is important for understanding performance requirements.

Implications for Alternative Assessment

There are a variety of implications for alternative assessment in combining language functions with higher order thinking skills to define academic language proficiency. In drawing these implications, I assume that students acquiring English are enrolled in a program that will incorporate at least some form of academic content such as a bilingual program or a content-based ESL program. If the special program for students acquiring English does not contain academic content, and is limited to a language focus, the student at some point will be included in content area classes and will be called upon to understand and produce academic language.

One of the first implications of the definition of academic language advocated in this paper is that the design of alternative as-

sessment originates with an analysis of the curriculum framework for the content areas rather than with an analysis of language or a language-based syllabus. The analysis of language demands in the context provided by content area instruction produces an understanding of the language that needs to be evaluated through alternative assessment. This analysis can take place in content area texts or by analyzing the language content area teachers use in classrooms. A second implication is that alternative assessment is best thought of as a form of domain assessment that has curriculum validity for the concepts, skills, and language used in performing academic tasks. As such, alternative assessment needs to be continuous in order to reflect students' understandings of and ability to use curriculum content introduced throughout the school year. A third implication is that alternative assessment needs to reflect the complexity of concepts, skills, and language that are integrated in performing academic tasks. Because it is difficult for any single assessment approach to capture this complexity, multiple assessment needs to be used in order to gain varied perspectives on the students' academic growth. A fourth implication is that new kinds of performance instruments are needed that will assess this complexity using authentic academic tasks in which the language functions and higher order thinking skills will be evidenced by students. Because of the authentic nature of these tasks, the assessment need not take time away from teaching but should be part of the instructional process.

One of the aspects of learning that we have come to appreciate through our studies of the application of learning strategies to instruction is the importance of the processes that students use in learning concepts and skills. Learning strategies are mental or overt procedures that students use to assist their own learning. In a CALLA program, strategies are taught directly in order to ensure that students will have a satisfactory repertoire of skills for learning academic content. Because learning strategies are among the stated outcomes of instruction, they are included among the objectives and, accordingly, are assessed. Thus, in a CALLA program, alternative assessment will include assessment of learning strategies and learning processes in addition to the assessment of academic and linguistic outcomes. This is another major difference between the assessment approach expressed here and the approach suggested by Damico.

Practical Applications to Instruction

The complex and varied requirements for alternative assessment of academic language proficiency call for a straightforward approach to the interpretation of data. A strong and visible role needs to be given to portfolio development in any discussion of alternative assessment for this reason. The design of the portfolio should be focused and systematic, and the interpretation of data in the portfolio

should enable users to integrate information from a variety of different sources (Moya & O'Malley, 1990). The design should consist of a five-step portfolio development process that includes the following stages:

1. Design - the statement of the purposes of the portfolio, and the selection of a committee of teachers and other staff to design the portfolio, collect the information, and review the data;

2. Focus - a statement of instructional goals that will be assessed using the portfolio information, and selection of alternative or other assessment instruments or data collection procedures to be included in the portfolio;

3. Data Collection - assignment of responsibilities for collecting the data in addition to the data collection schedule;

4. Interpretation - design of procedures for integrating the information obtained from multiple assessment, relating it to the goals, and making it useful in instruction; and

5. Evaluation - evaluation of the portfolio process, reliability of the scoring, and the portfolio's usefulness in instructional decision making for individual students or in meeting other purposes established for the portfolio.

It is in the fourth stage of the portfolio development process that the committee formed to design and use the data from the portfolio specifies the relationships among instructional objectives, the evidence and nature of student progress, and the specific instruments that do or do not support the conclusion that the student has progressed toward the objectives. Thus, the instructional use of alternative assessment is embedded in the portfolio design. Without the portfolio, the teacher is left with an unmanageable collection of alternative assessment information that is difficult to relate to the instructional intent.

Conclusion

I have presented a different view of alternative assessment from that suggested by Damico in order to highlight the way in which I believe language needs to be assessed in schools. I do not suggest that Damico's analysis is flawed, simply that there are other ways of analyzing the assessment of language minority students in schools. My major differences with Damico concerned the rationale for using alternative assessment, the definition of language proficiency, the breadth of skills that should be assessed, and the procedures for interpretation of data from alternative assessment.

Despite these differences, there are many commonalities. We concur on the importance of alternative assessment in the education of students acquiring English and in many of the characteristics that alternative assessment should possess. Among these are authenticity, functionality, validity, and both descriptive and explanatory power. Although I did not have the opportunity to elaborate on the forms of alternative assessment I advocate, I am a supporter of performance assessment, direct observation, anecdotal reports, checklists, and rating scales and many of the other types of assessment Damico describes. Let us hope that these common points and the strengths of each approach will lead to the improvement of assessment and instruction for language minority students acquiring English.

References

Anderson, J.R. (1985). Cognitive psychology and its implications. (2nd ed.). New York: Freeman.

Chamot, A.U., & O'Malley, J.M. (1987). The Cognitive Academic Language Learning Approach (CALLA). TESOL Quarterly, 21(3), 227-49.

Chamot, A.U., & O'Malley, J.M. (forthcoming). The CALLA Handbook: How to implement the Cognitive Academic Language Learning Approach. Reading, MA: Addison-Wesley.

Cummins, J. (1984). Bilingualism and special education: Issues in assessment and pedagogy. Clevedon, U.K.: Multilingual Matters.

Gagné, E.D. (1985). The cognitive psychology of school learning. Boston, MA: Little, Brown.

Garner, R. (1987). Metacognition and reading comprehension. Norwood, NJ: Ablex.

Jones, B.F., & Idol, L. (1990) (Eds.). Dimensions of thinking and cognitive instruction. Hillsdale, NJ: Lawrence Erlbaum.

Moya, S., & O'Malley, J.M. (1990, March). Monitoring literacy development: Looking beyond basic skills. Paper presented at the annual meetings of the Teachers of English to Speakers of Other Languages, San Francisco, CA.

Oller, J.W. Jr., & Damico, J.S. (1991). Theoretical considerations in the assessment of LEP students. In E.V. Hamayan & J.S. Damico (Eds.), Limiting bias in the assessment of bilingual students (pp. 77-110). Austin, TX: Pro-ed.

O'Malley, J.M., & Chamot, A.U. (1990). Learning strategies in second language acquisition. Cambridge: Cambridge University Press.

46

Response to Jack Damico's Presentation

Cecilia J. Navarrete
University of New Mexico

Central to the evaluation of bilingual education programs, and Title VII programs in particular, are the instruments and procedures used to assess the progress of language minority students. Dr. Damico's descriptive assessment procedure of communicative abilities offers a performance-based approach to address those evaluation needs. In applying performance-based approaches such as the descriptive assessment procedure in bilingual education we must keep in mind several factors:

Factor 1: Purpose of Assessment

First and foremost is our motives or purposes for engaging in language assessment as a useful activity. In other words, why do we want to engage in language testing in the first place? Dr. Edward DeAvila argues that, from a historical bilingual education perspective, the need to assess the language proficiency of students came about as a result of the *Lau vs Nichols* Supreme Court ruling. This ruling made school districts accountable for providing an equal education to language minority students. It was followed by amendments to the Bilingual Education Act, which provided federal funding not only to assist schools in preparing language minority students to effectively participate in school but also to assess the projects' effectiveness for their participants.

As a result of these decisions, four major evaluation purposes for bilingual education programs have evolved:

(1) identification of LEP students;

(2) placement of LEP students into appropriate programs;

(3) reclassification or exit of program students;

(4) evaluation of students' progress for instruction and evaluation of the program effectiveness).

These purposes, while part and parcel of bilingual programs, at times, function independently from each other and at times are incompatible. For example, if you consider the process for identifying students, the information used most is categorical data. These are the types of data where students are identified into levels such as non-English proficient (NEP), limited English proficient (LEP), and

fluent English proficient (FEP). If other data, such as the results on vocabulary, story comprehension, or writing tests, are thrown away or not used it becomes virtually impossible to do any program planning based on student needs or measure any student progress. The reason is because all one can do with categorical data is count the number of students in each group. Therefore, it is critical that, when collecting information on individual students, procedures be considered for aggregating data into a meaningful set of indices that allow for true assessment of the students' achievements.

Overall, in looking at purposes of assessment in bilingual education and assessment approaches such as Dr. Damico's, it is important to seek a balance between what is needed at the classroom, federal, and state level. The greater the gap between the purposes of bilingual education and assessment, be it alternative or traditional standardized assessments, the greater the likelihood of distorting our understanding of the relationship between language and schooling, at least as defined by the *Lau* decision -- the framework upon which bilingual education is based.

Factor 2: Issues of Validity

In thinking further about the purposes of bilingual education assessment, it is crucial to determine how we propose to validate the kind of assessments we choose -- that is, how we propose to demonstrate what we measured for some important or real sense. Alan Davis, in his recent book Principles of Language Testing (1990), emphasizes the need to assemble evidence about any test we choose. In this light, recent work by Richard Stiggins (1990), as well as by Robert Linn, Eva Baker, and Stephen Dunbar (1990) offer criteria that are consistent with traditional psychometric standards for judging the technical adequacy of performance-based measures. These are standards that should be considered before using any performance-based approach.

The major value of these criteria is that they are aimed at maximizing the validity of performance-based assessments by including design features such as clarifying the purpose of the assessment, identifying the consequences or specifying uses to be made with the results, and defining in explicit, observable terms the tasks and performance criteria to be considered in the assessment. In general, "authentic" performance assessments of students' performance on instructional tasks must: be accurate and viable; include the fundamental constructs of measurement; and demonstrate how they will contribute to the improvement of instruction and learning, especially for LEP students.

Factor 3: Practical Constraints

Other factors to consider in bilingual evaluations are the obvious practical constraints on language assessment that affect validity. The most obvious practical problem is what Melville called the need for "time, cash, and patience," which for bilingual programs means ensuring that the facilities, the materials, the personnel and so on are available in the numbers and at the times they are required. For example, an assessment that requires individual audio recordings of hundreds, even thousands of LEP students, all to be assessed within a specified time period could prove to be difficult if there is a shortage in the number of qualified and experienced data collectors. Other practical difficulties which need to be foreseen are those such as noise conditions, materials, and/or equipment. Overall, a test or assessment device should be simplified as much as possible and limit its requirements of people, time, and materials.

Another practical constraint has to do with the issue of reliability, especially as it pertains to performance-based assessment. By reliability I am referring to the consistency by which observations, judgments, and results are interpreted. As we seek effective approaches to acquire acceptable levels of reliability there is a minimum set of standards recommended by EAC-West:

(1) design clear, observable scoring criteria in order to maximize the raters' (e.g., teachers or evaluators) understanding of the performance to be evaluated;

(2) ensure training for inter- and intra-rater reliability when more then one person is involved in the scoring process;

(3) allow time to test the observation instrument and its ability to pick up the information desired;

(4) maintain objectivity in assessing student work by periodically checking the consistency of ratings given to students' work in the same area;

(5) keep consistent and continuous records of students to measure their development and learning outcomes;

(6) check judgments by using multiple measures including standardized and other performance-based assessments.

I mention the use of standardized tests as part of the multiple measures package because their are many who are uncertain whether to support or oppose its use. Many bilingual educators have criticized them for not being applicable to their student population or

program aims -- and justifiably so. On the other hand, these types of tests continue to be administered annually by school districts, and they provide a ready source of achievement or linguistic data.

For those who are struggling with the decision on how best to use data that are available in their school district, I recommend reading an article by Blain Worthen and Vicki Spandel in the February 1991 issue of Educational Leadership. They address some of the most common criticisms of standardized testing and offer suggestions on how to avoid the pitfalls of over interpretation and misuse of standardized tests. Some pitfalls they address are:

(1) using a single test score to make important decisions about students;

(2) failing to supplement test scores with other information (for example, the teacher's knowledge of the students); and

(3) assuming tests measure all the content skills or behaviors or interests.

Overall, Worthen and Spandel point out that when standardized tests are used correctly, they do have value, but they provide only part of the picture and do have limitations.

Conclusion

Obviously, there is no quick fix answers to the assessment and testing dilemma of LEP students. However, there are steps we can take to make our evaluations practical, viable, and accurate. We can: (1) maintain a clear understanding of the purposes of bilingual education programs -- from the classroom level to the policy level; (2) Educate ourselves and those involved with LEP students about evaluation and assessment; (3) Carefully avoid any misuses of tests or performance-based assessments; and (4) realize that, no matter which assessment instrument we select, each will have its limitations -- none will be able to provide us with all the answers we are seeking. There are panaceas to assessment. As we look at the types of assessment offered by Dr. Damico we must keep in mind the constraints and limitations of such assessments as well as take into consideration the purposes of assessment for our LEP students.