DOCUMENT RESUME

ED 236 911 FL 013 984

AUTHOR Seidner, Stanley S., Ed.
TITLE Issues of Language Assessment. Volume II: Language Assessment and Curriculum Planning.
INSTITUTION Illinois State Board of Education, Springfield.
SPONS AGENCY Department of Education, Washington, DC.
PUB DATE Aug 83
NOTE 208p.; For related documents, see ED 221 023-033.
PUB TYPE Viewpoints (120) -- Collected Works - General (020)

EDRS PRICE MF01/PC09 Plus Postage.
DESCRIPTORS *Bilingualism; Child Language; Communicative Competence (Languages); *Curriculum Development; *Language Proficiency; *Language Tests; *Limited English Speaking; Second Language Instruction; *Second Language Learning; *Testing

ABSTRACT

These 25 papers address issues of second language assessment and curriculum planning. The papers are divided into three categories: theoretical foundations, assessment approaches, and research and policy. Among the specific topics discussed are: whether tests measure language or intelligence, additive versus subtractive bilingualism, the implications of image production in two languages for testing the bilingual child, a bilingual perspective on language and cognition, the use and misuse of instruments, a review of the IDEA Oral Language Proficiency Test, use of native language tests for program planning purposes, assessment considerations for limited English proficient vocational students, test and spontaneous language behavior of bilingual preadolescents, influence of public policy on language assessment, and estimating bilingual education populations. (RW)

ISSUES OF LANGUAGE ASSESSMENT
VOLUME II: LANGUAGE ASSESSMENT AND CURRICULUM PLANNING

Edited by
Stanley S. Seidner, Ph.D.
Associate Dean for Research and Development
Executive Director, Language Institute and Minorities Resource Center
National College of Education

The activity which is the subject of this report was supported in whole or in part by the U.S. Department of Education and the Illinois State Board of Education. However, the opinions expressed herein do not necessarily reflect the positions or policies of these agencies, and no official endorsement by these agencies should be inferred.

Developed and Published by:
Illinois State Board of Education

Disseminated by:
Evaluation Dissemination and Assessment Center
Dallas, Texas
and
National Clearinghouse on Bilingual Education
Rosslyn, Virginia

TABLE OF CONTENTS

Part III: Research and Policy

WHERE DO WE GO FROM HERE?

The era spanning the nineteen-seventies through the nineteen-eighties might at first suggest revolutionary and dynamic breakthroughs in the field of language assessment. Monumental capacities of computer-enriched statistical approaches have seemingly yielded a plethora of research on the one hand, with veiled inferences of expanding existing knowledge, ad infinitum. From another perspective, the metaphorical geometric rate of amassing this data has perhaps outdistanced our arithmetic ability to process this mass and conceptualize adequate frameworks. However, a corollary to this statement might still be equally argued and applicable to language assessment, among other areas, in that a few visionary prophets may have outdistanced the parameters of current technology and await the time when reification of constructs (sort of speaking) bridges the gap between the theoretical and empirical.

I am reminded somewhat of Zeno's paradox presented by Troike in an earlier volume of this series (Seidner, 1982). In reaching our hypothetical goals, we appear to cover half the distance, then half of the remaining distance, ad infinitum. In essence, all victories, then, become partial victories. But what do they mean? It would seem that this question typifies every era of human endeavor. Think of the immediate sensation caused by Fechner, for example, who converted Weber's Law into a psychological scale (on this, see Boring, 1942, 1961). Nevertheless, the ensuing activity and controversy in Germany had prompted William James (1890) to write:

> "And what good came of it at last?"
> Quoth little Peterkin.
> "Why that I cannot tell," said he,
> "But twas a famous victory!"

The radical impossibility of disassociating experimental approaches from a current theoretical framework compounds individual attempts to transcend the figuratively stated arithmetic rate. Here, Claude Bernard (1865) would also warn of an extreme attachment to existing frameworks, where men "with excessive faith in their theories," are ill-disposed toward making new discoveries and shedding bias from their recorded observations.

I would suggest here that the lack of consensus in definitions of terms alternately impedes and accelerates this arithmetic rate, compared to the current geometric influx of information. A brilliant exchange between John Oller and John Garroll brought forth the latter's personal communication on the relationship of the "g" factor to language proficiency, and the observation that final conclusions would rest upon specific definitions of intelligence and language proficiency (Oller, 1983). Traditionally, theoretical frameworks have developed cautiously for the most part under the scrutiny of orthodoxy. However, the existence of varied definitions might enhance developments in other related areas. If Conant's (1947) concept of the "horror vacui" dictated, in turn, that contradictory evidence in itself would not supplant an established theoretical framework until a better one arose, a newly formulated framework would have to reflect some elements of established tenet. We might find then that the definition of communicative competence advanced by Canale and Swain (Seidner, 1982) illustrates a demarcation from perspectives of Munby (1978), Hymes (in Pride and Holmes, 1972) or Savignon (1972) in some respects, but forms a consensus of agreement along other lines. The operational definition of communicative competence (arbitrarily labeled Z), viewed here fundamentally as the verbal account of particular manipulations and computations, becomes subject to what Carnap viewed as testability, confirmability and verifiability (i.e. where verifiability is characterized with an incompleteness in attaining a definitive and final establishment of "truth"). If we then view a hypothetical assessment instrument of communicative competence, associated with our definition Z, it is the *interpretation* of the data emanating from the particular procedure which is validated, (as indicated by Cronbach in Thorndike, 1971) in opposition to the viewing of validity, for example, as a specific characteristic of the test. An overall view of progress in the historical development in language assessment, then, might be likened to the spiral-like process attributed to Rousseau in the quest of some absolute truth or perfection (De Balla, 1973). Each advancement is subject to the flat plateau of the spiral and awaits the processing of data and emergence of an adequate conceptual framework. The time factor in proceeding along the plateau and through the spiral to the next level varies in its speed and is dependent upon a myriad of factors which cannot be fully explored here. In essence, we encounter the transformation made by Boring (1936) of Occam's "razor" from "entia non sunt multiplicanda praeter necessitatem" to "entia sunt multiplicanda propter necessitem."

If we agree with Day, McCollum, et al. in Erickson and Omark (1981) that discrete-point testing emerged as the natural result of structural linguistic approaches (influenced by behavioral psychology) to language acquisition, the advance along the plateau through the figurative spiral to the next one was relatively a chronologically shorter one. Perhaps the same may be true for adherents of what Carroll (1961) termed as integrative philosophy. An example of this trend may emerge again with some recent stirrings, associated with the reexamination of the so-called unitary factor hypothesis by Farhady (1980, and Oller, 1983) as well as by Briere (1980) among others. If a shortly forthcoming breakthrough were, per se, to occur from this direction in the design and implementation of new language assessment instrumentation, would it be mostly attributed to the aforementioned advances in computerized statistical applications? Indeed, the studies cited in the example appeared to rely upon sophisticated statistically packaged approaches. First appearances might suggest that it constitutes a determining factor in accelerating spiral-like advancement. Consider the comparison of approaching

factor analysis from the actual raw data or from a correlation matrix. Statistical programming affords the individual readily available approaches to address the commonalities in the diagonals of correlation matrices. The inclusion of an oblique (or correlated) method of rotation to "reconcile" highly loaded factors, or perhaps the more popularly used orthogonal, (or uncorrelated), varimax method (developed by Kaiser, 1958) provide an added sophistication in manageable statistical calculations. If these and other approaches appear to confirm the hastening of developments along the spiral, I would suggest a resistance factor — or, perhaps, check and balance — where research, utilizing the normal curve, conforms to the confines of the predetermined standard (See Seidner, 1982b). Returning to Zeno's paradox, the seemingly quickened advance along the plateau which spirals upward might be no more than illusory, in view of the attainment of the half, of the half, of the half, which is perpetual. Given another perspective, the computerized applications might indeed prove a quagmire in instances of statistical applications hindering, or perhaps shaping, the direction of theoretical formulation, detracting from necessary foundations of the fields of the origin. Where do we go from here?

Papers by John Oller and Jim Cummins (Part I) offer us some prospects for new frontiers and breakthroughs. Oller presents an intriguing blend of biological, psychological and linguistic approaches to the question of language-based intelligence. Cummins presents us with a theoretical framework, proposing the conceptualization of language proficiency along several continua. Muriel Saville-Troike's study (Part III) substantially adds to our understanding of communicative and academic competence in teaching and testing situations. Benji Wald (also Part III) offers important findings on the relationship between spontaneous and test speech, utilizing a sample of 10-12-year-old Hispanic bilinguals. These and other contributions are found in the three categories within this volume: Foundations, Assessment Approaches, Research and Policy. As with Volume I, it is with the thought of bringing current issues to the forefront that the Second Language Assessment Institute was conducted in, Chicago, Illinois. This volume was conceived as the second in a series to include deliveries by participants and invited papers from theoreticians and practitioners in the field. Sponsors of the Institute, which will be repeated annually, include the National Clearinghouse for Bilingual Education; Educational Testing Services; the Evaluation, Dissemination and Assessment Center (Dallas, Texas); the Bilingual Education Service Center (Illinois); the Illinois State Bilingual Office; the Illinois Resource Center; and National College of Education.

Behind the scenes in the production of any volume of this nature are the unsung heroes who commit countless hours of toil and dedication. It is with grateful thanks and appreciation that acknowledgement is made to Jean Honeyman, Manager, Staff Support Services, Illinois State Board of Education; to Delores Johnson, Word Processing Supervisor; and to Becky Copelin, Typesetter, Staff Support Services, who so willingly gave so much time and expertise in the typesetting of this project. Bill Becker, Manager, Illinois State Board of Education, Internal Office Supporting and Printing Section, and Bernie Neff, Assistant Manager, Printing and Graphics, are invaluable in providing much needed material support and creative ideas. Special gratitude goes to Linda McElroy who spent many hours poring over the manuscripts and is deserving of the accolade, "an editor's editor." Cornelia Powell and Steve Rothenberg added their much appreciated talent in helping design the cover and other graphics. I would also like to express my indebtedness to Maria Medina Seidner, Manager, Bilingual Education Section, for her assistance and support in the preparation of this volume. Finally, my appreciation goes to the contributors of this volume for their excellent presentations and papers.

Editor
Stanley S. Seidner, Ph.D
Associate Dean for Research and Development
National College of Education

Bibliography

Bernard, Claude. *Introduction a la Medecine experimentale*. Paris: Plon, 1865.

Boring, Edwin. "Psychophysiological Systems and Isomorphic Relations," *Psychological Review*, No. 43, 1936, pp. 565-587.

_____. *Sensation and Perception in the History of Experimental Psychology*. New York: Appleton-Century, 1942.

_____. "The Beginning and Growth of Measurement in Psychology," *Isis*, No. 52, 1961, pp. 238-257.

Briere, Eugene. "Testing Communicative Language Proficiency." Paper, 14th Annual Language Convention, TESOL, San Francisco, 1980.

Carroll, John. *Testing the English Proficiency of Foreign Students*. Washington D.C.: Center for Applied Linguistics, 1961.

Conant, J.B. *On Understanding Science*. New Haven: Yale University Press, 1947.

DeBalla, Borisz. "Is There Progress in History?" Jamaica, New York: St. John's University, 1973. (Mimeo)

Erickson, Joan Good and Donald R. Omark. *Communication Assessment of the Bilingual Bicultural Child*. Baltimore: University Park Press, 1981.

Farhady, Hossein. "Justification, Development, and Validation of Functional Language Tests." Unpublished Ph.D. Dissertation, U.C.L.A., 1980.

James, William. *Principles of Psychology*. New York: Henry Holt, 1890. Vol. I.

Kaiser, H.F. "The Varimax Criterion for Analytic Rotation in Factor Analysis," *Psychometrika*, No. 23, 1958, pp. 187-200.

Munby, J. *Communicative Syllabus Design*. Cambridge: Cambridge University Press, 1978.

Oller, John. *Issues in Language Testing Research*. Rowley: Newbury House, 1983.

Pride, J.B. and J. Holmes. *Sociolinguistics*. Harmondsworth, England: Penguin, 1972.

Savignon, Sandra. *Communicative Competence: An Experiment in Foreign Language Teaching*. Philadelphia: Center for Curriculum Development, 1972.

Seidner, Stanley S. *Issues of Language Assessment. Vol. I: Foundations and Research*. Springfield: Illinois State Board of Education, 1982.

_____. *Ethnicity, Language and Power from a Psycholinguistic Perspective*. (Reprint). Bruxelles: Centre de Recherche sur le Plurilinguisme, 1982b.

Thorndike, Robert L. *Educational Measurement*. Washington D.C.: American Council on Education, 1971.

Part I
Foundations

CONCEPTUAL AND LINGUISTIC FOUNDATIONS OF LANGUAGE ASSESSMENT

Jim Cummins
The Ontario Institute for Studies in Education

Many of the most contentious debates in the areas of psycholinguistics and educational psychology during the past 20 years have revolved around the issue of how "language proficiency" is related to academic achievement. Disagreement about appropriate ways of conceptualizing the nature of language proficiency underlies controversies as diverse as the extent to which "oral language" is related to the acquisition of reading (e.g., Wells, 1981), the extent to which learning disabilities are in reality language disabilities (e.g., Vellutino, 1979), and the extent to which poor school achievement of low socioeconomic status (SES) and minority group students is caused by differences in the language use patterns of these students in comparison to middle class students (e.g., Labov, 1970).

The question of what constitutes "language proficiency" and the nature of its cross-lingual dimensions is also at the core of many hotly debated issues in the areas of bilingual education and second language pedagogy and testing. Researchers have suggested ways of making second language teaching and testing more "communicative" (e.g. Canale & Swain, 1980; Oller, 1979) on the grounds that a communicative approach better reflects the nature of language proficiency than one which emphasizes the acquisition of discrete language skills. Issues such as the effects of bilingual education on achievement, the appropriate age to begin teaching L2, and the consequences of different patterns of bilingual language use in the home on minority students' achievement are all intimately related to the broader issue of how L1 proficiency is related to the development of L2 proficiency. This issue, in turn, clearly cannot be resolved without an adequate conceptualization of the nature of "language proficiency."

I shall first briefly review the current state-of-the-art in bilingual language proficiency assessment and then describe some of the detrimental consequences for minority students' academic development which derive from popular assumptions about the nature of "English proficiency." Finally, I will outline a theoretical framework for relating language proficiency to academic development and suggest some applications of the framework for both assessment and pedagogy.[1]

Language Proficiency Assessment in Bilingual Programs

A cursory examination of the many tests of language proficiency and dominance currently available for assessing bilingual students. (see, e.g., DeAvila & Duncan, 1978; Dieterich, Freeman and Crandell, 1979) reveals enormous variations in what they purport to measure. Of the 46 tests examined by DeAvila and Duncan (1978), only four included a measure of phoneme production, 43 claimed to measure various levels of lexical ability, 34 included items assessing oral syntax comprehension and nine attempted to assess pragmatic aspects of language.

This variation in language tests is not surprising in view of the lack of consensus as to the nature of language proficiency or communicative competence." For example, Hernandez-Chavez, Burt and Dulay (1978) have outlined a model of language proficiency comprising 64 separate components, each of which, hypothetically at least, is independently measurable. By contrast, Oller and Perkins (1980) have argued that:

> "a single factor of global language proficiency seems to account for the lion's share of variance in a wide variety of educational tests including nonverbal *and* verbal IQ measures, achievement batteries, and even personality inventories and affective measures... the result to date are...preponderantly in favor of the assumption that language skill pervades every area of the school curriculum even more strongly than was ever thought by curriculum writers or testers." (p.1).

This global dimension is not regarded by Oller (in press) as the only significant factor in language proficiency, but the amount of additional variance accounted for by other factors is relatively modest.

The considerable evidence that Oller and his colleagues (e.g. Oller & Streiff, in press) have assembled to show that academic and cognitive variables are strongly related to at least some measures of all four general language skills (listening, speaking, reading, and writing) raises an important issue for the assessment of entry and exit criteria in bilingual programs: to what extent *should* measures of language proficiency be related to measures of academic achievement? In other words, to what extent does the construct of language proficiency overlap with the constructs of "intelligence" and academic achievement?

This theoretical question has rarely been asked; instead, researchers have either asked only the empirical

[1] The need for a theoretical framework explicitly designed to relate language proficiency to academic achievement was brought home to me at the Language Proficiency Assessment Symposium (LPAS) (Rivera, in press) not only as a result of criticisms of the distinction which I had introduced between basic interpersonal communicative skills (BICS) and cognitive/academic language proficiency (CALP) but, more importantly, by the lack of any resolution of the issues to which that distinction was addressed. The present theoretical framework is essentially an elaboration and, it is hoped, a clarification of the BICS-CALP distinction. In addition to the many participants at the LPAS who made valuable suggestions, I would like to acknowledge my debt to John Oller Jr. and to Merrill Swain for many useful discussions on these issues.

question of how language proficiency *is* related to achievement (often expressed in terms of the relation between "oral language" and reading) or else ignored the issue entirely, presumably because they do not consider it relevant to language proficiency assessment in bilingual education. However, the theoretical issue cannot be avoided. The relationship of language proficiency to academic achievement must be considered in view of the fact that a central purpose in assessing minority students' language dominance patterns is to assign students to classes taught through the language in which it is assumed they are most capable of learning, and in which they will most readily acquire academic skills. If measures of language proficiency bear no relationship to students' acquisition of academic skills, their relevance in the context of entry and exit criteria is open to question. This issue requires theoretical resolution, rather than empirical, because, as will be discussed below, some language measures correlate highly with achievement, while others show a negligible relationship. Without a theoretical framework within which language proficiency can be related to the development of academic skills, there is no basis for choosing between alternative tests which are clearly measuring very different things under the guise of "language proficiency."

Essentially, what is at issue are the criteria to be used in determining the validity of language proficiency measures in the specific context of bilingual education. Whether we are talking about content, criterion-related, construct, face, or ecological validity, our procedures for determining validity are always based on a theory regarding the nature of the phenomenon being measured. In many cases, however, this theory has remained implicit in language test development for bilingual students, and where the theory has been made explicit, the construct of language proficiency has usually been regarded as independent of the constructs of intellectual and academic abilities.

Thus, it is reported (see Oakland, 1977, p. 199) that on the *Basic Language Competence Battery* there is little or no increase in scores across the elementary grades among native speakers. This is interpreted as evidence for the construct validity of the battery in that it is indeed measuring "language knowledge," rather than intellectual abilities or educational achievement. In arguing against "language deficit" theories, many sociolinguists (e.g., Labov, 1970; Shuy, 1977) have similarly asserted that language proficiency is independent of cognitive and academic performance. Shuy (1977, p.5), for example, states that "rather compelling evidence rejects every claim made by those who attempt to show linguistic correlates of cognitive deficit."

One apparent implication of the theoretical position that "language proficiency" is independent of intellectual abilities and academic achievement is that language measures such as the integrative tests (e.g. oral cloze, dictation, elicited imitation) used in the research of Oller and others (see Oller & Perkins, 1980, Oller & Streiff, in press) would have to be rejected as invalid to assess the construct of language proficiency because of their strong relationships to achievement and IQ.

Many theorists would regard any form of contrived test situation as inadequate to assess language proficiency, arguing instead for procedures which assess children's language in naturally occurring communicative situations (e.g. Cazden, Bond, Epstein, Matz, & Savignon, 1977; Dieterich et al., 1979). For example, Dieterich et al. argue in relation to an elicited imitation task that "it mirrors no real speech situation and is thus of questionable validity in assessing proficiency" (1977, p. 541).

Although the requirement that proficiency measures reflect "naturally occurring speech situations" is a basic principle of validity for many theorists, few pursue the issue to inquire whether or not the communicative demands of natural face-to-face situations are identical to the communicative demands of classroom situations. In classrooms, students' opportunity to negotiate meaning with the interlocutor (teacher) is considerably reduced as a result of sharing him or her with about 25-30 other students, and there is considerable emphasis on developing proficiency in processing written text where the meaning is supported largely by linguistic cues, rather than the richer "real-life" cues of face-to-face communication.

These issues are being raised not to argue against the assessment of "language proficiency" in naturally occurring situations, but rather to show the need for a theoretical framework which would allow the construct of language proficiency to be conceptualized in relation to the acquisition of academic skills in bilingual programs. The urgency of this need can be seen from the fact that the most commonly used tests of language proficiency and dominance for minority students clearly embody different theoretical assumptions in regard to the relationship between language proficiency and achievement. The Language Assessment Scales (LAS) (DeAvila & Duncan, 1977), for example, are reported to consistently show moderate correlations with academic achievement, whereas the Bilingual Syntax Measure (BMS) (Burt, Dulay, & Hernandez-Chavez, 1975) and the Basic Inventory of Natural Language (BINL) (Herbert, 1975) tend to show much lower correlations with achievement (see Rosansky, 1981 for a review). All of these tests showed lower correlations

---

[2] Much of the vehemence with which researchers have rejected the verbal components of standardized IQ and achievement tests as valid measures of either "language proficiency" or cognitive abilities stems from the blatant misuse of such measures with low socioeconomic status (SES) and ethnic minority students (see for example Cummins, 1980). However, the fact that SES or cultural differences on such measures can be explained by acculturation to middle class majority group norms does not account for differences between individuals *within* SES or cultural groups on cognitively demanding culture-specific measures of proficiency. In other words, it is logically invalid to argue that a particular phenomenon (e.g. cognitive development) does not exist because some of the tools used to measure that phenomenon (e.g. IQ tests) have been abused.

with achievement than teachers' ratings of students' chances for academic achievement if instructed only in English (Urbaun, Spencer & Rivas, 1980). This teacher variable accounted for 41 per cent of the variance in reading achievement scores and the BINL, BSM, and LAS added only zero, one and four percent, respectively to the prediction of reading achievement.

Apart from the issue of their relationship to academic achievement, the validity of these tests can be questioned on several other grounds. For example, Rosansky (1981) points out that data elicited by the BSM-English were unrelated to data elicited from taped naturalistic conversation of the same individuals. The LAS Spanish language classification is reported to considerably underestimate the Spanish proficiency of native Spanish speakers as assessed by other teacher ratings or detailed ethnolinguistic analysis of children's speech in a range of settings (Mace-Matluck, 1980).

This brief survey of assessment issues in bilingual education suggests that a major reason for the confused state of the art is that the developmental relationships between language proficiency (in L1 and L2) and academic performance have scarcely been considered, let alone resolved. The confusion about the assessment of "language proficiency" is reflected in the varied criteria used to exit language minority students from bilingual programs.

"English Proficiency" and Exit Criteria

Lack of English proficiency is commonly regarded by policy-makers and educators as the major cause of language minority students' academic failure in English-only programs. Thus, it is assumed that students require bilingual instruction only until they have become proficient in English. Logically, after students have become "proficient in English," any difficulties they might encounter in an English-only program cannot be attributed to lack of English proficiency.

If we combine this apparent logic with the fact that immigrant students generally appear to acquire a reasonably high level of L2 fluency within about 1½ – 2 years of arrival in the host country (Cummins, 1980c; Snow & Hoefnagel-Hohle, 1978), then one might assume that two years of bilingual education should be sufficient for students to make the transition to an English-only program. This line of reasoning is frequently invoked to justify exiting students out of bilingual programs after a relatively short period. It is assumed that because students can cope adequately with the communicative demands of face-to-face situations and appear quite fluent in English, their English proficiency is sufficiently well-developed to cope with the communicative demands of the regular English-only curriculum on an equal basis with native English speaking students.

There is considerable evidence to suggest that the logic is false. Bilingual programs which have been successful in developing a high level of English academic skills in language minority students have usually maintained instruction in L1 throughout elementary school. Usually it is only in the latter grades of elementary school that students approach grade norms in English reading skills (see Cummins, 1981 for a review). In a similar way, it has been shown (Cummins, in press a) that it took immigrant students who arrived in Canada after the age of six, 5-7 years on the average, to approach grade norms in academically related aspects of English proficiency. Thus, it clearly takes considerably longer for language minority students to develop age-appropriate *academic* skills in English than it does to develop certain aspects of age-appropriate English face-to-face communicative skills. It follows that students exited on the basis of teacher judgments or language tests which primarily assess face-to-face communicative skills are likely to experience considerable academic difficulty in an English-only program, and many will manifest the well-documented pattern of cumulative deficits.

The dangers of unanalyzed notions of what constitutes "English proficiency" can be illustrated by an example from a Canadian study in which the teacher referral forms and psychological assessments of 428 language minority students were analyzed (Cummins, 1980c). This particular child (PR) was first referred in grade 1 by the school principal who noted that:

> "PR is experiencing considerable difficulty with grade 1 work. An intellectual assessment would help her teacher to set realistic learning expectations for her and might provide some clues as to remedial assistance that might be offered."

No mention was made of the child's English-as-a-second-language (ESL) background; this only emerged when the child was referred by the grade 2 teacher in the following year. Thus, the psychologist does not consider this as a possible factor in accounting for the discrepancy between a Verbal IQ of 64 and a Performance IQ of 108. The assessment report read as follows:

> "Although overall ability level appears to be within the low average range, note the significant difference between verbal and nonverbal scores ... It would appear that PR's development has not progressed at a normal rate and consequently she is, and will continue to experience much difficulty in school. Teacher's expectations (at this time) should be set accordingly."

What is interesting in this example is that the child's face-to-face communicative skills are presumably sufficiently well developed that the psychologist (and possibly the teacher) is not alerted to her ESL background. This leads the psychologist to infer from her low verbal IQ score that "her development has not progressed at a normal rate" and to advise the teacher to set low academic expectations for the child since she "will continue to experience much difficulty in school."

There is ample evidence from many contexts (e.g. Mercer, 1973) of how the attribution of deficient cognitive skills to language minority students can become self-fulfilling.

In many of the referral forms and psychological assessments analyzed in this study the following line of reasoning was invoked:

> Because language minority students are fluent in English, their poor academic performance and/or test scores cannot be attributed to lack of proficiency in English. Therefore, these students must either have deficient cognitive abilities or be poorly motivated ('lazy').

In a similar way, when language minority students are exited from bilingual programs on the basis of fluent English communicative skills, it appears that their subsequent academic difficulties cannot logically be attributed to "lack of English proficiency." Thus, educators are likely to attribute these difficulties to factors within the student such as "low academic ability" (IQ).

These misconceptions derive from the fact that the relationships between "language proficiency" and academic development have not been adequately considered, either among native English-speaking or language minority students. In the remainder of this paper, a theoretical framework is developed for conceptualizing these relationships.

A Theoretical Framework[3]

On the basis of the foregoing analysis of the confusions which exist both in current language proficiency assessment techniques and in procedures for exiting students from bilingual programs, three minimal requirements for a theoretical framework of language proficiency relevant to bilingual education in the United States can be outlined: First, such a framework must incorporate a developmental perspective such that those aspects of language proficiency which are mastered early by native speakers and L2 learners can be distinguished from those that continue to vary across individuals as development progresses; second, the framework must be capable of allowing differences between the linguistic demands of the school and those of interpersonal contexts outside the school to be described; third, the framework must be capable of allowing the developmental relationships between L1 and L2 proficiency to be described.

Current theoretical frameworks of "communicative competence" (e.g. Canale, 1981; Canale & Swain, 1980) do not (and were not intended to) meet these requirements. Canale (1981) distinguishes grammatical, sociolinguistic, discourse and strategic competencies, but states that their relationships with each other and with world knowledge and academic achievement is an empirical question yet to be addressed. Although this framework is extremely useful for some purposes, its applicability to bilingual education is limited by its static nondevelopmental nature and by the fact that the relationships between academic performance and the components of communicative competence in L1 and L2 are not considered. For example, both pronunciation and lexical knowledge would be classified under grammatical competence. Yet L1 pronunciation is mastered very early by native speakers, whereas lexical knowledge continues to develop throughout schooling and is strongly related to academic performance.

The framework outlined below is an attempt to conceptualize "language proficiency" in such a way that the developmental interrelationships between academic performance and language proficiency in both L1 and L2 can be considered. It is proposed only in relation to the development of academic skills in bilingual education and is not necessarily appropriate or applicable to other contexts or issues. Essentially, the framework tries to integrate the earlier distinction between basic interpersonal communicative skills (BICS) and cognitive/academic language proficiency (CALP) into a more general theoretical model. The BICS-CALP distinction was intended to make the same point that was made earlier in this paper, namely, academic deficits are often created by teachers and psychologists who fail to realize that it takes language minority students considerably longer to attain grade/age-appropriate levels in English academic skills than it does in English face-to-face communicative skills. However, as is pointed out in other papers in this volume, such a dichotomy oversimplifies the phenomena and risks misinterpretation. It is also difficult to discuss the crucial developmental issues in terms of the BICS-CALP dichotomy.

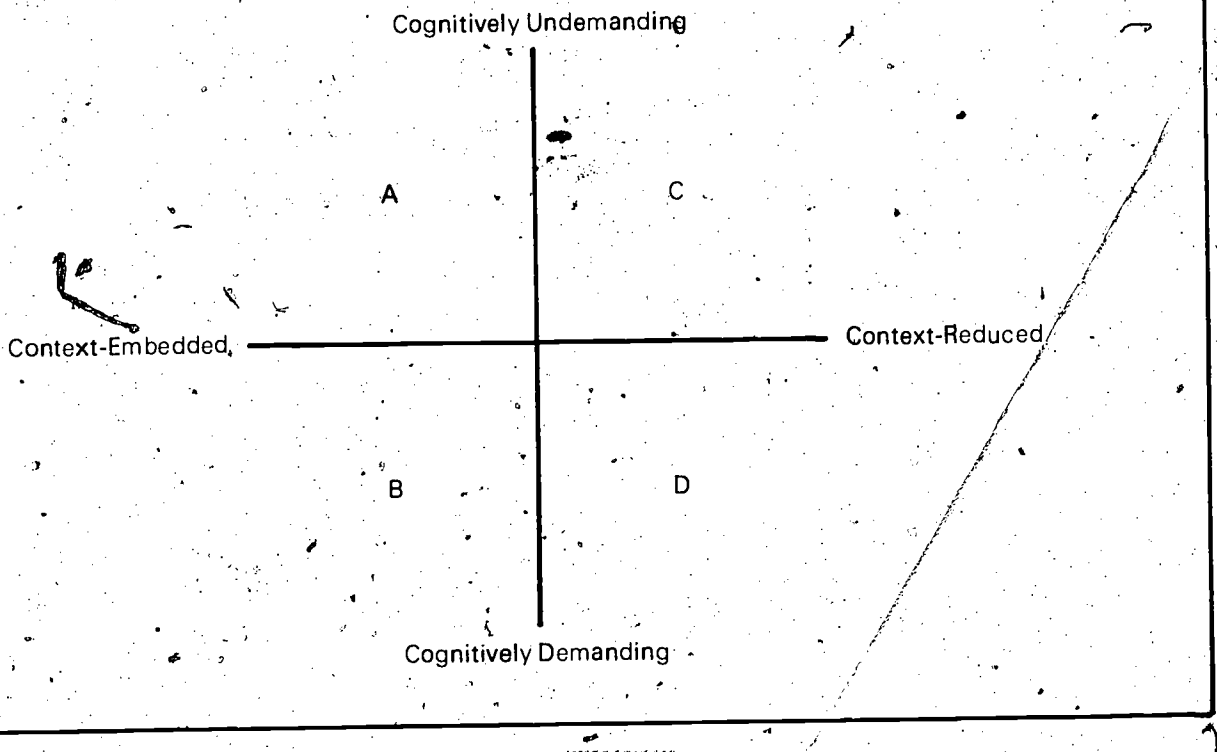
The framework presented in Figure 1 proposes that in the context of bilingual education in the United States "language proficiency" can be conceptualized along two continuums. First is a continuum relating to the range of contextual support available for expressing or receiving meaning. The extremes of this continuum are described in terms of "context-embedded" versus "context-reduced" communication. They are distinguished by the fact that in context-embedded communication the participants can actively negotiate meaning (e.g. by providing feedback that the message has not been understood) and the language is supported by a wide range of meaningful paralinguistic and situational cues; context-reduced communication, on the other hand, relies primarily (or at the extreme of

[3] This theoretical framework should be viewed within a social context. The language proficiencies described develop as a result of various types of communicative interactions in home and school (see e.g. Wells, 1981). The nature of these interactions is, in turn, determined by broader societal factors (see Cummins, 1981c). In order to emphasize the social nature of "language proficiency", this term will be used interchangeably with "communicative proficiency" in describing the framework.

the continuum, exclusively) on linguistic cues to meaning and may in some cases involve suspending knowledge of the "real world" in order to interpret (or manipulate) the logic of the communication appropriately.[4]

Figure 1

Range of Contextual Support and Degree of Cognitive Involvement in Communicative Activities



In general, context-embedded communication derives from interpersonal involvement in a shared reality which obviates the need for explicit linguistic elaboration of the message. Context-reduced communication, on the other hand, derives from the fact that this shared reality cannot be assumed, and thus, linguistic messages must be elaborated precisely and explicitly so that the risk of misinterpretation is minimized. It is important to emphasize that this is a continuum and not a dichotomy. Thus, examples of communicative behaviors going from left to right along the continuum might be: engaging in a discussion, writing a letter to a close friend, writing (or reading) an academic article. Clearly, context-embedded communication is more typical of the everyday world outside the classroom, whereas many of the linguistic demands of the classroom reflect communication which is closer to the context-reduced end of the continuum.

It is useful to distinguish between internal and external contextual factors in relation to the horizontal continuum. External context refers to aspects of language activities or tasks which are more or less objectively specifiable along the embedded-reduced continuum. Thus, inherent text characteristics make reading and writing less context-embedded than face-to-face communication. However, the location of any particular task on the continuum will also be greatly influenced by internal contextual factors such as degree of familiarity and acceptance of the task or activity. In this way, the role of sociolinguistic factors can be accommodated into the framework (see discussion in Rivera, in press).

The vertical continuum is intended to address the developmental aspects of communicative proficiency in terms of the degree of active cognitive involvement in the task or activity. Cognitive involvement can be

[4] The term "context-reduced" is used rather than "disembedded" (Donaldson 1978) or "decontextualized" because there is a large variety of contexual cues available to carry out tasks even at the context-reduced end of the continuum. The difference, however, is that the cues are exclusively *linguistic* in nature.

conceptualized in terms of the amount of information that must be processed simultaneously or in close succession by the individual in order to carry out the activity.

How does this continuum incorporate a developmental perspective? If we return to the four components of communicative competence (grammatical, sociolinguistic, discourse, and strategic) discussed by Canale (1981), it is clear that within each one, some subskills are mastered more rapidly than others. In other words, some subskills (e.g. pronunciation and syntax within L1 grammatical competence) reach plateau levels at which there are no longer significant differences in mastery between individuals (at least in context-embedded situations). Other subskills continue to develop throughout the school years and beyond, depending upon the individual's communicative needs, in particular, cultural and institutional milieux.

Thus, the upper parts of the vertical continuum consist of communicative tasks and activities in which the linguistic tools have become largely automatized (mastered) and thus require little active cognitive involvement for appropriate performance. At the lower end of the continuum are tasks and activities in which the communicative tools have not become automatized and thus require active cognitive involvement. Persuading another individual that one's own point of view, rather than hers/his, is correct or writing an essay on a complex theme are examples of such activities. In these situations, it is necessary to stretch one's linguistic resources (i.e., grammatical, sociolinguistic, discourse and strategic competencies) to the limit in order to achieve one's communicative goals. Obviously, cognitive involvement, in the sense of amount of information processing, can be just as intense in context-embedded as in context-reduced activities.

As mastery is developed, specific linguistic tasks and skills travel from the bottom towards the top of the vertical continuum. In other words, there tends to be a high level of cognitive involvement in task or activity performance until mastery has been achieved or, alternatively, until plateau level at less than mastery levels has been reached (e.g. L2 pronunciation in many adult immigrants, "fossilization" of certain grammatical features among French immersion students, etc.). Thus, learning the phonology and syntax of L1, for example, requires considerable cognitive involvement for two- and three-year-old children, and therefore these tasks would be placed in quadrant B (context-embedded, cognitively demanding). However, as mastery of these skills develops, tasks involving them would move from quadrant B to quadrant A since performance becomes increasingly automatized and cognitively undemanding. In a second language context, the same type of developmental progression occurs.

The third requirement for a theoretical framework applicable to bilingual education is that it permit the developmental interrelationships between L1 and L2 proficiency to be conceptualized. There is considerable evidence that L1 and L2 proficiency are interdependent, i.e. manifestations of a common underlying proficiency (see Cummins, 1981). The evidence reviewed in support of the interdependence hypothesis primarily involved academic or "context-reduced" language proficiency because the hypothesis was developed explicitly in relation to the development of bilingual academic skills. However, any language task which is cognitively demanding for a group of individuals is likely to show a moderate degree of interdependence across languages. Also, other factors (e.g. personality, learning style, etc.) in addition to general cognitive skills are likely to contribute to the relationship between L1 and L2 and thus some cognitively undemanding aspects of proficiency (e.g. fluency) may also be related across languages.

As far as context-reduced language proficiency is concerned, the transferability across languages of many of the proficiencies involved in reading (e.g. inferring and predicting meaning based on sampling from the text) and writing (e.g. planning large chunks of discourse) is obvious. However, even where the task-demands are language-specific (e.g. decoding or spelling), a strong relationship may be obtained between skills in L1 and L2 as a result of a more generalized proficiency (and motivation) to handle cognitively demanding context-reduced language tasks. Similarly, on the context-embedded side, many sociolinguistic rules of face-to-face communication are language-specific, but L1 and L2 sociolinguistic skills may be related as a result of a possible generalized sensitivity to sociolinguistic rules of discourse.

In conclusion, the theoretical framework appears to permit the complexity of L1-L2 relationships to be conceptualized at the same time as it provides a more adequate rationale for the essentially simple point that academic skills in L1 and L2 are interdependent. The framework also provides the basis for a task-analysis of measures of "language proficiency" which would allow the relationships between language measures and academic performance to be predicted for any particular group of individuals. In general, the more context-reduced and cognitively demanding the language task, the more it will be related to achievement. However, although there are intrinsic characteristics of some language tasks which make them more cognitively demanding and context-reduced, these task characteristics must be considered in conjunction with the characteristics of the particular language users (e.g. L1 and/or L2 proficiency, learning style, etc.). For example, skills that have become automatized for native speakers of a language may very

well be highly cognitively demanding for learners of that language as an L2. Thus, we would expect different relationships between achievement and certain language tasks in an L1 as compared to an L2 context.[5]

Assessment of Entry and Exit Criteria Revisited

The theoretical framework can readily be applied to the issue of the assessment of entry and exit criteria. The problem highlighted earlier was that if language minority students manifest proficiencies in some context-embedded aspects of English (quadrant A), they are often regarded as having sufficient "English proficiency" both to follow a regular English curriculum and to take psychological and educational tests in English. What is not realized by many educators is that because of language minority students' ESL background, the regular English curriculum and psychological assessment procedures are considerably more context-reduced and cognitively demanding than they are for English-background students. In other words, students' English proficiency may not be sufficiently developed to cope with communicative demands which are very different from those of face-to-face situations.

What assessment procedures should be used for entry and exit in bilingual programs? Given that the purpose of language proficiency assessment in bilingual education is *placement* of students in classes taught through the language which, it is assumed, will best promote the development of academic skills, it is necessary that the procedures assess proficiencies related to the communicative demands of schooling. However, in order to be valid, the procedures should also reflect children's previous experience with language. Because the child's language experiences prior to school have been largely in context-embedded situations, the assessment procedures for entry purposes should involve cognitively demanding context-embedded measures which are fair to the variety of L1 (and L2) spoken by the child. However, for exit purposes, it is recommended that cognitively demanding context-reduced measures be used because these more accurately reflect the communicative demands of an all-English classroom. If children are unable to handle the context-reduced demands of an English test, there is little reason to believe that they have developed sufficient "English proficiency" to compete on an equal basis with native English-speaking children in a regular English classroom.

These suggestions derive from a theoretical analysis of the relationships between language proficiency and academic performance and clearly require empirical confirmation. However, without a theoretical framework for conceptualizing these relationships, legitimate empirical questions cannot even be asked. An example of a commonly posed empirical question which is essentially meaningless when asked in a theoretical vacuum is the issue of the relationship between "oral language proficiency" and reading. Within the context of the present framework, "oral language proficiency" could equally refer to cognitively undemanding context-embedded skills as to cognitively demanding context-reduced skills. As one would expect on the basis of the present analysis, there is little relationship between these two aspects of "oral language proficiency"; also, reading skills are strongly related to the latter, but unrelated to the former (see e.g. Cummins, 1981).

Two other applications of the present framework can be briefly noted. The first relates to language pedagogy. A major aim of schooling is to develop students' ability to manipulate and interpret cognitively demanding context-reduced text. However, there is considerable agreement among theorists (e.g., Smith, 1978) that the more initial reading and writing instruction can be embedded in a meaningful communicative context (i.e. related to the child's previous experience), the more successful it is likely to be. The same principle holds for L2 instruction. The more context-embedded the initial L2 input, the more comprehensible it is likely to be, and paradoxically, the more successful in ultimately developing L2 skills in context-reduced situations. A major reason why language minority students have often failed to develop high levels of L2 academic skills is because their initial instruction has emphasized context-reduced communication insofar as instruction has been through English and unrelated to their prior out-of-school experience.

In summary, a major pedagogical principle for both L1 and L2 teaching is that language skills in context-reduced situations can be most successfully developed on the basis of initial instruction which maximizes the degree of context-embeddedness, i.e. the range of cues to meaning. For language minority students, this principle implies a language experience, rather than phonics approach, to initial reading instruction (whether L1 or L2).

A *third* application is related to the influence of home preschool experiences on children's acquisition of literacy skills in school. Wells (1981), in a ten-year longitudinal study, has identified two broad types of communicative activities in the home which strongly predict the acquisition of reading skills in school. One is the extent to which there is "negotiation of meaning" (i.e. quality of communication) between adults and children; the other is the extent to which literacy-related activities are promoted in the home (e.g., reading to children). There is no clear-cut relationship between SES and the former, but a strong relationship between SES and the latter.

[5] It should be pointed out that the framework in no way implies that language pedgagogy should be context-reduced. There is considerable evidence from both first and second language pedgagogy (e.g. Smith, 1978; Swain 1978) to support the principle that context-reduced language proficiency can be most successfully developed on the basis of initial instruction which maximizes the degree of context-embedment. In other words, the more instruction is in tune with the experience and skills the child brings to school (i.e. the more meaningful it is), the more learning will occur. This is one of the reasons that bilingual education is, in general, more successful than English-only programs with language minority students.

These results have two clear implications in terms of the present framework. First, the strong relationship observed between both literacy activities and negotiation of meaning in the home and the later acquisition of reading in school supports the principle proposed above that context-reduced communicative proficiency can be successfully developed on the basis of prior context-embedded communication, or to put it another way, the more opportunity the child has to process comprehensible linguistic input (Krashen, 1981) and to negotiate meaning, the greater the range of input which will become comprehensible.

The second implication of Wells' findings is that many low SES students experience greater initial difficulties in school literacy acquisition because the school makes little or no pedagogic adjustments for the fact that these students may have had relatively few literacy-related experiences prior to school. Children who have had extensive prior context-embedded literacy-related experiences would seem to be in a better position to handle a more phonically oriented (i.e. less meaningful) approach than children who lack such experiences. (This, however, is not to imply that such an approach is desirable for these children.) It would thus seem especially important to make low SES and/or ESL students' initial exposure to literacy in school as meaningful and context-embedded as possible, perhaps through a language experience or similar approach. Although most middle-class children may be able to handle more context-reduced approaches to teaching reading, it is possible as Smith (1978) suggests, that this may not be the case for children who develop learning disabilities. In other words, it is possible that at least some of these "disabilities" are pedagogically induced.

The *fourth* application derives from a large number of studies showing strong relationships between L1 and L2 academic skills (e.g., Cummins, 1981). In terms of the framework, we would say that context-reduced cognitively demanding aspects of language proficiency are interdependent across languages. Thus, an immigrant child with well-developed skills in his/her L1 will likely develop academic skills in L2 to a similar degree. Assessment of a child's L1 proficiency clearly has important implications for placement.

In summary, the major reason for the confusion in regard to assessment procedures for entry and exit criteria in bilingual education is that neither the construct of language proficiency itself, nor its relationship to the development of cognitive and academic skills has been adequately conceptualized. The extreme positions (1) that language proficiency is essentially independent of cognitive and academic skills, implied by some sociolinguists, and (2) that language proficiency is largely indistinguishable from cognitive and academic skills, suggested by much of the psychometric research reviewed by Oller and his colleagues, arbitrarily identify particular aspects of the construct of language proficiency with the totality of the construct. In the present paper, it has been argued that language proficiency cannot be conceptualized as one static entity or as 64 static entities. It is constantly developing along different dimensions (e.g. grammatical, sociolinguistic, discourse and strategic dimensions) and being specialized for different contexts of use among monolingual English-speaking, as well as language minority, children. In academic contexts, certain aspects of language proficiency develop in specialized ways to become the major tool for meeting the cognitive and communicative demands of schooling.

A major implication of the present framework is that recognition of the very different communicative proficiencies required of children in school encounters as compared to the one-to-one, face-to-face interaction typical of out-of-school contexts is a first step towards the development of theoretically and empirically viable entry and exit procedures.

References

Burt, M.; Dulay, H. & Hernandez-Chavez, E. *Bilingual Syntax Measure.* New York: Harcourt, Brace, Jovanovitch, 1975.

Canale, M. "From Communicative Competence to Communicative Language Pedagogy." In J. Richards & R. Schmidt (Eds.) *Language and Communication.* New York: Longman, 1981.

Canale, M. and Swain, M. "Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing." *Applied Linguistics*, 1980, *1*, 1-47.

Cazden, C.; Bond, J.; Epstein, A.; Matz, R.; Savignon, S. "Language Assessment: Where, What and How." *Anthropology and Education Quarterly*, 1977, *8*, 83-91.

Cummins, J. "The Influence of Bilingualism on Cognitive Growth: A Synthesis of Research Findings and Explanatory Hypotheses." *Working Papers on Bilingualism*, No. 9, 1-43, 1976.

Cummins, J. "Educational Implications of Mother Tongue Maintenance in Minority Language Groups." *The Canadian Modern Language Review*, 1978, *34*, 395-416.

Cummins, J. "Linguistic Interdependence and the Educational Development of Bilingual Children." *Review of Educational Research*, 1979, *49*, 222-251.

Cummins, J. "The Cross-Lingual Dimensions of Language Proficiency: Implications for Bilingual Education and the Optimal Question." *TESOL Quarterly*, *14*: 175-187, 1980a.

Cummins, J. "The Exit and Entry Fallacy in Bilingual Education." *NABE Journal, 4*, 25-60. 1980b.

Cummins, J. "Psychological Assessment of Minority Language Children." Research Report, OISE. 1980c.

Cummins, J. "The Role of Primary Language Development in Promoting Educational Success for Language Minority Students." In California State Department of Education, *Schooling and Language Minority Students: A Theoretical Framework*. Los Angeles: Evaluation, Dissemination and Assessment Center, 1981a.

Cummins, J. "Age on Arrival and Immigrant Second Language Learning in Canada: A Reassessment." *Applied Linguistics*, 1981, *2*, 132-149.

DeAvila, E.A. & Duncan, S.E. *Language Assessment Scales — LAS I*. (2nd ed.) Corte Madera, CA: Linguametrics Group, Inc., 1977.

DeAvila, E.A. and Duncan, S.E. "A Few Thoughts about Language Assessment: The LAU Decision Reconsidered." *Bilingual Education Paper Series*, National Dissemination and Assessment Center, Vol. I, no. 8, 1978.

Dieterich, T.G.; Freeman, C.; and Crandall, J.A. "A Linguistic Analysis of Some English Proficiency Tests." *TESOL Quarterly*, 1979, *13*, 535-550.

Donaldson, M. *Children's Minds*. Glasgow: Collins, 1978.

Duncan, S.E. and DeAvila, E.A. "Bilingualism and Cognition: Some Recent Findings." *NABE Journal*, 1979, *9*, 15-50.

Herbert, C.H. *Basic Inventory of Natural Language* (BINL). San Bernadino, CA: Checkpoint Systems, 1975.

Hernandez-Chavez, E.; Burt, M. & Dulay, H. "Language Dominance and Proficiency Testing: Some General Considerations." *NABE Journal*, 1978, *3*, 41-54.

Krashen, S. *Second Language Acquisition and Second Language Learning*. New York: Pergamon Press, 1981.

Kessler, C. & Quinn, M.E. "Positive Effects of Bilingualism on Science Problem-Solving Abilities." In J.E. Alatis (ed.) *31st Annual Georgetown University Round Table on Languages and Linguistics*. Washington, D.C.: Georgetown University Press, 1980.

Labov, W. *The Study of Nonstandard English*. Champaign, Illinois: NCTE, 1970.

Mace-Matluck, B.J. "A Longitudinal Study of the Oral Language Development of Texas Bilingual Children (Spanish-English): Findings from the First Year." Paper presented at the National Conference in the Language Arts in the Elementary School. San Antonio, TX. March 1980.

Mercer, J. *Labelling the Mentally Retarded*. Berkeley: University of California Press, 1973.

Oakland, T. *Psychological and Educational Assessment of Minority Children*. New York: Brunner/Mazel, 1977.

Oller, J.W., Jr. *Language Tests at School*. New York: Longman, 1979.

Oller, J.W., Jr. "Language Testing Research (1979-80)." In R. Kaplan (ed.) *Annual Review of Applied Linguistics*. Rowley, Mass.: Newburn House, 1981.

Oller, J.W., Jr. and Perkins, K. *Research in Language Testing*. Rowley, Mass. Newbury House, 1980.

Oller, J.W., Jr. & Streiff, V. *The Language Factor: More Tests of Tests*. The University of New Mexico, Mimeo, in press.

Rivera, C. (Ed.) *Language Proficiency and Academic Achievement*. Washington, D.C.: Center for Applied Linguistics, in press.

Rosansky, E. "Future Perspectives on Research in Oral Language Proficiency Assessment." Paper presented at the InterAmerica Symposium on Language Proficiency Assessment, Airlie House, Virginia, March, 1981.

Shuy, R.W. "On the Relevance of Recent Developments in Sociolinguistics to the Study of Language Learning and Early Education." *NABE Journal*, 1979, *4*, 51-71.

Skutnabb-Kangas, R. & Toukomaa, P. *Teaching Migrant Children's Mother Tongue and Learning the Language of the Host Country in the Context of the Sociocultural Situation of the Migrant Family*. Helsinki: The Finnish National Commission for UNESCO, 1976.

Smith, F. *Understanding Reading* (2nd edition) New York: Holt, Rinehart and Winston, 1978.

Snow, C.E. & Hoefnagel-Hohle, M. "The Critical Period for Language Acquisition: Evidence from Second Language Learning." *Child Development*, 1978, 49, 1114-1128.

Swain, M. "French Immersion: Early, Late or Partial?" *The Canadian Modern Language Review*, 1978, *34*, 577-586.

Toukomaa, P. & Skutnabb-Kangas, T. *The Intensive Teaching of the Mother Tongue to Migrant Children of Preschool Age and Children in the Lower Level of Comprehensive School.* Helsinki: The Finnish National Commission for UNESCO, 1977.

Ulibarri, D.; Spencer, M.L. and Rivas, G.A. "Language Proficiency and Academic Achievement: A Study of Language Proficiency Tests and Their Relationship to School Ratings as Predictors of Academic Achievement." *NABE Journal*, Vol 3, 1981, 47-80.

Vellutino, F. *Dyslexia: Theory and Research.* Cambridge; M.I.T. Press, 1979.

Wells, G. *Learning through Interaction: The Study of Language Development.* Cambridge University Press, 1981.

ARE WE TESTING FOR INTELLIGENCE OR LANGUAGE?

John W. Oller, Jr.
University of New Mexico

In the first part of this decade, at least three books will appear which deal principally with the topic of language testing issues.[1] During 1981, the British Council sponsored *Issues in Language Testing* edited by J. Charles Alderson and Arthur Hughes, and the U. S. Department of Education and the Illinois State Board of Education sponsored *Issues of Language Assessment: Foundations and Research* edited by Stanley S. Seidner. In 1982 yet another volume, *Issues in Language Testing Research*, is slated to appear under the Newbury House imprint. Among the unresolved questions in all three volumes, is the one posed here: Are we testing for intelligence or language?

What is language? What is intelligence?

On the definition of language proficiency, there seems to have been some progress toward acceptable theoretical models — see Cummins (this volume, and also in press); Bachman and Palmer (in press); Canale (in press) and their references. There is some agreement now that both holistic and analytic models of language proficiency are useful.

However, on the differentiation of language proficiency from intelligence, there seems to be less agreement (see Carroll, in press). There are at least two aspects to the problem. On the one hand, there is the theoretical issue of saying what either of the constructs at stake actually *is*, and on the other hand, there is the question of what is *measured* by tests intended to quantify one or the other construct. The theoretical and the measurement aspects of the problem are related, but are not identical. It is conceivable that intelligence could *be* distinct from language proficiency deep within the human psyche, and yet, that tests aimed at intelligence might still largely *measure* acquired language proficiency. Another possibility, which seems intriguing, is that intelligence itself may be dependent on a kind of grammatical system — a logic of experience. This system might be distinct from any particular language, and yet, be dependent on the development of proficiency in *some* particular language in order to mature normally. Other possibilities have also been suggested and superior options no doubt remain to be worked out. As Carroll (in press) suggests, there is a good deal of relevant research, but it does not yet justify a definitive demarcation of intelligence and language proficiency.

Therefore, setting aside the measurement aspect momentarily, it may be useful to consider the theoretical question. To begin with, three speculative premises are put in the form of "what-if" postulates. With these premises in mind, some evidence is considered on the nature of intelligence and its relation to language processes.

Three What-If's...

(i) What if cognition itself were dependent on a kind of language system? Call this system *Basic* and say it is a language for representing and interpreting the events of experience.[2] Suppose we assume for the sake of argument that Basic is preprogramed and matures on a biological schedule along the lines of Chomskian "innate ideas." Basic might be construed as a grammar specifying the structure of texts of experience.

To illustrate how such a grammatical system might work, consider the following example. Imagine that you are in a certain American embassy in the Middle East. You work there. You are aware of the crowds that are milling around outside. The noise and tension is building. Suddenly, a window high above your desk shatters into a spray of fragments. A brick hurtles toward your head. You duck. The brick clatters to the floor...

It may be argued that various textual structures are taken into account in order for you to react intelligently by ducking your head. The immediate proposition which presents itself is the brick crashing through the window and proceeding on a trajectory that will pass through the space now occupied by your head. You are suddenly aware that the brick (a kind of grammatical subject in propositional logic) has just broken the window (a direct object) and is proceeding (a transitive

---

[1] This chapter is based on study undertaken during academic 1979-80 while the author was on sabbatical leave from the Department of Linguistics at the University of New Mexico. The material contained here has been used as the basis for several lectures. These included talks at the Annual Meetings of Colorado TESOL in September, 1980; the New York ESOL and Bilingual Education Associations in October, 1980; the Ontario Institute for Studies in Education in February, 1981; New Mexico State University in Las Cruces during the month of April, 1981; and the American University in Washington, D. C., as the fourth in the Hugo J. Mueller Lecture Series, in April, 1981. Thanks are expressed especially to Mark Clarke and John Haskell for their comments and encouragement which spurred the completion of the *Language Learning* (December, 1981) version under the title, "Language as Intelligence?" Other readers or listeners whose comments have helped to shape the final product have included Jack S. Damico, G. Richard Tucker, Mary Ann Hood, Frank Smith, and Helga Delisle.

[2] Of course, this is not to say that the language of experience is *basic* in the sense of being the lowest order system. Clearly it depends on the existence of many lower order systems which feed information into it. Basic presupposes the existence of a body and all of its sensory-motor systems. It also anticipates a great deal about the nature of event structures external to the organism. It both specifies and takes into account the knowledge base that is progressively materializing over the temporal course of experience. In this sense, ordinary experience may be construed as a very linear text, or better, a series of texts, which come to the organism over the course of time. The interpretation of this series of texts is a sort of reading process which, however, is *not* perfectly linear in nature.

predicate associated with the subject) toward your own head (another potential direct object). You anticipate the consequence that the brick is about to hit you in the head (an undesirable proposition) which you negate by removing your head from the trajectory of the brick.

The point of the example is to suggest that perhaps the meaningfulness of the actions and interpretations of mundane experience may be contingent upon a grammatical system which utilizes textually complex relations. Among them would be subject-predicate relations, negation, and conjunction as well as textual operations of inference concerning antecedent propositions (the fact that someone threw the brick) and consequent propositions (the brick is about to hit you in the head, but if you duck, it will miss).

All of this no doubt could be made much more elaborate and explicit, but the point is simply to try to show the surprising plausibility of the idea that cognition may be based on a deep language system — not any particular language (e.g., Chinese, Russian, English, Navajo, or what have you) — but on a universal logic of experience which matures on a biological timetable given sufficient nutrition and normal exposure.

(ii) What if grammars of particular languages (English, Spanish, Chinese, etc.) were mapped into the texts of Basic (that is, the language of experience) via a kind of *translation* process. Basic would underlie the intermediate particular language system and would link that language to the event structures (i.e., what we have termed the "texts") of experience. In its turn, the intermediate particular language system would underlie the manifest forms appearing ultimately as utterances or texts in that natural language system (English, or whatever).[3]

(iii) What if "non-verbal intelligence" were actually some aspect of the capacity to negotiate texts in Basic, while "verbal intelligence" were more closely linked to the intermediate grammar of a particular language and its manifest forms (speech, writing, or sign)?[4] Over the course of development, Basic could be expected to develop along with the acquisition of proficiency in the particular language system.

If these three uncommon assumptions were allowed, just for the sake of argument, an act of intelligence could be defined as a representational problem of fitting texts either to the event-structures of experience or to other texts. For instance, Basic texts are fitted to the facts of experience, and texts of a particular language are similarly fitted (or in Piaget's term, "equilibrated") to Basic. In the case of verbal acts, we might say that the facts or events manifest in experience are represented in Basic and then may be translated via a particular language into manifest forms such as utterances or written surrogates.

It follows that a criterion of "goodness of fit" is logically bound up in this definition of intelligence. The negotiation of event-structures is "intelligent" just to the extent that the implicit and explicit propositional values fit and are appropriate to the facts of experience. That is, action can be judged to be intelligent only, it would seem, in relation to some notion of truth, validity, correctness, functionality, or value.

How Do You Eat an Elephant?

The question that we started out with now becomes more problematic than before. To illustrate how, it may be useful to consider the riddle posed in the immediately preceding subhead: How *do* you eat an elephant? The answer is so obvious that it may easily be overlooked: You eat an elephant one bite at a time.

The riddle is instructive in at least two ways. For one thing, it reveals something significant about the temporal nature of experience and language. Experience comes to us in small bites, moment by moment, day by day. Language too is highly sequentialized and broken up into units consisting of sequences of subunits and so forth. So, we eat our respective elephants just one bite at a time.

However, the riddle can also be instructive in another way. It illustrates something about the process of fitting text to the facts of experience. In it we gain some idea of what is entailed by the notion "goodness of fit."

For instance, why did the riddler choose an elephant? Why not a mountain?

The riddle would have lost something. For one thing, mountains are not normally edible. Besides, the riddler knew that humans are generally carnivorous, that they kill for food. He understood that his audience knew English and knew about elephants being edible. He also was aware of the fact that most English speakers don't usually eat elephants. Thus, they might well be decoyed away from the correct answer by the problem of how to prepare elephant meat — something they would not likely have ever done.

---

[3] Here and throughout this paper the term "text" is used in the sense of an "actual" system, rather than a "virtual" system (see De Beaugrande, 1980). A virtual system merely has potential for realization in experience, while a text actually occurs in the stream of real experience. De Beaugrande and others (e.g., see John Dewey, 1916) argue that actual systems are interpretable and meaningful in ways that virtual systems cannot be. Note that this is radically at odds with the earlier claims of some linguists who argued that a sentence, for instance, could not have any meaning in a context that it did not have in isolation. To the contrary, the argument for viewing texts as actual rather than virtual systems suggests that temporally progressive experience invests elements of text with meaning that they could not have apart from that particular and unique *real* context.

[4] The term "nonverbal intelligence" as it is used here is intended to refer to what exists, without assuming that it can be measured.

Hmmmmm. So, elephants are not commonly used for food by the audience. Neither are grasshoppers, so why not a grasshopper?

It just wouldn't make as good a riddle. For one thing, grasshoppers are not large enough to be a very challenging quarry, neither are they a great challenge to eat—at least not because of their size.

Well, then, why not ask: How do you eat a truckload of grasshoppers? One grasshopper at a time. Gulp!

It's shocking in a way, but it misses the riddler's point which was something to the effect that a gigantic problem can be dealt with in small phases.

Isn't there also some shock value in the idea of eating an animal that is often used as a beast of burden? The riddler could have achieved a similar effect by choosing a horse, but horses seem far less formidable, and to most English speakers, eating one seems a little too close to eating a dog or cat. The idea borders on cannibalism.

Doesn't the choice of an elephant also bring with it a whole mystique that would otherwise be lost? Doesn't it call to mind scenes of African lions, dark jungles, Bengal tigers, and Indian gurus?

And what about the existing lore concerning elephants? Consider the expression, "She has a memory like an elephant." Or what about the well-worn parable of the blind men and the elephant? And don't forget the elephant riddles of the last two decades (e.g., "How do you hide an elephant in a cherry tree?" Or "How do you get four elephants in a Volkswagen?").

So much to provide just a glimmering of the significance of the notion "goodness of fit."

Is it possible that intelligence is really a characteristic of symbol systems capable of generating texts which fit the facts of experience? Jean Piaget observed some decades ago that "every definition of intelligence comes sooner or later to lean on biology or logic" (1947, p. 3). The approach here leans heavily on biology and also on a logic of natural experience — the sort that Dewey (1916) had in mind in his *Essays on Experimental Logic*. It is also possible to make a psychometric case for the definition of intelligence considered here, but this is attempted in another work (Oller and Streiff, in press). With all of the foregoing in mind, some of the genetic, cytological, and neurological evidence bearing on the proposed definition will first be considered. Then, having sketched in some of the evidence supporting the proposed theoretical framework, the measurement problem will be dealt with briefly through an analysis of a few sample items drawn from "non-verbal" and "verbal" IQ test items.

Biological Evidence

According to a biological approach, perhaps the primary characteristic of intelligence is *adaptability*. Piaget (who incidentally began his career as a biologist) describes the capacity of an organism to "assimilate" its environment to itself and to "accommodate" itself to the environment wherever change is needed and the environment itself is unyielding. He refers to the whole process as "equilibration."

Nowadays, biologists are turning increasingly to the texts which make this sort of intelligent action or equilibration possible. In 1977, the eminent biologist, Brian Clark, commented that the deciphering of the "genetic code" in the mid 1960s was "one of the most significant advances in this century in the biological sciences" (p. iii). It had been known for several decades that proteins played a crucial role in the development of living things, but it was not until the 1960s that the structure of proteins was systematically linked to the code written in the DNA molecules of cell nuclei. Since then, biologists have come to refer to the macro-molecules of DNA and also to the proteins as "texts."

Is it possible, therefore, that there might be a language basis for the existence and structure of organisms? Could the order we see in living organisms be dependent on a language system?

Martynas Ycas, another eminent biologist, wrote in 1969:

> It is now both useful and commonplace to regard the proteins as a topologically linear 'text' written in a 20-symbol alphabet, the amino acids. This text, in turn, is specified by (or encoded in) the nucleic acids (DNA and RNA), another text whose alphabet has four symbols, the bases. Some of these nucleic acids are the self-replicating genetic material, and thus the information specifying proteins is transferred from generation to generation. The coding problem is to determine how one text specifies the other (p. 1).

Shortly after the cracking of the code in 1966, another noted biologist, Carl Woese, wrote:

> The cell has a certain vocabulary of *words* . . . with which to construct *messages* in nucleic acid *language*. These messages are then translated into amino acid language by means of a dictionary or code book . . . The cell seems to employ a *tape-reading* process in the synthesis of protein (his italics; 1967, p. 4).

Additional biological evidence for some sort of language system comes from observations of communications that occur within cells and between them and the external world. How are such communications possible? How, for instance, does a cell know its functions? How is its equilibrated knowledge coded, stored, retrieved, and utilized?

For example, consider the flight of an amoeba from light or heat. When the source of the disturbance is introduced into the medium, the amoeba does not tend to move toward it or even parallel to it, but away from it. In some manner, it seems, the amoeba takes account of a state of affairs external to itself, adjusts its internal states, and thus alters its relation to the external state so as to restore a condition more favorable to its continued existence. To describe what happens inside the organism in purely mechanical (physico-chemical) terms does not really do away with the most difficult aspect of the problem. Even if a mechanical basis is found for the movements of the organism, there is still the problem of how the internal events are set in systematic correspondence with external states of affairs so as to enhance the organism's likelihood of survival.

Of all people, the author of the precursor of the modern intelligence tests, Alfred Binet, made an interesting foray into this area as early as 1888. While looking for certain of his writings on intelligence testing, I stumbled onto a book entitled, *The Psychic Life of Micro-Organisms*, published in 1888.[5]

In that remarkable little book, Binet claimed, among other things, that single-celled organisms lead a surprisingly complex mental life. For instance, he argued that they possess mental capabilities that enable them to display emotions such as fear, sexual arousal, memory, and instinct.

To illustrate the capacity for fear, he wrote:

> There is not a single ciliate infusory that cannot be frightened . . . If a drop of acetic acid be introduced beneath the glass slide, in a preparation containing quantities of infusoria, the latter will at once be seen to flee from all directions like a flock of frightened sheep (p. vi).

Perhaps his most interesting argument is found in his observations about the sexual behavior of micro-organisms. He described the process of courtship, pursuit, and union in paramecia as having all flavor of similar rituals in higher organisms, including the game of "hard-to-get." The chase, as he described it, ends in a "blending" of the two organisms "at an important part of their persons in a very intimate fashion" (p. 69). Although Binet could not have known it in 1888, the organisms were actually exchanging their genetic texts. He did see this process, however, as a microcosm of what higher organisms perform in their own sphere.

He quoted Balbiani (a famed biologist of that era) who said:

> I believe that the spermatozoids do not move about blindly, but . . . act in obedience to a kind of inner impulsion, to a sort of volition which directs them to a definite object (p. 78).

If this is the case, are we not moved to ask on what basis the sperm cells know what they are after? How is the plan coded inside the organism? How does the organism take account of things external to itself so that it knows when to start executing its plan and when the plan has been culminated? Could the plan of the sperm cells be simpler than the sort of logic underlying propositions that relate agents to objects via certain intermediate actions? (On human goals and plans, see Schank and Abelson, 1977).

Binet was not unaware of the considerable problem which his observations inevitably brought to light. What sort of internal mechanism would explain the complex communications that he was led to infer? He put the difficulty like this:

> What would be necessary to explain, is how and in consequence of what mechanism of structure, one form of molecular movement, corresponding to a given excitation, is followed by a certain other form of molecular movement corresponding to an act likewise determined (p. 65).

Binet correctly anticipated the prevailing inclination of present day theorists toward a mechanistic explanation. However, he seemed to be arguing that such an explanation would not succeed in eliminating the need for a rationalistic account. Why else did he title his book *The Psychic Life of Micro-Organisms?* (See note 5.) Binet observed that the nucleus is "an essential factor in the cell's vitality" (p. 100), a seat of government from which the protoplasm receives its "delegation of physiological powers" (p. 101). His remarks raise images of politics as much as of biology, and they suggest that even the simplest one-celled organisms are like miniature states requiring communication, transportation, and defense.

Certainly, some significant governmental problems exist between the cells of multicellular organisms. How do the various types of cells communicate with each other? For instance, when you cut your finger, how are the orders issued to reconstruct the damaged tissue? How do the cells cooperate in transporting the necessary materials to the site to do the reconstruction? And, in general, how are the differences in function established and maintained across the numerous cells of any complex organism?

In a popular article, "Pattern Formation in Biological Development," Lewis Wolpert noted that "there is a

[5] Actually, *mental* would probably be a better modern translation of the French term *psychique*, thus changing the title to *The Mental Life of Micro-Organisms*

widespread belief that cells have complex conversations with each other during development" (1978, p. 167). However, he hastened to point out that, for his own part, he believed the conversations to be relatively brief, and possibly even "boring." Nonetheless, he admitted that "a large number of experiments indicate that most patterns arise as the result of cell to cell interaction" (p. 164).

The question, as Wolpert noted, is what kind of physico-chemical mechanisms is employed in intercellular communication? The fact that such communication takes place is not questioned. One of the things that must be communicated from cell to cell during embryological development is information about location in the developing organism. Apparently biologists believe that this sort of information is crucial to the differentiation of cells and the formation of the various organs of the body.

The mechanism for intercellular communications, however, remains a mystery. Is it conceivable that any basis could exist which will not entail textual complexities of the sort which are inferred for intracellular communication? Does the genetic basis of life suggest anything other than some sort of language for intercell communication? And what about the case of biological organization in higher organisms?

More specifically, what about man?

At least since the germinal work of Karl Lashley (1951), it has been known that human cerebral functions exhibit certain text-like properties. Lashley pointed out that a movement as simple as "reaching and grasping" is dependent on the sort of hierarchical organization characteristic of the intricate structuring of sounds, syllables, words, phrases, and clauses of discourse. In his *Programs of the Brain* (1978), the Oxford biologist, J. Z. Young writes:

> To understand the language of the brain it is necessary to know how the nerve cells combine, like letters or the phones of speech, to produce units that have meaning, like words . . . If grammar is the system that regulates the . . . use of language, we might say that the brain operates a sort of metalanguage with a metagrammar, which regulates . . . the conduct of life, including speech (p. 46).

How is it that the temporal coordination of hierarchical sequences is accomplished? Even the simplest sorts of actions — e.g., tying shoe laces, buttoning a shirt, opening a door, starting a car — all involve delicately orchestrated sequences of events. Returning to the "what-ifs" stated at the outset, it might be proposed that the ordering of events is accomplished via a kind of universal language system. At times there may be interaction between this universal Basic system and the grammar of a particular language, but there are good reasons to suppose that this is not always necessary. The question then is, how does the brain relate holistic impressions with highly grammatically structured sequences of events? Is it possible that the differentiation of functions in the two hemispheres of the brain may shed some light on this process?

With respect to hemispheric specialization, Bogen, a neurosurgeon and Professor of Medicine at the University of Southern California, writes:

> the right hemisphere is specialized for processing time-independent stimulus configurations and the left hemisphere for time-ordered stimulus sequences (1977, p. 141).

Of course, this does not mean that these functions are completely separated. However, the left hemisphere is better at organizing and interpreting hierarchically arranged sequences of words, numbers, or other categories, while the right is better at handling Gestalts, images, or complex totalities. It seems that the left hemisphere excels at taking things apart (analyzing), while the right excels at handling global patterns (synthesizing).

It has become popular to relate the different specializations of the two hemispheres to the most widely attested types of intelligence. "Verbal intelligence" is associated with the left hemisphere, while "non-verbal intelligence" is attributed largely to the right hemisphere. Verbal processing is highly temporalized. On the other hand, taking account of a Gestalt, or a global pattern, seems to require a certain simultaneity that must largely disregard the temporal aspect of elements within the whole.

The picture is complicated by the fact that both hemispheres seem capable of performing both types of function to some extent. Also, it seems that the linking of Gestalts to temporalized text-like structures is an integral aspect of ordinary experience.

Although the notion that brain functions must occur in particular sites of tissue became popular immediately after Broca's discovery of the so-called "speech center" in 1861, more recent research reveals more or less global brain activity during performance of demanding tasks.

Based on observations of blood flow to the various regions of the cortex in both hemispheres, Lassen, Ingvar, and Skinhoj (1978) conclude:

> It appears that for the brain to 'understand' the surrounding world, to perceive its meaning and to take action in difficult tasks, the cerebral cortex must be activated not only locally, but also totally (p. 71).

As Pribram (1971) has observed, there are many brain functions that seem to be distributed globally. He cites the case of our ability to write with the non-dominant hand. The question is, how can the information about writing which is presumably stored in the left hemisphere (which controls the right hand) be transmitted to the right hemisphere and from there to the left hand. Or, as Pribram asks, how is it that a

right-hander will be able to write his name with his left foot in bold strokes in the sand at the sea shore? How can specialized tasks be communicated in toto from one brain region to another?

In an attempt to solve this intriguing puzzle and others associated with it, Pribram proposes that the elusive "engram" (the brain's unit of memory) may actually be a wave form that has holographic properties. If Pribram's idea is correct, the brain represents the world in such a way that its coded information has a holistic (analogical) correspondence to experience, rather than a feature by feature (digital) correspondence. Among other things, Pribram's "holographic memory" would help to explain why it is that as much as 90% of the tissue associated with a particular brain function may be destroyed, and yet, the function may still remain intact.

But what then must become of the dichotomy that divides intelligence into the supposed verbal functions of the left hemisphere and the nonverbal of the right? The question seems to be whether the brain divides up the work in the supposed manner between left and right hemispheres, or whether this is not somewhat of an oversimplification.

It seems that many of the most common uses of language require both temporal (sequential analytic) and simultaneous (Gestalt-synthetic) operations. Aren't these logically aspects of the same textual processes? Imagine, for instance, what sorts of brain functions might be involved in describing how to row a boat. Or, consider what mental processes are involved in describing the stars in the sky on a particularly dark and clear night. Don't both of these uses of language require analytical *and* synthetic mental functions? They both seem to require handling of hierarchically sequentialized elements of language as well as the processing of holistic and fluid impressions of experiences. In fact, isn't it the case that the verbal descriptions must be pragmatically mapped onto the holistic impressions?

To account for such simple brain processes as these, it would seem that a theory is required that sees both "verbal" and "nonverbal" aspects of intelligence as parts of a more inclusive totality. It was apparently this sort of perspective which led UCLA brain scientist, Harry Jerison, to argue that "a nonverbal test of intelligence may be a contradiction in terms" (1977). In the same context he said, "language and language-related performances are inevitably dominant in most human performances" (p. 59).

Is it not possible that, in fact, experience itself may be essentially textual in nature? That is, in order for it to be meaningfully processed, could it perhaps be necessary to perceive the complex linkages of events in their apparent temporal progression through some sort of propositionally complex language system? Is it possible that the translation of nucleic acid language into protein language might actually provide a useful analogy for the way experience is interpreted through the use of natural language systems? Could talking about experience be a process like *translation* from one language into another? In fact, is it possible that talking about experience is more like translating between language systems than like relating a language system to something that is essentially nonpropositional in nature?

So much for biological evidence on the theoretical framework. To relate all of the foregoing to the problem of intelligence testing, we now turn to an examination of some typical "intelligence" test items.

What Do IQ Items Require?

A logical analysis of the structures required by tasks in so called "intelligence" tests would seem to have substantial bearing on the determination of what intelligence tests measure. It also bears heavily on what may be called "the language hypothesis" – the idea that nonverbal and verbal intelligence tests probably measure acquired language skills more than anything else (e.g., Oller and Perkins, 1978). One empirical approach to the testing of this hypothesis would be to go to the test-takers themselves and get them to introspect about how they are able to solve the problems in the tests. Roth (1978) used this approach. Another method is to examine the variance in tests by factorial methods (Carroll, in press; Bachman and Palmer, in press).

Still another method is to examine the items that are included in certain so called "intelligence" tests (Gunnarsson, 1978). The purpose of this sort of analysis would be to determine what mental operations must accompany the solution of commonly used types of items. The question is: *What sorts of internal representations are essential to the solution of typical intelligence test items, and what kinds of operations must be performed on those internal representations?* Possibly such an analysis will help to lead us beyond the claim of Arthur Jensen (1969) that intelligence is "what intelligence tests measure" (p. 8).

First we will examine some of the most commonly used "intelligence" test items of the nonverbal type, and then we will have a look at some "intelligence" items of the verbal type.

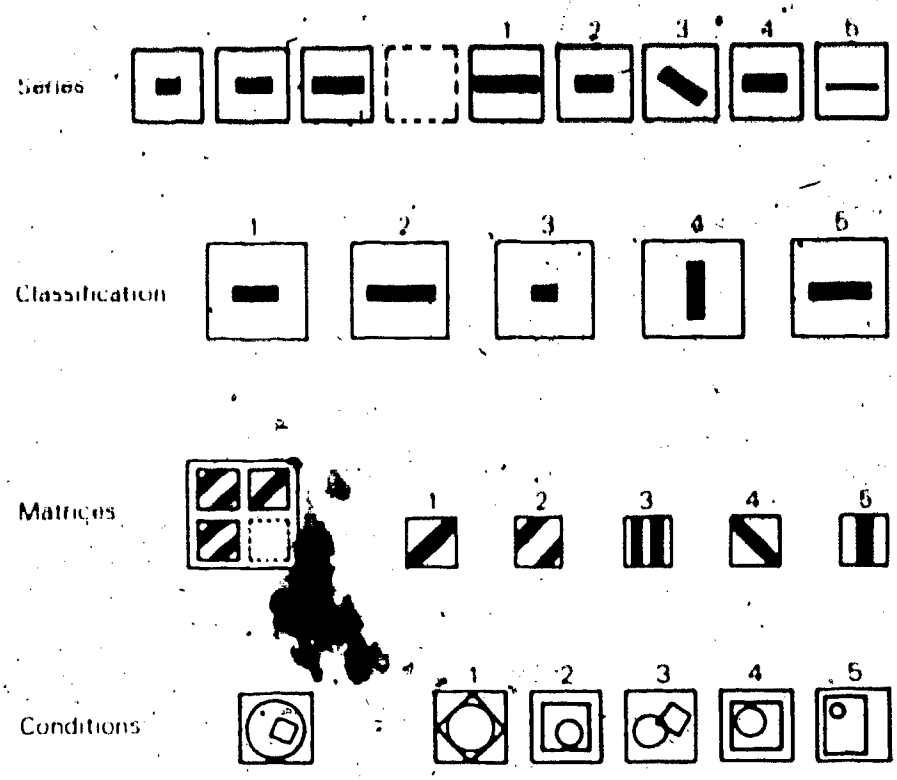
Nonverbal IQ Items. Among exemplary nonverbal IQ tests are Cattell's Culture Fair Intelligence Tests and Raven's Progressive Matrices. Jensen (1980) refers to these as nearly "pure" measures of "general, intelligence."

Figure 1 contains examples of the four item types that appear in the Cattell tests. The third row in the figure, labeled "Matrices," represents the sort of items appearing the Raven's tests as well. Thus, an examination of these four item types should provide a fairly representative idea what typical nonverbal items require. These samples come from a pamphlet prepared by the publishers of the Cattell tests to inform users

about the test and its applications. Since the items shown are fairly simple ones by comparison to those that actually appear in the tests, whatever mental processes are required by these should be equally necessary to the more complex items that actually appear in the tests.

Figure 1

Sample items from *An Introduction to IPAT (Cattell's) Culture Fair Intelligence Testing*



For each type of item, there is a separate set of instructions. For instance, in order to solve items of the "Series" type, it is necessary to select the alternative on the right that best completes the sequential pattern that is displayed on the left. The examinee must note, among other things, that the rectangular dashes on the left become longer as one progresses through the series from left to right. It may also be noted that the increase in length is constant from one dash to the next.

What sorts of mental operations are necessary in order to solve items of the "Series" type? In addition to understanding the instructions themselves, which may present some difficulty to the lower range of examinees, there are other textual operations that seem logically indispensable to the solution of such items. Setting aside the question of how the examinees distinguish between the lead-in material (the item stem) and the possible answers (the multiple-choice alternatives), may we not infer that they must internally represent the dashes on the left in some manner or other? Perhaps the representations remain visual throughout the solution of the problem, but can it not be argued even if this is so that these representations must nonetheless serve as propositional subjects with which certain predicates must be associated?

For instance, necessary logical predicates of the second dash include (it would seem) the fact that it is the second one in the series (i.e., that it is positioned to the right of the first), that it is longer than the first dash, that it is oriented on the horizontal, and that it has the same width as the first. And no doubt other facts can be adduced. Each of these facts can be understood as a proposition for which the subject-predicate relation is an essential ingredient.

The question is whether these facts can be "known" in any sense at all without some form of propositionally complex logic in which to represent them. In addition, there is the implicit proposition (or fact) that the first and second dashes are objects of the same type, and that they form a series of similar objects with the third dash. This requires that the "series" be treated as a logical subject which is distinct from its individual elements. Further, there is the proposition that the increase in length from dash one to two to three is relatively constant. From this fact it may be inferred that

the next item in the logical series must have the same properties as the preceding ones. In other words, it may be inferred that the same predicates that have applied to preceding subjects will apply to the next logical subject in the series. (A necessary presupposition for all of this would be the proposition that things will be orderly; i.e., that there is a logically discoverable answer to the test item.)

The item solver may rule out choice 2 because it is the same approximate length as the second dash in the series. He knows from the proposition that the dashes in the series become progressively longer from left to right that the correct answer must be longer than the third dash in the series. Choice 3 may be ruled out because the dash in that alternative is not oriented on the horizontal. Here it seems that an implicit negation is required (i.e., not horizontal). (Again, it is acknowledged that the mental operations may, in fact, be performed on visual images, but does this eliminate the necessity of deep textual structures?) Choice 4 can be eliminated because it does not satisfy the predicate of being longer than dash three in the series, and choice 5 is too narrow. It does not satisfy the predicate of being the same width as the other dashes in the series.

While it might be correct to argue that not *all* of the propositions contained in the preceding remarks are necessary to the solution of the series problem, can it be argued that *none* of them is required? On the contrary, would it not make sense to suppose that many textual forms may be necessary which we have not highlighted in the previous analysis? Moreover, is it not also probable that many if not all of the textual *structures* illustrated will be required for the solution of any such problem?

The second type of item in Figure 1 seems to be somewhat simpler than the first. The directions tell the examinee to indicate which of the pictured elements does not go with the rest. Which one does not fit? Again, the dashes, it would seem, are logical subjects with which certain predicates must be associated. In this example, the predicate of horizontal orientation is necessary to the solution. The propositional operation of negation is equally essential. The problem is to find which element does not share the predicate(s) shared by all of the others. There may be some other way to represent the problem, but is there any method of representation which does not implicitly contain the propositional complexities of predication and negation?

In the third type of problem, labeled "Matrices", Cattell was obviously influenced by the matrix format of the Raven nonverbal IQ test. The problem is to select the element from the several choices that best completes the matrix pattern. The matrix type of item is potentially a much more complex version of the "Series" type. With a matrix, there are more dimensions of pattern variation and hence a greater number of propositional structures are possible. In the item at hand, on the third row of Figure 1, the test taker may note that there are various progressions in the matrix which may help to determine just what the missing element must be like. There is in fact a progression from left to right in each row, and from top to bottom in each column. There is even a progression across the diagonals that may be of some help. The question, it would seem, is what predicates must be associated with each of the elements of the pattern in order to uniquely determine the missing element of the whole pattern? The order in which the progressions are noticed may not be too important, but isn't the fact that at least some of them must be noticed undeniable?

In the rather simple problem under consideration, it is sufficient to notice that the two patterns on the left side of the matrix are similar both in the direction of their cross-hatched lines, and in the number of those lines. From this it may be inferred that the ones on the right must also be similar, and that therefore the correct choice is the one which matches the pattern in the top right hand corner of the matrix. Thus, choice 1 is the correct answer. Other strategies would lead to the same result, but is there any conceivable strategy that does not depend on complex textual operations? That is, is it possible to solve such problems without performing mental operations that set up subject-predicate relations which are true of the patterns in question? Or is it possible to avoid the negation of certain predicates or the conjoining of textual meanings in complex ways?

In the fourth type of item, "Conditions," the task is more complicated than in the previous items. The problem is to note the relation(s) between the dot and the various other elements enclosed in the figure to the left and then find the figure on the right where the dot can be placed in the same relation(s) to similar elements. In the example item, the dot is inside the circle but outside the square. The only place where it can be placed in a similar relationship to a circle and square on the right is in Choice 3. In all of the other cases, the circle is completely contained within the square, and it is not possible to place the dot inside th circle without also placing it inside the square at the same time.

As with previous examples, we set aside the question of how the examinee represents the fact that the larger square is not to be considered part of the immediate problem, and all of the other propositions that are contained implicitly or explicitly in the instructions to items of this type. Nevertheless, it seems necessary for the problem solver to employ a vocabulary (perhaps in Basic) specifying such elements as DOT, CIRCLE, and SQUARE. It may not be necessary for these elements to be represented in the words of any particular natural language, but isn't a verbal representation highly likely?

Roth's research (1978) seems to support the conclusion that normal children do in fact use verbal representations for such problems. But what if they did not? What if the representations were entirely visual or in some other mental form, and not in words associated with a particular natural language? Would this eliminate the necessity of textual operations?

There must also, it seems, be an implicit (or explicit) vocabulary to distinguish the predicates INSIDE, NOT INSIDE (or OUTSIDE), and no doubt, others as the problems become more complex. Moreover, there are implicit superordinate operations such as NOT and AND. To solve the problem, the examinee, it would seem, must be able in some manner or other to represent internally propositions of the form, DOT INSIDE CIRCLE, as well as DOT NOT INSIDE SQUARE. Moreover, these propositions must be conjoined by some sort of AND operator to create a proposition that has the rough form, DOT INSIDE CIRCLE AND DOT NOT INSIDE SQUARE.

Such textual forms have to be fitted to the visually presented facts. Can we say that they must be pragmatically mapped onto those facts? Further, is it not reasonable to infer that the actual pragmatic mappings be *much* more complex than the ones sketched out here?

Verbal IQ Items. Some may argue that any analysis of "verbal" IQ items will be trivially relevant to the hypothesis that intelligence may be based in a kind of language system. However, there are some good reasons for looking closely at "verbal" IQ items. For one, if the language hypothesis is correct, it should be possible to demonstrate that in important ways "verbal" items are similar to "nonverbal" items in the textual operations that they require. For another, it should also be possible to examine the extent to which such items are distinct from so-called "language proficiency" items.

"Verbal" IQ items are, as Gunnarsson (1978) showed, sometimes excessively difficult to distinguish from items appearing in so-called "Reading" tests, and other categories of "Achievement" tests. Almost always, however, IQ items of the "verbal" type require the interpretation and/or production of words, or statements in a *particular language*. To this extent, it is already difficult to follow the logic that Jensen, Herrnstein, and others do in claiming that such tests primarily measure innate ability, rather than acquired language proficiency. It was the purpose of *Language in Education: Testing the Tests* (Oller and Perkins, 1978) to question that logic. In spite of some legitimate criticisms of the statistical methods used in that work, it seems that the objections raised there against certain claims of the IQ testers remain largely unanswered.[6]

At any rate, just what kinds of mental operations are required by IQ items of the "verbal" type? A typical item, one which appeared in the original Binet tests, requires the examinee to say what is wrong with a given statement such as the following:

> The judge told the prisoner: "You are to be hanged at dawn. Let this be a warning to you."

The correct answer is any statement to the effect that, "He can't be warned if he is dead." The crux of the problem is a conflict between the necessary consequence that hanging will kill the prisoner and the equally necessary fact that warning implies an opportunity to repeat whatever offense the prisoner has committed. Most adults immediately laugh upon hearing the item. But it costs a little thinking to explain why, and even more effort to see the extensive textual fabric on which our conclusion depends.

We may differentiate the network of inferentially related propositions into three categories: first, there are presuppositions related to prior events or states of affairs which are crucial if we are to associate any meaning at all with the asserted propositions; second, there are associated propositions which aid our comprehension, but which are neither before nor after the stated assertions; and, third, there are implications which are in some sense subsequent to (or consequent upon) the asserted propositions.

For instance, because the judge hands down a death sentence, we may infer that the prisoner has been convicted of a capital crime. This inference is not a necessary one, but it is a logical presupposition which normal adults will accept as relevant and probably correct. A more obviously necessary presupposition is that the prisoner has been convicted of committing some *serious* crime.

Associated textual meanings would have to include the fact that a prisoner is someone who is detained against his will in a societal institution known as a prison; that a judge is someone empowered by the society with certain authorities within its legal system. In fact, the presupposed propositions may also have associated propositions. For instance, the presupposition that the trial of the prisoner has already occurred when the sentence is being handed down is linked to the presupposition that the trial occurred in a courtroom, and that the verdict was probably reached by a jury. Associated with the sentence of hanging is the proposition that it involves placing a rope around the victim's neck, with the victim's hands restrained, and then, suspending the victim by the rope so that he will strangle or die of a broken neck. Implied propositional meanings would include the consequence that hanging kills the victim and that a warning implies an opportunity to repeat the previous offense.

No doubt the foregoing analysis leaves out much more of the textual network of associated meanings than it makes explicit. However, it does suggest that an IQ item of the type analyzed involves a good deal of textual reasoning of the very sort that goes on in the interpretation of ordinary discourse (for more on this topic, see Rumelhart, 1976; and Schank and Abelson, 1977). Is it not an open question whether there are any meanings at all which are not linked to the fabric of experience via the sort of textual network illustrated with respect to the hanging example?

---

[6] For those criticisms, see the various contributions to *Issues in Language Testing Research* (Oller, in press). Also see especially Chapter 22 of that volume where some of the criticisms are answered and others are accepted as valid.

Here is another example from the Binet-type verbal IQ test. What is wrong with the statement?

> In a train wreck, the cars at the end are usually the ones most severely damaged.

Surprisingly, perhaps, most adults not only do not laugh at this item (in contrast to their reaction to the one about the hanging), but in fact, they meet it with a sort of blank stare. Many do not see anything wrong with it at all on first hearing. However, a bit of thinking will show what the test authors had in mind.

To solve the puzzle, it is necessary to consider what usually happens in a train wreck, and then to ask whether it is likely that the cars in the back will be the ones to receive most of the damage.

In fact, trains travel on tracks. This is an associated proposition. That is, it is associated with the assertion about trains. Ordinarily, they travel in a forward linear direction. This too is an associated proposition which is a direct consequence of the linear nature of the tracks. What is more, it may be helpful in solving the problem. From this it can be inferred that the cars at the front are the ones which will logically receive the greater force of an impact if there is a collision.

This inference is an implication (by the definition given above) of the nature of trains and their usual linear motion in a forward direction. A further implication is that the cars at the back are therefore the least likely to be damaged, since they are the ones which are the most protected due to the cushioning effect of the intervening cars.

Of course, it is true that the cars at the end will receive the greatest impact if a stationary train is struck by another train from behind. However, the one striking it will tend to receive the most damage in its lead cars. And, in any case, to reason that the statement about damage to trains is correct or to reason on such a basis that it is incorrect demonstrates in both instances the need to map propositions into the episodic facts of experience. As in the hanging example, the problem solver requires access to much of the textually interpreted fabric of experience. There may be a multitude of strategies for working out these relationships, but this does not deny the importance of inferential access to them.

Here is one final example. In another typical item type, the examinee is presented with certain asserted information and then is asked to answer a question based on that information. For instance, the examinee might be told:

> Mary is taller than George but she is shorter than Sam or Harry. Who is the shortest?

Just what sort of mental computations are necessary to the solution of this problem? It seems simple enough on the surface, but there is more to it than immediately greets the analyst.

Consider first how a solution might be achieved through a visual strategy. Suppose we first imagine a female named Mary who is visibly taller than some male named George. (At this point, two propositional inferences about sex which go beyond the asserted information have already been made. We have also arrived at the proposition that both Mary and George are humans. (It is difficult to imagine how such inferential meanings might be derived from a strictly visual strategy, but let us continue.) Then, we may simply add two other male persons to our visual image, namely Sam and Harry. (Again, sex assignment is based on propositional inferences about the likely values of the two names in questions.) It is asserted that they are both taller than Mary. (However, a slight difficulty again arises for a strictly visual strategy since Sam and Harry are not compared with each other. How can we know for certain what height to assign to either of them? Though, it makes no difference to the problem at hand, height assignment to Harry and Sam would seem to be necessary to placing them in our visual image. Let's assume this is accomplished in some manner or other.) To answer the question, "Who is shortest?", we may now simply scan the image and pick out George.

Or, it would seem, a quicker nonvisual strategy would be to reason that since George is shorter than Mary and Mary is shorter than Sam or Harry, George must be shorter than Sam or Harry, and therefore, the shortest of the four. This inference works because height is a transitive relationship.

No doubt, there are other strategies that will arrive at the same correct solution, but are there any that do not rely heavily on inferred propositional meanings? In fact, isn't it helpful (perhaps even necessary for some problems) to know what the question is in advance of beginning to interpret the relevant facts? Otherwise, how is it possible to know what to do with the asserted propositional meanings? At least some problem solvers report that they need to have another pass at the asserted meanings after they know what it is they are looking for. For instance, consider how the problem would be changed if the question were, "Who is the next to the shortest person?", or "The third from the tallest?" Doesn't the need to know what the question is suggest that comprehension itself needs to be guided by some sort of propositionally formulated plan? (This would also appear to hold for the nonverbal item types discussed above.)

In addition to the reasoning already sketched out, there are some other textual inferences that are lurking in the background just out of view. For instance, we assume that the measurement of the persons in the given assertions is done with all of them in a standing position, legs straight, head erect, etc. Further, we assume that hats don't count, that shoes are taken off, and that it is unfair for anyone to stand on their tip-toes. We do not expect the measurement to be done with some of the persons in kneeling or sitting position while others are standing, and so forth. Beyond these presupposed meanings there must be a great deal of additional textual structure.

In the light of even this brief analysis, does it not seem clear that the solution of both "verbal" and "nonverbal" items may be dependent on the utilization of propositional meanings of considerable complexity? Is it not clear that presuppositions, associations, and implications must be supplied by the problem solver through abstract inferential reasoning? Further, doesn't all of this entail the utilization of some deep language system which has grammatical capabilities allowing such operations as predication, negation, conjunction, modification, and the like?

Educational Implications

With the three "what-if" premises in mind that were stated earlier, let us return to the question in the title. Are we testing for intelligence or for language? Suppose that some universal Basic language really were the foundation of intelligence. Further, suppose that the interaction between this system and the progressive acquisition of one or more particular language system were crucial to the accessibility and maturation of the deeper Basic system. From such a vantage point, it might well be that "intelligence" tests of all sorts are actually measures of the ability to equilibrate complex texts in relation to given facts. Human intelligence may achieve its marvelous feats of understanding by translation back and forth between the texts of a particular language and representations in some deeper Basic language system (perhaps rendered as Gestalts at some level). Within this framework, adaptation, assimilation, and accommodation could all be defined as processes of fitting potential propositional meanings to the facts of experience (which themselves are already expressed in a kind of language system).

In the final chapter of *The Human Brain*, Wittrock (1977) contends that "imagery" is crucial to the acquisition of "verbally presented" content. He cites the work of Allen Paivio and others showing that creating a meaningful image seems to have a direct and highly positive impact on retention and recall in many types of learning tasks. Could this be because of the fact that the creation of the appropriate sorts of images engages the main text generating machinery of human intelligence? According to the framework developed here, meaningfulness may depend on the linking of propositional meanings in one form with similar meanings expressed in different forms. It suggests that perhaps consciousness is dependent on translation or movement from one language system to another.

Paivio and others have demonstrated that the recall of a verbally presented form may be enhanced by causing the subject to link it up with a visual image that somehow expresses the same (or a similar) meaning. Could it be that the linking itself is the principal factor? Moreover, is it possible that this linking facilitates retention and recall because it requires an active fitting of textual meanings to each other?

In two dramatic demonstrations, Keislar and McNeil (1962) and Wittrock (1963) showed that children who dealt with iconic presentations of various facts of molecular kinetics succeeded surprisingly well in recalling the principles a year later. In Wittrock's study, two-thirds of the children were able to remember the criterial concepts a year later. It was surprising that they were able to do this since they had only received two to four weeks of instruction. More importantly, they were believed to be too young to learn the concepts in question in the first place.

Could the translation of texts in one language (Basic) to texts in another language (e.g., English or whatever) be the essence of intelligent action? Is experience textual in nature? Can it be understood without inferential operations that relate incoming facts to precedents and consequences? What if intelligence were language-based after all?

References

Alderson, J. Charles and Hughes, Arthur. *Issues in Language Testing*. London: British Council, 1981.

Bachman, Lyle F. and Palmer, Adrian S. "The Construct Validation of the FSI Oral Interview." In J. W. Oller, Jr., ed., *Issues in Language Testing Research*. Rowley, Massachusetts: Newbury House, in press. Reprinted from *Language Learning* 31, 1981, 67-86.

Bogen, Joseph E. "Some Educational Implications of Hemispheric Specialization." In M.C. Wittrock, Jackson Beatty, Joseph E. Bogen, Michael S. Gazzaniga, Harry J. Jerison, Stephen D. Krashen, Robert Nebes, and Timothy J. Teyler. *The Human Brain*. Englewood Cliffs, New Jersey: Prentice Hall, 1977. 133-152.

Binet, A. *The Psychic Life of Micro-Organisms*. Chicago: Open Court, 1888.

Canale, M. "On Some Dimensions of Language Proficiency." In J. W. Oller, Jr., ed., *Issues in Language Testing Research*. Rowley, Massachusetts: Newbury House, in press.

Carroll, John B. "Psychometric Theory and Language Testing." In J. W. Oller, Jr., ed., *Issues in Language Testing Research*. Rowley, Massachusetts: Newbury House, in press.

Cattell, R. B. *An Introduction to IPAT (Cattell's) Culture Fair Intelligence Testing*. Champaign, Illinois: Institute for Personality and Ability Testing, n.d.

Clark, Brian F.C. *The Genetic Code*. London: E. Arnold, 1977.

30

Critchley, M. "Speech and Speech-Loss in Relation to Duality of the Brain." In V. B. Mountcastle, ed., *Interhemispheric Relations and Cerebral Dominance*. Baltimore: Johns Hopkins, 1962. 203-213.

Cummins, J. "Is Academic Achievement Distinguishable from Language Proficiency." In J. W. Oller, Jr., ed., *Issues in Language Testing Research*. Rowley, Massachusetts: Newbury House, in press.

De Beaugrande, Robert. *Text, Discourse, and Process*. Norwood, New Jersey: Ablex, 1981.

Dewey, John. *Essays in Experimental Logic*. New York: Dover, 1916.

Fodor, J.A. *The Language of Thought*. Cambridge, Massachusetts: Harvard, 1975.

Fodor, J. A. "Fixation of Belief and Concept Acquisition." In M. Piatelli-Palmarini, ed., *The Debate Between Jean Piaget and Noam Chomsky*. Cambridge, Massachusetts: Harvard, 1980. 142-149.

Fodor, J. A.; Bever, T.G.; and Garrett, M. F. *The Psychology of Language: An Introduction to Psycholinguistics and Generative Grammar*. New York: McGraw Hill, 1974.

Gunnarsson, Bjarni. "A Look at the Content Similarities between Intelligence, Achievement, Personality, and Language Tests." In J. W. Oller, Jr. and K. Perkins, ed., *Language in Education: Testing the Tests*. Rowley, Massachusetts: Newbury House, 1978. 17-35.

Herrnstein, R. *IQ in the Meritocracy*. Boston: Little Brown, 1973.

Jensen, Arthur R. "How Much Can We Boost IQ and Scholastic Achievement?" *Harvard Educational Review* 39, 1969, 1-123.

Jensen, Arthur R. *Bias in Mental Testing*. New York: Free Press, 1980.

Jerison, Harry J. "Evolution of the Brain." In M. C. Wittrock, Jackson Beatty, Joseph E. Bogen, Michael S. Gazzaniga, Harry J. Jerison, Stephen D. Krashen, Robert Nebes, and Timothy J. Teyler. *The Human Brain*. Englewood Cliffs, New Jersey: Prentice Hall, 1977.

Kieslar, E. and McNeil, J. "Teaching Science and Mathematics by Auto-instruction in the Primary Grades: An Experimental Strategy in Curriculum Development." In J. F. Coulson, ed., *Programmed Learning and Computer-Based Instruction*. New York: Wiley, 1962.

Lashley, Karl. "The Problem of Serial Order in Behavior." In L.A. Jeffress, ed., *Cerebral Mechanisms in Behavior*. New York: Wiley, 1951. Reprinted in S. Saporta, ed., *Psycholinguistics: A Book of Readings*. New York: Holt, 1961. 180-197.

Lassen, Niels A.; Ingvar, David H.; and Skinhoj, Erik. "Brain Function and Blood Flow." *Scientific American*, 239, 1978, 62-71.

Lenneberg, Eric. *Biological Foundations of Language*. New York: Wiley, 1967.

Oller, J. W., Jr. "Language as Intelligence?" *Language Learning*, 31, 1981, 465-492.

Oller, J. W., Jr., ed., *Issues in Language Testing Research*. Rowley, Massachusetts: Newbury House, in press.

Oller, J. W., Jr. and Perkins, K., eds., *Language in Education: Testing the Tests*. Rowley, Massachusetts: Newbury House, 1978.

Oller, J. W., Jr. and Streiff, Virginia. *The Language Factor: More Tests of Tests*. in press.

Piaget, Jean. *The Psychology of Intelligence*. Totowa, New Jersey: Littlefield Adams, 1947.

Piatelli-Palmarini, Massimo, ed., *Language and Learning: The Debate between Jean Piaget and Noam Chomsky*. Cambridge, Massachusetts: Harvard, 1980.

Pribram, Karl. *Languages of the Brain*. Englewood Cliffs, New Jersey: Prentice-Hall, 1971.

Raven, J. C. and Lewis, H. K. *Raven's Progressive Matrices*. New York: Psychological Corporation, 1938-1965.

Roth, D. "Raven's Progressive Matrices as Cultural Artifacts." In W. S. Hall and M. Cole, eds., *Quarterly Newsletter of the Laboratory of Comparative Human Psychology*, 1, 1978, 1-15.

Rumelhart, D. E. "Notes on a Schema for Stories." In D. G. Bobrow and A. Collins, eds., *Representation and Understanding: Studies in Cognitive Science*. New York: Academic, 1975. 211-236.

Schank, Roger and Abelson, R. B. *Scripts, Plans, Goals and Understanding*. New York: Academic, 1977.

Seidner, Stanley S., *Issues of Language Assessment: Foundations and Research*. Springfield: Illinois State Board of Education, 1982.

Wittrock, M. C. "Response Mode in the Programming of Kinetic Molecular Theory Concepts." *Journal of Educational Psychology*, 54, 1963, 89-93.

Wittrock, M. C. "The Generative Process of Memory." In M. C. Wittrock, Jackson Beatty, Joseph E. Bogen, Michael S. Gazzaniga, Harry J. Jerison, Stephen D. Krashen, Robert Nebes, and Timothy J. Teyler. *The Human Brain*. Englewood Cliffs, New Jersey: Prentice Hall, 1977. 153-184.

Woese, Carl R. *The Genetic Code: The Molecular Basis for Genetic Expression*. New York: Harper and Row, 1967.

Wolpert, Lewis. "Pattern Formation in Biological Development." *Scientific American* 239, 1978, 154-164.

Ycas, Martynas. *The Biological Code*. Amsterdam: North Holland, 1969.

Young, J. Z. *Programs of the Brain*. Oxford: Oxford University, 1978.

ADDITIVE VERSUS SUBTRACTIVE FORMS OF BILINGUALISM: CONFUSION REGARDING PROGRAMS OF IMMERSION*

Wallace E. Lambert
McGill University

Introductory Overview

Immersion programs for teaching second languages and developing bilingual skills were initiated in public schools in Canada some fifteen years ago. Since then their development, modification and implementation have been described in technical journals along with an accumulation of evaluative studies of their impact on pupils who proceed from one year in the program to another. Information about the programs and their surprisingly favorable outcomes have caught the attention of educators in various nations, especially in the USA. In order to assess their possible applicability to the USA educational scene, it is important that a clear description of a typical "IMMERSION" program be widely available, along with differentiations between "immersion," "submersion," and "second-language-teaching" approaches. This is one aim of the present paper.

The second is to highlight the fact that immersion programs were devised basically for native English-speaking pupils in Canada or in the U.S.A. who are certain that their development of skills in the English language would never be left in jeopardy, either by the educators in charge of the immersion programs or by the home, community and media environments in which they lived. With the assurance of a solid English-language linguistic environment, the immersion approach exposes pupils to as much of the second language as possible (short of living in a foreign setting) while monitoring its effects and the effects of having little or no home-language instruction on children in the early elementary years of schooling. The programs were explicitly not meant for Canadian children with some language other than English as the home language (in particular not those who have French as the home language) who, one might expect, could profit as well from a "reversed immersion" program where they would encounter only or mainly English language instruction. This reversal was painstakingly avoided because of a well-grounded fear on the developers' part that in North America (Canada and Quebec included), French or any other home language other than English is vulnerable to neglect and replacement. Thus, a French-speaking Canadian in a reversed (i.e., English-based) immersion program might move towards bilingualism for a short time, but basically he/she would be starting a slow "subtraction" of French and its replacement by English which, in the North American setting, can too easily be viewed as a more useful, prestigious or otherwise more "valuable" language. The cognitive and educational difficulties associated with this subtraction/replacement process need to be understood.

Accordingly, attention will be directed to a distinction that has been made, on the basis of available research studies, between "additive" and "subtractive" forms of bilingualism and the confusions that would arise if immersion programs, which were designed to be "additive," were reversed and inadvertently made "subtractive." Potential solutions to the problem of linguistic and cultural subtraction, in the case of ethnolinguistic minorities, are available.

The Origins, the Nature, and the Outcomes of Immersion Programs

Origins. It is important to understand the origins of immersion programs in the Canadian setting because they are much more social-psychological than they are linguistic or educational at their cores. Basically, immersion education started because of the desire on the part of the original group of parents, educators and researchers, who got the first experimental programs underway, to make Canada a more fair and more interesting society for both "founding peoples," French-speaking Canadians and English-speaking Canadians. Although Canadian in content, this development is pertinent to American society because similar social processes run their courses in both settings. In Canada these are more visible because of socio-political movements towards independence or separation on the part of the French-Canadians.

Although there are many Canadian/American parallels, there are still important differences. For instance, Canada's constitution has clear provisions for the protection of the language and culture of both French- and English-speaking subgroups, and although the government has a policy favoring multiculturalism, it does not provide extended support for education being conducted in any of the numerous other home languages spoken in Canada. Since World War II, non-English speaking immigrants make up a sizeable proportion of Canada's population. To its great credit, the USA has federal laws requiring educational help — involving teaching via the home language of pupils — for all non-English speaking ethnic groups who, it is recognized, are placed at a disadvantage in schools and in occupations that presume native competence in English. However, the USA shows no signs of recognizing or appreciating the *de facto* bilingual character of contemporary America which has nearly as many families with Spanish as the home language as there are people in the total population of Canada. And the English-Spanish bilingual character of contemporary America is only one strain, for there are various other equally vital ethnolinguistic groups; each

contributing to a fascinating multicultural American society. There is, then, much more to be done to capitalize on this ethnolinguistic richness in both America and Canada.

French-speaking Canadians have had a long history of finding themselves second-class citizens in a social world which has reinforced Anglo-American values and the English language. The second-class status showed itself in the form of French-speaking Canadians playing subordinate roles to English Canadians, the dominant subgroup in Canadian society, comparable to English-speaking whites in the USA. Not only have French Canadians been grossly underrepresented in the upper levels of Canada-wide status hierarchies, but even in the Province of Quebec, where they constitute some 80 percent of the population, French Canadians have not, relative to English Canadians, made it occupationally or economically, and their style of life has been ignored, ridiculed and blamed as the cause of their social and economic position. The trouble is that this type of thinking becomes contagious and over time even members of the marked minority group begin to believe they are inferior in some sense and blame themselves for their inferiority (see Lambert, 1967). It takes much reflection in frustrating situations of this sort to see through the sophistry and realize that one's ethnic or social-class group is in no way inherently inferior, but simply that those with the power advantages have learned well how to keep the advantages and that their social-class cushion makes keeping power relatively easy for them. Stereotyping or otherwise marking minority groups — people they really know very little about — becomes an effective way for the majority group to keep others out of the power sphere.

As social psychologists, we began to study this state of affairs in Canada some 25 years ago just as two extreme solutions to the "French Canadian problem" were coming into vogue: 1) French Canadians should pull up their socks and compete — meaning they should master English and Anglo-American ways — while toning down their French Canadian-ness; 2) French Canadians should pull apart or separate — meaning they should form a new independent nation where they could be masters of their own fate and where the French Canadian language and culture could be protected. Both alternatives worried us because one meant giving up a style of life that was precious, and the other meant closing a society through separation, "closing" in the sense that Karl Popper (1966) uses the term in describing socio-political attempts to create a conflict-free subworld where the "good old ways" will be protected. Instead we viewed the French Canadian way of life as something valuable for Canada as a whole — a nation whose potential and fascination rest in its multicultural/multilingual makeup — whether or not it was appreciated as such by the majority of English or French Canadians.

The Nature of Immersion Schooling. So we became interested in reducing, if possible, the ignorance of French Canadian-ness and in enhancing an appreciation for it among Anglo-Canadian children. This then became the guiding purpose for the research initiated at McGill on "early immersion" schooling (see Lambert & Tucker, 1972; Swain, 1974; Genesee, 1978-79), wherein English-speaking children, with no French-language experience in their homes and little if any in their communities, enter public school kindergarten or grade 1 classes that are conducted by a monolingual French-speaking teacher. This "early immersion" or "home-to-school language-switch" program as we call it, is kept exclusively French through grade 2 and only at grade 2 or 3 in English introduced, in the form of a language arts program, for one period a day. By grade 4, particular subject matters are taught in English (by a separate English-speaking teacher) so that by grade 5 and 6, some 60% of instruction is conducted in English (see Lambert, 1979).

The concept of immersion schooling was based on a very important and fundamental premise — that people learn a second or third language in the same way they learn their first and that languages are best learned in contexts where the person is socially stimulated to learn the language and is exposed to it in its natural form.

From the first encounter, the immersion teacher used only the target language. She clearly, patiently and repetitively focuses on the development of a basic vocabulary in the new language, relying, with young children, on plastic art materials, songs and animated stories. But from the start, the learning of language *per se* is made quite incidental to learning how to do new and interesting things, with the new language as the verbal accompaniment. Later, new ideas of a scientific, mathematical or problem-solving nature are given the main focus, and again the amassment of skill in the new language appears to be incidental, except for short daily periods of language arts which focus on the new language itself (see Lambert & Tucker, 1972). The teachers' aim is to cover fully the content matters expected of any child in a conventional program at that grade level.

Immersion classes normally comprise only Anglophone pupils; thus, the whole class experiences immersion as a group. In some cases, a few children who are native speakers of the target language are introduced into the otherwise Anglophone class, and their presence can be useful in many ways. However, one has to choose carefully those who are to represent the other ethnic group; a biased selection could easily upset the development of positive attitudes toward the other group (see Lambert, 1982).

We refer to situations where a very few Anglophone children are placed in an otherwise all French language class (one comprised of Francophone pupils only), as "submersion" instead of immersion. The chances for more rapid development of expressive language skills in the foreign language are obvious in submersion in comparison to immersion programs, but being few and being different might present social adjustment problems not encountered in immersion classes.

Immersion and submersion programs both focus on subject-matter mastery and make language learning incidental, and this similarity makes both quite different from second-language teaching programs (e.g., French-as-a-second-language) where the new language is the focus, where content-matter mastery is not a main goal and where only small amounts of time are devoted to the second-language component. That component is also usually covered by a specialist, rather than the grade teacher. Thus, immersion programs are much more intense and comprehensive than second-language programs, and since no specialists are involved, the costs of immersion programs are hardly any different from normal costs since the class teacher is also the language specialist and the class size (e.g., 30-32 pupils to a teacher in Canada) is usually kept normal.

Immersion differs from typical "bilingual education" programs (as these are conducted in North America) because no bilingual skills are required of the teacher (who presents herself as a monolingual in the target language) and who, therefore, never switches languages, reviews materials in the other language, or otherwise encourages bilingualism in the teacher-pupil interactions. In immersion programs, bilingualism is therefore developed through monolingual instructional routes.

Outcomes: Educational and Cognitive. The consistent findings from fifteen years of careful research on children in immersion programs permit several conclusions which bear not only on the linguistic consequences of these programs, but the psychological and social consequences as well. 1) Immersion pupils are taken along by *monolingual* teachers to a level of functional bilingualism that could not be duplicated in any other fashion short of living and being schooled in a foreign setting. Furthermore, pupils arrive at that level of competence 2) without detriment to home-language skill development; 3) without falling behind in the all-important content areas of the curriculum, indicating that the incidental acquisition of French does not distract the students from learning new and complex ideas; 4) without any form of mental confusion or loss of normal cognitive growth; and 5) without a loss of identity or appreciation for their own ethnicity. Most important of all in the present context, 6) they also develop a deeper appreciation for French Canadians and a more balanced outlook towards them by having learned about them and their culture through their teachers and through their developing skill with the language of French Canadians.

Instructive illustrations of the effect of immersion schooling on students' cognitive development are now available, even though, because their sample sizes are small, they should be viewed as suggestive, rather than definitive at this point. One is a study by Scott (1973) in Montreal who compared two subgroupings of children who were in all respects a single homogeneous group until grade 1, at which time one subgroup entered an immersion program, while the second subgroup of closely comparable youngsters had not been given the immersion opportunity. Scott worked with data collected over a seven-year period from two groups of English-Canadian children, one which had become functionally bilingual in French during the time period through "immersion schooling" in French, while the second group had followed a conventional English-language education program. Scott focused on the possible effects that becoming bilingual might have on "divergent thinking," a special type of cognitive flexibility (see Guilford, 1950, 1956). Measures of divergent thinking provide subjects (in this case, children) with a starting point for thought — "think of a paper clip" — and then ask them to generate a whole series of permissible solutions — "tell me all the things one could do with it." Some researchers have considered divergent thinking as an index of creativity (e.g., Getzels and Jackson, 1962), or at least an index of a rich imagination and an ability to scan rapidly a host of possible solutions. Whatever the term should be, the results, based on a multivariate analysis, showed that the functionally bilingual immersion pupils, as of grades 5 and 6, were substantially higher scorers than the non-immersion controls with whom they had been equated for IQ and social class background at the first-grade level. Although the numbers of children in each group are small, this study supports a causal link between bilingualism and flexibility, with bilingualism one of the factors that seem to enhance flexibility.

Supportive evidence is also found in quite independent studies of the cognitive effects of immersion programs on Anglophone Canadian children in other areas of Canada, namely the work of Barik and Swain (1976) in Ottawa and Toronto and Cummins (1975, 1976) in Edmonton, Alberta. Both of these studies show increases in IQ or in divergent thinking scores that can be attributed to the development of bilingual skills transpiring in immersion schooling.

There is, then, an impressive array of evidence accumulating that argues plainly against the common sense notion that becoming bilingual — having two linguistic systems within one's brain — naturally divides a person's cognitive resources and reduces his/her efficiency of thought and/or language. Instead, one can now put forth a very strong argument that there are definite cognitive, educational and social advantages to being bilingual. These advantages are experienced as much by children from working class socioeconomic backgrounds as the more advantaged and for children with various levels of measured IQ, including children with diagnosed learning difficulties.

Outcomes: Sociocultural Awareness. What is exciting about this program, over and above its educational and cognitive impact, is that it opens children's minds to an otherwise foreign and possibly threatening outgroup. It also provides certain sociopolitical insights that monolingual mainstreamers would likely never have. For example, the immersion children come to the realization that peaceful democratic coexistence among members of distinctive ethnolinguistic groups calls for something more than simply learning one another's languages (see Blake, Lambert, Sidoti, & Wolfe, 1981;

Cziko, Lambert, Sidoti, & Tucker, 1979). Having learned the other language well and having learned to appreciate the other cultural group, children with immersion experience, compared to control, realize that effective and peaceful coexistence calls as well for opportunities for both ethnic groups of young people to interact socially on an equitable basis. This is a very sophisticated insight.

Immersion Programs in the USA

Thus, a new approach to bilingual education is now available, and since it works as well in other parts of Canada where few if any French Canadians are encountered in social life (see Swain, 1974), it or some variation of it can be expected to work equally well in the USA. In fact, there are currently some ten or more communities in the USA where comparable early immersion programs for mainstream English-speaking children are underway (in Spanish, French and German, so far), and from all available accounts, they are working splendidly (see Cohen, 1976; Samuels & Griffore, 1979; Montgomery County Public Schools, 1976; Derrick, 1980; Grittner, 1981; O'Connell, 1981; Sidoti, 1981). Part of the reason for their success is that school administrators and principals, after an initial period of skepticism and wariness, become extremely pleased and proud of the outcomes. Furthermore, the costs of the programs are surprisingly low compared to second-language-teaching programs because the regular teachers' salaries go to the new "foreign speaking" teachers.

But what really counts as success is the pride and progress reflected by teachers, parents and pupils. For example, Frank Grittner, the Supervisor of Second Language Education for the State of Wisconsin, has collected data on third grade English-speaking children (few with German ethnic backgrounds) in a German immersion program where they were taught through German for three years. That particular immersion program was related to a plan for desegregation, and thus, some 40% of the pupils involved were black. At the end of grade 3, 100 percent of the German immersion pupils scored in the average to above average range on the Metropolitan Achievement Test for Reading *(in English)* compared to 70 percent for Milwaukee schools in general and 77 percent for U.S. norm groups. Likewise for Mathematics Test scores (also tested through English), the respective averages were 92, 71 and 77 percent. Similar outcomes are available for English-speaking American children in a French immersion program in Holliston, Massachusetts, as of the end of grade 2 (O'Connell, 1981). In New York City where an interesting program of partial immersion in Spanish for English-speaking pupils has been tried out, the end-of-year parental responses and evaluations are extremely favorable. What characterizes the parents' reaction is the delight they show that their children are learning about Spanish-speaking people and developing an appreciation for them at the same time as they acquire the basics of the language. They are pleased not for "instrumental" reasons but for "integrative" ones, i.e., intergroup harmony is initiated, not that their children can profit in the business world by knowing Spanish (Sidoti, 1981).

Furthermore, there is strong evidence to show that monolingual Anglo-Canadian children can handle easily a "double immersion" program wherein two totally unknown languages (in this case French and Hebrew) are used as the main languages of instruction from kindergarten through the elementary grades (see Genesee & Lambert, 1982). Incidentally, the striking success of these double-immersion programs in Montreal schools make one think twice about Canadian policy makers who give verbal support to promote multiculturalism, but stop short of providing at least some instruction via home languages. The point is that ethnic minorities in Canada might easily handle and enjoy education that is trilingual — French, English and home languages — just as the Jewish children in the double immersion programs not only manage, but also enjoy education that is French, Hebrew and English.

The variants of the immersion program that might be valuable and relevant when applied in the USA are limited only by one's imagination. For instance, the New York City variant is a partial immersion program that can be increased in time devoted and in scope to satisfy large numbers of pupils with a variety of language options (see Sidoti, 1981). Then there is an extremely interesting "Language to Share" program, wherein ethnic minority adolescents are trained to be junior teachers of their home language to pupils two or three years younger than themselves. Similarly, there are possibilities for "language exchange" programs (Lambert, 1978b), wherein speakers of English, for example, who are interested in learning particular foreign languages are paired up with peers who have those foreign languages as their home languages and who exchange two or so hours per week in teaching their home language informally while receiving English instruction in return. The exchanges are coordinated by a small group of master teachers.

What this all means is that there is now available an effective means of developing a functionally bilingual citizenry. The evidence, both scientific and anecdotal, is consistent enough that many of us are confident that such programs will slowly but surely be implemented in the USA and other major sites. Our concerns now shift to another domain, i.e., that immersion programs will be instigated without sufficient attention given to certain social-psychological processes that come into play simultaneously with language learning. Such processes are at the base of what we refer to as "additive versus subtractive" forms of bilingualism.

Additive Versus Subtractive Forms of Bilingualism

In the descriptions just given, it is important to keep in mind which segments of the Canadian and American societies have played the major role in immersion programs. In each setting, it has been the English-speaking Canadian and the English-speaking American mainstream families — those segments of the respective societies that are most secure in their

own ethnic and linguistic identity, but the ones most in need of knowledge about and appreciation for other ethnic and linguistic groups. To the extent that mainstream children are sensitized to and educated in other languages and cultures, the better the chances are of developing a rich, harmonious, pluralistic society. The better too are the chances of improving the self-views of ethnolinguistic minority children who are immensely heartened and complimented when they realize that mainstream children are making sincere gestures to learn about them, their language, and their ways of life.

We have referred to this process of developing the bilingual and bicultural skills of English-speaking Canadian or American children as an "additive" form of bilingualism, implying that these children, with no fear of ethnic/linguistic erosion, can add one or more foreign languages to their accumulating skills and profit immensely from the experiences — cognitively, socially, educationally, and even economically (see Lambert & Tucker, 1972; Lambert, 1978a). Developing strong skill in a second, socially relevant language would enlarge these pupils' repertory of skills, and rather than detracting from their English home-language base, their skill with English appears to be enhanced. For these children and their parents, it becomes clear that the learning of the second language in no way portends the slow replacement of the first or "home" language, as would be the case for most linguistic minority groups in North America who are pressured to develop high-level skills in English at the expense of their home languages. Thus, we refer to these experiences as examples of "additive" bilingualism, and we draw a sharp contrast with the "subtractive" form of bilingualism experienced by ethnolinguistic minority groups, who, because of national educational policies and/or social pressures of various sorts, feel forced to put aside or subtract out their ethnic languages for a more necessary, useful and prestigious national language (Lambert, 1974).

In the subtractive case, one's degree of bilinguality at any point in time would likely reflect a stage in the gradual disuse of the ethnic home language and a fading of that language's associated cultural accompaniments and their replacement with another more "necessary" language and a new cultural accompaniment. This form of bilingualism can be devastating because it gradually makes the child's basic conceptual thinking — to the extent that it is linked with or colored by the home language — no longer relevant or functional. In a sense, the child is expected to redevelop basic concepts with a new linguistic accompaniment, while trying to keep up with native English-speaking peers. In other words, these youngsters are placed in a psycholinguistic limbo where neither language is useful as a tool of thought and expression, a type of "semi-lingualism," as Skutnabb-Kangas & Toukomaa (1976) put it. Although not yet fully understood, this subtractive phenomenon could become a central issue in the psychology of bilingualism.

The case of French and English in Montreal is interesting because both additive and subtractive features are involved. For Anglophone Quebecers, learning French is clearly additive in nature, with no fear of a loss of identity or of French eradicating English-language competence. Since Francophone Quebecers comprise some 80 percent of the population and have their own French-language school system from kindergarten to the most advanced professional institutions, learning English might also be thought of as additive. From a North American perspective, however, Quebec is a small French-speaking enclave that is continuously bombarded by English language media, with pressures on its children to prepare themselves for life in an otherwise English-speaking semi-continent. For Francophone Canadians outside Quebec, the chances of keeping French alive as a home, school and work language are slim. This fear of a subtractive loss of Frenchness is real for many French-speaking Quebecers as well; a too ardent move toward Englishness might well subtract out Frenchness.

The research of Taylor, Meynard, & Rheault (1977) indicates how sensitive certain French-speaking Canadian subgroups are to a possible loss of ethnic identity when they are either forced or enticed to use English instead of French as a language of work and/or a language of thought. Thus it becomes clear how learning the other group's language through immersion schooling means one thing for the ethnically and linguistically comfortable Anglophone Canadian, but a quite different thing for the ethnically and linguistically insecure Francophone Canadian. In one case, immersion-in-French offers numerous advantages to Anglophones, through an additive process, while in the other case immersion-in-English would be a menacing reversal of immersion schooling, which, through a subtractive process, could place a social minority group's language and cultural identity in jeopardy.

It was these examples of different expectations and worries to be found in Francophone versus Anglophone communities in Canada that led Lambert and Tucker (1972) to propose a "general guiding principle: In any community where there is a widespread desire or need for a bilingual or multicultural citizenry, priority in early years of schooling should be given to the language or languages least likely to be developed otherwise, in other words, the language(s) most likely to be neglected" (1972, p. 216). This would mean having *both* French- and English-speaking Canadian children start their schooling in French in Canada, the language more likely to be neglected or bypassed; this would provide an additive immersion program for Anglophone children while protecting Francophone children from the subtractive drift to English. Once their home language is established as an active language of thought and expression, the Francophone children could gradually start a part-time immersion-in-English program to assure them an eventual bilinguality, comparable to that of the young Anglophones learning French through immersion. Lambert and Tucker believe that this principle holds for any setting; it is a matter of

determining which language(s) have high status and utility and which low in the given community, and then substituting new languages for English and French in the Canadian example.

Tranforming Subtractive Bilingualism into Additive Bilingualism

The guiding principle just mentioned can be applied in another manner. One starts with the reasonable proposition that the major aim of education in North America should be to brighten the outlook for ethnolinguistic minority-group children by preparing them to compete fairly with mainstream children in educational and occupational pursuits. As potential bilinguals, they certainly have the cognitive and linguistic potential as the research already mentioned shows. The best way I can see to release that potential is by transforming their subtractive experiences with bilingualism and biculturalism into additive ones.

We already have a few research-based examples of how this transformation might work. The first is the case of Franco-Americans in northern New England who recently were given a chance to be schooled partly in their home language (Dube and Herbert, 1975a and 1975b; Lambert, Giles and Picard, 1975; Lambert, Giles and Albert, 1976). Some 85% of families in the northern regions of Maine have kept French alive as the home language or as one of the two home languages, even though traditionally all schooling has been conducted in English. We participated in an experiment wherein a random selection of schools and of classes in the area were permitted to offer about a third of the elementary curriculum in French and a second sample of schools, with children of comparable intelligence scores and socioeconomic backgrounds, served as a control or comparison in that all their instruction was in English. After a five-year run, the children in the "partial French" classes clearly outperformed those in the control classes in various aspects of *English* language skills and in academic content, such as math, learned partly via French. At the same time, French had become for them something more than an audio-lingual language because of the reading and writing requirements of the French schooling. These results mean that the French-trained Franco-American children were given a chance to be fully bilingual and this had repercussions on their cognitive abilities and bettered their opportunities to compete in occupations or professions that call for high-level educational training. They had been lifted from the typical low standing on scholastic achievement measures that characterizes so many ethnolinguistic groups in North America.

An important element in this transformation was the change in the self-views of the French-trained youngsters who, we found, began to reflect a deep pride in being French and a realization that their language was as important a medium for education as English (Lambert, Giles and Picard, 1975). Similar community-based studies are underway in the American Southwest, and these, too, are based on the belief that ethnolinguistic minorities need a strong educational experience in their own languages and traditions before they can cope in an "all American" society or before they will want to cope in such a society.

A second example of a transformation of subtractive to additive bilingualism is provided by Carolyn Kessler and Mary Quinn (1980). In their study, Spanish-speaking Texas grade 6 students were given the opportunity in elementary school to learn subject matters via Spanish while learning English, that is, like the Franco-Americans in the first example, to use their home language — the language through which their basic conceptual thinking developed in infancy — as one of the linguistic media for further conceptual growth. The Hispanic-American students were compared with a much more privileged sample of middle class, white monolingual English-speaking American pupils of the same age. Both groups were given an extensive training program in "science inquiry" through films and discussion of physical science problems and hypothesis testing. In tests given after the training, it was found that the Spanish-English bilinguals generated hypotheses of a much higher quality and complexity than did the monolinguals. This problem-solving quality was also reflected in the language used, as indexed by a "syntactic complexity" measure, so that the bilinguals were clearly using more complex linguistic structures as well. They also found substantial correlations between their measures of hypothesis quality and syntactic complexity, providing thereby an important link between problem-solving capacity and linguistic skills.

The research by Kessler and Quinn jibes nicely with other findings. For example, Padilla and Long (1969) found that Spanish-American children and adolescents can acquire English better and adjust more effectively to the educational and occupational demands of American society if their linguistic and cultural ties with the Spanish-speaking world are kept alive and active from infancy on. There are in fact numerous recent examples (Hanson, 1979; McConnell, 1980; Rosler & Holm, 1980; Troike, 1978) that point in the same direction. G.R. Tucker (1980) recently summarized these studies and concluded that there is "a cumulative and positive impact of bilingual education on all youngsters when they are allowed to remain in bilingual programs for a period of time greater than two or three or even five years and when there is an active attempt to provide nurturance and sustenance of their mother tongue in addition to introducing teaching via the language of wider communication" (pp. 5-6).

Conclusions

A new form of education is developing around immersion programs, already well known in Anglophone communities across Canada and, now being scrutinized carefully for possible adoption by educators, parents and researchers in the USA. The innovative feature is that we now realize that solid education and better-than-normal cognitive development can take place in classes where a second

totally foreign language is used as the major or only medium of instruction and learning from the earliest school grades on, only gradually giving way to conventional English language instruction from grade 3 or so on, but nonetheless kept as an active alternative language of instruction and learning through high school.

These conclusions hold, without exception as far as we can determine, only for English-speaking children in Canada and the USA, the segments of American society for whom immersion-in-another-language programs were developed. It is now clear that Anglophone youngsters can keep up with or surpass control children in conventional classes in their performance on curriculum content, and there are no negative effects on English language development, cognitive development or general educational attainment. Instead, their English skills and cognitive growth are strengthened, relatively and their ethnic identity is likewise better rooted and made open to other ethnicities.

All of these advantages are seen as a form of enrichment derived from the "addition" of highly advanced skills with a new language and a deep understanding of a different ethnolinguistic group. The enrichment is additive because these children have no fear or worry in North America of losing English as their basic language of thinking, expressing and problem solving. Nor need they have worries that they will lose their American identity in the process.

In contrast, immersion programs were not designed or meant for ethnolinguistic minority groups in North America who have some language other than English as the main home language. To place these children in an all-English instructional program would be to reverse the immersion in a harmful, subtractive way. Their personal identities, their early conceptual development, their chances of competing or succeeding and their interest in trying to succeed would all be hampered by a reversed immersion-in-English program.

Fortunately, there are practical and valuable alternatives now available to help these children transform a potentially subtractive form of bilingual development to an additive one, with advantages similar to those enjoyed by the mainstream Anglophone in normal immersion programs. These transformations are not based on typical "bilingual education" models, but rather on a dual-track education model that emphasizes the use of the non-English home language as the major instructional language in the early grades and, when the time is right, introduces a separate English-language instructional component when it is certain that the child's home language has taken roots and is a secure base for adding the new language. This stage may not be reached until grade two or three. This alternative brings two traditionally disparate social groupings — Anglophone mainstreamers and ethnolinguistic minorities — together around an exciting new form of bilingualism and biculturalism.

For some, especially the older generations, such an emphasis on bilingual or multilingual/multicultural development may seem un-American or socially dysfunctional and impractical. For others, especially the young in spirit, it may represent the emergence of a new coping style that is taking shape in contemporary North America where sizeable subgroups of young people are not only inquisitive about the rich multiethnic societies they find themselves in, whether in Canada or the USA, but also anxious to become actively involved in this ethnic richness.

References

Barik, H.C., & Swain, M. "A Longitudinal Study of Bilingual and Cognitive Development." *International Journal of Psychology*. 1976, *11*, 251-263.

Blake, L., Lambert, W.E., Sidoti, N., & Wolfe, D. "Students' Views of Intergroup Tensions in Quebec: The Effects of Language Immersion Experience." *Canadian Journal of Behavioural Science*, 1981, *13*, 144-160.

Cohen, A. "The Case for Partial or Total Immersion Education." In A. Simoes, Jr. (Ed.), *The Bilingual Child*. N.Y.: Academic Press, 1976.

Cummins, J. *Cognitive Factors Associated with Intermediate Levels of Bilingual Skills*. Unpublished manuscript, Educational Research Centre, St. Patrick's College, Dublin, 1975.

Cummins, J. "The Influence of Bilingualism on Cognitive Growth: A Synthesis of Research Findings and Explanatory Hypotheses." *Working Papers on Bilingualism*, 1976, *9*, 1-43.

Cziko, G.A., Lambert, W.E., Sidoti, N., & Tucker, G.R. "Graduates of Early Immersion: Retrospective Views of Grade 11 Students and Their Parents." In R. St. Clair & H. Giles (Eds.), *The Social and Psychological Contexts of Language*. Hillsdale, N.Y.: Erlbaum Associates Inc., 1980.

Derrick, W.J. Personal communication, 1980.

Dubé, N.C. & Herbert, G. *St. John Valley Bilingual Education Project*. U.S. Department of Health, Education and Welfare, Washington, D.C., 1975(a).

Dubé, N.C. & Herbert, G. *Evaluation of the St. John Valley Title VII Bilingual Education Program, 1970-1975*. Mimeo, Madawaska, Maine, 1975(b).

Genesee, F. "Scholastic Effects of French Immersion: An Overview After Ten Years." *Interchange*, 1978-1979, *9*, 20-29.

Genesee, F. & Lambert, W.E. "Trilingual Education for Majority Language Children." *Child Development*, in press, 1982.

Getzels, J W & Jackson, P W *Creativity and Intelligence* New York: Wiley & Sons, 1962

Gritner, I Personal communication, 1981

Guilford, J P "Creativity" *American Psychologist* 1950, 5, 444-454

Guilford, J P "The Structure of Intellect" *Psychological Bulletin*, 1956, 53, 267-293

Hanson, G *The Position of the Second Generation of Finnish Immigrants in Sweden The Importance of Education in the Home Language* Symposium Report Spit, Sweden, 1979

Kessler, C & Quinn, M *Bilingualism and Science Problem-Solving Ability* Unpublished paper presented at the 14th Annual International Convention of Teachers of English to Speakers of Other Languages, San Francisco, CA, 1980

Lambert, W E "The Social Psychology of Bilingualism" *Journal of Social Issues*, 1967, 23, 91-109

Lambert, W E "Some Cognitive and Sociocultural Consequences of Being Bilingual" In J E Alatis (Ed), *International Dimensions of Bilingual Education* Washington, D C Georgetown University Press, 1978a

Lambert, W E "An Alternative to the Foreign Language Teaching Profession *Interchange*, 1978b, 9, 95-108

Lambert, W E *A Canadian Experiment in the Development of Bilingual Competence The Home-to-School Language Switch Program* Mimeo, McGill University, Psychology Department, 1979

Lambert, W E *The Effects of Bilingual-Bicultural Experiences on Children's Attitudes and Social Perceptions* Mimeo, Department of Psychology, McGill University, 1982

Lambert, W E, Giles, H, & Albert, A *Language Attitudes in a Rural City in Northern Maine*, Mimeo, McGill University, 1976

Lambert, W E, Giles, H, & Picard, O "Language Attitudes in a French-American Community" *International Journal of the Sociology of Language*, 1975, 4, 127-152.

Lambert, W E & Tucker, G R *Bilingual Education of Children The St Lambert Experiment* Rowley, Mass: Newbury House, 1972.

McConnell, B B *Effectiveness of Individualized Bilingual Instruction for Migrant Students* Ph D dissertation, Washington State University, 1980.

Montgomery County Public Schools, Maryland *End of the School Year Report on the French Language Immersion Program at Four Corners*, 1976

O'Connell, J Personal Communication, 1981

Padilla, A M & Long, K K *An Assessment of Successful Spanish American Students at the University of New Mexico* Paper presented to the annual meeting of the AAAS, Rocky Mountain Division, Colorado Springs, 1969

Popper, Karl *The Open Society and Its Enemies* 2 Vols London: Routledge and Paul Kegan, 1966

Rosier, P & Holm, W *Saad Naaki Bee Na'nitin Teaching by Means of Two Languages — Navajo and English — at Rock Point Community School* Center for Applied Linguistics, Washington, D C, 1980

Samuels, D D, & Griffore, R J "The Plattsburgh French Language Immersion Program: Its Influence on Intelligence and Self-Esteem" *Language Learning*, 1979, 29, 45-52

Scott, S *The Relation of Divergent Thinking to Bilingualism Cause or Effect?* Unpublished research report, McGill University, 1973

Sidoti, N Personal communication, 1981

Skutnabb-Kangas, R & Toukomaa, P *Teaching Migrant Children's Mother Tongue and Learning the Language of the Host Country in the Context of Sociocultural Situation of the Migrant Family* The Finnish National Commission for UNESCO Helsinki, 1976

Swain, M "French Immersion Programs across Canada" *The Canadian Modern Language Review*, 1974, 31, 117-128

Taylor, F M, Meynard, R & Rheault, E "Threat to Ethnic Identity and Second-Language Learning" In H Giles (Ed), *Language Ethnicity and Intergroup Relations* London: Academic Press, 1977.

Troike, R C "Research Evidence for the Effectiveness of Bilingual Education" *NABE Journal*, 1978, 3, 13-24.

Tucker, G R *Comments on Proposed Rules for Nondiscrimination under Programs Receiving Federal Financial Assistance through the Education Department* Washington, DC: Center for Applied Linguistics, 1980.

IMAGE PRODUCTION IN TWO LANGUAGES: ITS IMPLICATIONS IN TESTING THE BILINGUAL CHILD

Antonio Simões, Jr.
New York University

Introduction in the First Person

As my hand pushes this pen, and my hand puts movements into words, and the words are developed into concepts and images, my values are in most ways written down. These thoughts can be transformed into images and are legitimated by some sort of logical process. My universe of knowledge is applied by language, thought, and action because what I do is "truth" that is known to me. But I am not alone on this earth, and I must live with other human beings whose images are similar to my own. Many times we, as historical living creatures, must get together and agree on our value systems in an organized way. Simply, and accepting all the complexities involved, we can call this union of agreement the institutions of our society. These entities cannot exist without human beings, human beings with feelings, human beings with collective histories, human beings with individual biographies, human beings with certain traditions. With these common agreements of what is and what should be, we as part of the human endeavor must rationalize our actions by creating a norm. Our logic seems correct and when drastic change is implied, change that may destroy our universe of thought, we resist. If it is change that "fits" somewhere into our existing schema, if it is change that can be legitimated in our domain, we accept it, sometimes with reluctance, and then we develop a "new" norm or universe of knowledge, but we forget one thing: We are creatures tied to a history, a language system, a social milieu, a culture which we cannot deny. But what happens if one possesses two cultures and one knows that the two values may be, images associated with the same word? Let us explore some of these issues in what I call the "Image Production in Two Languages: Its Implications in Testing the Bilingual Child."

For a decade I have talked about the concept of image production and language use to my colleagues, to my students, and to my friends with no real satisfactory in-depth determination of what my field (curriculum theory-bilingual education) had to offer as it related to knowledge-learning, knowledge-knowing, and thought. I knew that eventually a more adequate conceptual schema defining the relationship of language use and image production would be developed, but until this happened, I continued to read in the social sciences for a satisfactory answer. It did not come.

As a curriculum theorist with an interest in phenomenology as it relates to language education, I have always been interested by how society distributes knowledge and how knowledge is valued or prescribed. Reading Ponte, Shutz, Heidegger, Camus, Sartre, Witgenstein, and others gave me a clue of what knowledge is and how it relates to how one functions in the world through language and thought, but I was never satisfied with the personal dilemma of "talking-about" image production in more than one language.

Linguistics helped me to "look-at" the structure of language; sociolinguistics attended to the interpretation of language, and the use of it in social situations; Berger and Luckmann (1967) in the *Social Construction of Reality* helped me to interpret the distribution of knowledge from the viewpoint of sociology; psychology attempted to describe and analyze from both a descriptive and statistical paradigm of language and behavior; anthropology looked at the nature of humanity and how it functions in different societies; and yet, this corpus of knowledge, as well as knowledge from the other social disciplines, failed to satisfy a definition of image production in more than one language.

Eleven years ago while working at Hunter College of the City University of New York in the Department of Curriculum and Teaching, I met a faculty member who was in a field called bilingual education. Although I was interested in cross-cultural studies, I read superficially in bilingual education. At that time, I was interested in culture and valued knowledge and I thought that the field of bilingual education could be seen as another perspective in cross-cultural studies in the bilingual situation. I never directly dealt with a teaching concept, which assumed that the social environment had to be viewed from a multilingual/multicultural perspective. I pondered the idea with some interest and finally accepted the challenge. At about the same time, I was invited to speak at a Portuguese conference in Massachusetts about Portuguese children who had entered a new culture. At that conference, I met a future friend who was interested in linguistics, especially in the area of language and thought in bilingual situations. This second encounter transformed my interest in bilingual education into a serious inquiry as I started to realize that this area of study had something new to offer in the curriculum field. What it was, I did not know, but I had an urge to find out. Finally, I immersed myself in the literature. I learned that some of it was sheer nonsense, but some of it opened the doors to a new paradigm; a paradigm that, for me at least, could possibly describe the relationship between language and thought in the domain of the distribution of knowledge, especially in the area of image production and testing.

One day in my office, I was reflecting about language and culture and I started to compare my cognitive systems in such a way that I could view the world or valued knowledge(s) from two different perspectives. I

was feeling through cognition and language-meaning and how it shifted from one culture to another. To say this was my first experience would be false. Most bilingual-bicultural individuals know this phenomena by early adulthood, but it was the first time in my life that I knew that I was onto something new. For me, at least, now I could seriously think out (or think through) the phenomena of being bilingual. This encounter made me realize that it was possible to have different images for the same words in different languages. I realized that norms were based on a collective agreement of a specific distribution of knowledge legitimated in a statistical paradigm. Language and thought could be combined into images. Images that represent a collective agreement of what is a "correct" response.

The purpose of this presentation is then to explore the issue of language and thought through image production in testing the bilingual child. For this paper, image production is defined as verbal expressions, or explanations, that are used to describe or explain a specific concept of behavior. In addition, using the concept of image production, using the concept of the compound/coordinate paradigm, and combining these mentioned notions in the domain of language usage and (cultural) behavior, this paper will attempt to explore assessment and testing issues in the field of bilingual education. The following assumptions will be analyzed for the purpose of investigation:

1) A one-to-one correspondence in language translation does not always take place especially when one uses images for the purpose of test construction.

2) Language assessment in a particular language may necessitate the "proper" images that are socially constructed from a particular language group or social class.

3) Image production may be a new way of analyzing language, bilingualism, assessment and testing.

4) The Hidden Curriculum and Political Ideology.

As noted in the introduction, the methodology for this paper will be borrowed from the reconceptualist movement in curriculum theory where most of the paper is written in the first person. This allows the author/researcher to explore and combine the existing data in research with a personal or biased point of view. This methodology also allows one to explore the affective, as well as the cognitive, domain in the field of language assessment. The theoretical framework of the chapter will be from the point of view that all language and knowledge is socially constructed. I will not attempt to dwell on the political aspects of language policy and testing in the schools. However, many of my examples may be interpreted as a political expression. Also, I will not attempt to dwell on ethnicity, but I cannot "lock myself out" of this process of thinking when one describes language, images through language and the relationship of language and cultural behavior or cultural reproduction. What I will attempt to do is to "flow through and around" possible ways of explaining what it is to be a "balanced" bilingual and how this phenomena could be used to analyze image production and knowledge use in testing the bilingual child.

What is important is to discuss the concept of *valued* knowledge and how valued knowledge affects the individual when he/she is aware that one concept used in two different language systems is valued (felt) in different fundamental ways. Let us now explore some of these issues from the point of view that image production may be a new way of developing a paradigm for further research in assessment and test construction for the bilingual child.

A One-to-One Correspondence in Language Translation Does Not Always Take Place Especially When One Uses Images for the Purpose of Translating.

Apple and King (1979) hint at the problem of valued knowledge through the definition of the hidden curriculum, but they limit their hypothesis to language control in a monolingual situation vis-a-vis social class, a fact of which many bilingual educators are well aware. They state: "Just as there is a social distribution of cultural capital in society, so there is a social distribution of knowledge within classrooms. For example, different 'kinds' of students get different 'kinds' of knowledge." For the curriculum theorist who is in the bilingual field, the above assumption is a self-evident political reality. It is known that knowledge and how it is used relates directly to pedagogical principles in a multilingual, multicultural school experience. In the field of bilingual education and in testing the bilingual child, the problem is the distribution of knowledge of the two cultures or two cognitive styles and how the educator can treat different value structures in an educational setting. Curriculum theorists must confront this issue so they can develop a viable resolution in the area of pedagogical methods and the area of two languages in the classroom, i.e., the relationship of language and the social distribution of language.

Why is the above discussion important? If the curriculum theorist examines schooling from a monocultural point of view, based on the premise that language distribution exists only in *one* social context, she/he is in a poor methodological and epistemological position to examine the meaning of how one can move in, out, and around, under or over a concept as it relates to multidimensional meaning in a social context. The assumption here is that *all* knowledge is socially constructed and, hence, all knowledge is valued in society. This also includes the physical sciences. Kuhn (1962) presents this problem through paradigmatic construction in the so-called "hard sciences." His basic contention is that all "scientific knowledge" is developed through social paradigms vis-a-vis language. If this line of thought is accepted, then one may be able to compare values from different phenomenolingual paradigms in the distribution of the social sciences.

Bilingual educators may have the bicognitive "tools" to analyze and implement interrelationships between and among the social and physical sciences. I realize that one may argue the point that one has to be knowledgeable in each discipline to take on such a task, a fact that is impossible. Yet, one can still be a generalist in the curriculum field and develop several models relating to testing in a multilingual and multicultural situation.

The question basically is the following: If a person is a "balanced" bilingual, and, therefore, "bicultural," does language shift involve a different social view of the world? For example, as a balanced bilingual possessing native-ability in two languages, I can conceptualize and commiserate with the many faceted functions of two different types of "family." When I use the term "family" in English, possibly in an industrial context, the term implies specific values that may or may not fit the Portuguese rural term "familia." In English, family is usually synonymous with the notion of nuclear family, i.e., the inclusion of only parents and siblings. In this technological-industrial context, I can extend this concept to a political level where family may be defined as the single parent family, married people without children, the homosexual family, and other concepts that do not fit the Portuguese concept of "familia." Other valued aspects of "family" are the divisions of labor between men and women, the use and respect of one's private psychological space, and the sharing of decisions in terms of money, buying a home, the relationships with the children and so on. The nuclear family values independent behavior from other people, especially when such behavior involves discussing personal and financial problems with people outside the clan.

The Portuguese term "familia," however, recontextualizes my value image system. My images and feelings about the "familia" now focus on an extended relationship. Although my immediate family is still a spouse and children, my parents are also considered a part of the "familia." My aunts, cousins, and godparents are seen as part of the "familia." Decision making does not necessarily belong to the father, but to the person who is "older and wise" and who is more influential in the clan. Children's roles are strictly defined, and there is rarely a distinction among childhood, adolescence, and adulthood. You are either a child or an adult. Women and men have different roles as to the division of labor in and outside of the home. "Para ajudar a familia," (to help the family) is such a strong commitment to the clan that any family member who fails to take on responsibility for helping out in a crisis is looked down upon. Another term that may seem like a simple analysis until examined further is the concept of automobile. When I say "Carro" in Portuguese, my thought processes or cognitive style perceives Fiats, VW's, Toyota's, etc; but when I shift my language to English, I automatically think of Fords, Chevrolets, Plymouths, etc. Further, if I think of luxury cars in both languages, something similar happens. In Portuguese, Mercedes Benz and BMW's become a part of my social-cognitive style; and in English, Oldsmobile Model 98's and Buick Electra 225's become a part of my phenomena of everyday life as it relates to luxury automobiles.[1]

The phenomenon of operating in two cultural cognitive styles makes me aware of the multitude of cultural variables that affect my behavior. When I attempt to apply these givens in the pedagogical field, I must think about the images in valuing the two different environments. Another example of bicognition may be seen in participation in a New Year's gathering. In the Portuguese working-class society, one way to participate in the ritual of leaving the old year behind is to attend a local social club. I am expected to bring my children along so the community may celebrate together by dancing and talking about the past year. The important aspect is that children do participate in all events. In contrast, my Anglo friends view this ritual in a different way. Usually, the children are not allowed to be part of this event, especially if the group decides to go out for dinner and dancing.

Whatever the psychological and sociological analysis is for the valued behavior, it is not relevant for this paper. What is important is that I can "know" and "feel" the two events as different ways of viewing the world.

A final example, of many, is that of couples walking down the street together. For the last several years, I have observed language use and how language functions in a social encounter within cultural norms. For example, when American English-speaking monolinguals socialize with another couple, the way they walk together gives some interesting clues to language and culture. In this case, most American English speakers walking down the street together as couples will converse as a foursome. Very rarely will the males and females segregate themselves in a one-to-one conversation, male-female, but the dialogue will tend to center itself back to the group. Whe two Portuguese couples get together and walk down the street, the phenomena of social relationships or social talk generally will change. In most cases, the males will talk to each other and the females will do the same. What is more fascinating is that usually the males will walk in front of the females and they will segregate the conversation. Although this behavior has many political and sexist implications, the reality is that language use and social relationships have a direct involvement in the reality of everyday life.

As Berger and Luckmann state:

> "Language forces me into patterns . . . . Language provides me with a ready-made possibility for the ongoing objectification of my unfolding experience. Put differently, language is pliantly expansive so as to allow me to

[1] Since the gasoline crisis, there has been a shift in my own value distribution of knowledge.

> objectify a great variety of experiences coming my way in the course of my life. Language also typifies experiences, allowing me to subsume them under broad categories in terms of which they have meaning not only to myself but also to my fellowmen." (1967)

This analysis, although limited in scope, provides the curriculum theorist with a clue on how image production, assessment, and test construction via the role of language relates to the social environment. When two or more languages are used, it is possible to contrast and "flow-through" the use of two or more sets of valued knowledge. What is more important is that the curriculum theorist cannot talk about knowledge and learning until she/he is able to specify what is considered "worthwhile" in the context of a specific culture.

Using the above examples, one may note separate image production is closely related to coordinate bilingualism, rather than compound bilingualism (Weinrich, 1953; Ervin and Osgood, 1954; Lambert, 1958). Lambert (1958) popularized the compound-coordinate paradigm through the use of mediation theory of semantics. If the person is a coordinate bilingual, his/her mediating semantic responses will elicit differing responses based on the language of the SIGN. Likewise, if the person is a compound bilingual, the mediating, semantic responses will be the same regardless of the SIGN language of the response. For coordinates, the corresponding pairs of terms in two languages signify one "semantene." Compound bilinguals are for those when corresponding terms signify a single semantene regardless of the language itself (Martinez-Boyd).

Image production becomes a useful tool in testing situations to reproduce items that are both valid and reliable. If one is in a coordinate situation where differing situations exist, the images of *each* sign (usually called cultural variables) may be identified for further test construction. If only one sign is in use (a compound situation), the variable(s) may be explored for further analysis: 1) That the specific sign is not known and an item may be developed in a test for discrimination; 2) The item (image) becomes a problem of cultural capital or distribution of knowledge where the bell-shaped curve may bias a specific social class. In either case, compound or coordinate, image production may be the avenue for further test production.

Language Assessment in a Particular Language May Necessitate the Proper Images That are Socially Constructed From a Particular Language Group.

The question is then, what is the proper image that may be reproduced for test or assessment procedure? It seems that image production is directly related to language production and valued knowledge of a specific culture. That is, each culture has its own collective agreement of what is "correct" and what are the "proper" responses for individuals in an assessment or testing situation. My examples of the family/familia or code switching in Portuguese and in English are perfect illustrations of different but "correct" images of two different cultural behaviors.

What this means is that educators must approach testing and image production by two sets of factors, or what I term system and context. In the system approach, the theorist will apply general theories of learning or general theories of behavior to a specific situation. For example, one may select and develop a curriculum around programmed instruction. Then the theorist must apply the *CONTEXT* (cultural variables) that will fit the valued knowledge within the specific culture (Simoes, 1979). Another model, of many, would be the use of Freud's exploration of human behavior vis-a-vis the superego, ego, and the id. Here the system is the conflict between the superego and the id which results in an outward behavior called the ego. Applying this system to context, however, is crucial for positive and negative learning prescriptions. Again, when *two* or more contexts are functioning within one logical system, the theorist has different ways of analyzing a paradigm. Another way to look at this issue is as Christian states that the child develops, between the "I" and the "me," a system of communication that is normally monolingual, representing a relative coherent and consistent system of concepts, but this situation may change rapidly when the "me" starts speaking another language. This change implies a new set of social rules for the "I," a new and different type of conversation between the "I" and the "me," a new self-concept for the person (Simões, 1976). This implies that educational theorists must be knowledgeable about the differences of image production and how this process is directly related to knowledge use and knowledge learning. When the theorists can be aware *and* know two or more contexts within one system, then there is a possibility of understanding the nature of assessment in testing in the bilingual child. The field of bilingual education may supply new information in this domain.

Image Production May be a New Way of Analyzing Language, Bilingualism, Assessment and Testing.

As noted, being a balanced bilingual gives the viewer of the world two structured valued knowledge systems and the feeling of what is different in two languages. To "feel different words," to "feel different domains of language," to "feel different social constructions of reality,"[2] to "feel different senses of touch," as they relate to cultural forms, to "feel or see different perspectives of the world," to "feel different forms of mathematics as the world is divided through language and signs vis-a-vis culture," *and so on*, gives the bilingual/bicultural person a way of analyzing the world

[2] I have used the term "reality" in the singular but used the term constructions in the plural. Either use of these terms could be analyzed from different perspectives.

through a feeling of relativity and how *all* knowledge is based on valued assumptions.

Now I ask: Can a monolingual person view the world from different perspectives and analyze and construct a world view from the point of view of interdisciplinary studies? Contrary to what one might expect, the answer seems to be affirmative. There are many ways to view language. Even within one language system, there are varied styles and dialects, and these dialects are usually divided along social class lines. A word in one social context may have a different meaning in another social context. Unfortunately, most educational institutions promote only one type of language system, called standard language, and other dialects are usually seen as an inferior language system which is frowned upon. Valued knowledge, language and the interrelationship of both domains *cannot* be talked about or be implemented as interrelated phenomena unless specific interrelationships are known by the theorist, be it from one language or phenomenon borrowed from one social class to another. In other words, language cannot divorce itself from the phenomena of reality or everyday life and different forms of language, be it a dialect or not, and cannot be imposed upon a single set of criteria in pedagogical situations.

What becomes a more important problem is the concept of compound bilingualism in the area of testing, assessment and image production. I do realize that there are problems with the use of the distinction of the compound/coordinate paradigm and there is extensive literature since the early days of Weinreich (1953), Ervin and Osgood (1954), Lambert (1958), and MacNamara (1970). One also has to realize that there are other schools of thought regarding language in a one-to-one correspondence or the "mixed language theory" (Huerta, 1977; Chimombo, 1978; Martinez-Boyd, 1978). One may use the concept of "additive" versus "subtractive" environments (Cummins, 1976) as a reference for the coordinate and compound distinction. The literature is rich indeed of how bilinguals perform on certain tasks in a continuum of language development (for example: Sienkiewicz, 1974; Simões, 1976; Hoffmann and Ariza, 1978; Redlinger, 1979; Ben-Zeev, 1977; and Doyle, 1977). The problem in our research data base is that we still do not know where the semantic system changes to different signs for the same concept or test item. Hence, there really is no distinction between a compound bilingual and a coordinate bilingual. It is really a continuum in language development in that "bilingualism" is only produced when two or more images are produced by the same sign. It is in this context where image production becomes a viable methodology for further research in language assessment and test construction. Language is in some aspects an acquisition of knowledge by a specific social class. Language does distribute knowledge and images or behavior. When two or more languages are acquired, the same principal still remains in effect. That is, language(s) are produced by a social distribution of knowledge.

The Hidden Curriculum and Political Ideology

Before closing, I think it is important to make reference to what I term the hidden curriculum and political ideology with reference to language use. It is a known reality that schools still seek to maintain a specific cultural capital through each social class.

To attempt to impose a language on students who do not speak it or to impose a specific valued knowledge system implicitly violates the civil rights of every child who is linguistically or socially different. The so-called accepted knowledge norm of the educational establishment may be hazardous to the children's right to fully participate in an integrated society. Haugen develops this theme when he talks about language. He states:

> The concept of 'norm' in reference to language is highly ambiguous and slippery. It may refer to a standardized language like French, codified in grammars and sanctified by an Academy, taught in schools, and written by authors, but spoken by no one, except under duress. Any deviation from such a norm is deemed to reveal one's lack of a proper education, and is regarded as barbarism if it is unintentional. But it may be acceptable if it is an intentional stylistic variation, either as a mockery of the lower classes or as a relaxation of standards . . . (1977)

Unfortunately, most curriculum theorists, if not all, cannot divorce themselves from political ideology and hence it is sometimes difficult to separate an educational or a language analysis from political thought. The use of image production in the analysis of bilingualism and biculturalism may suggest a new restructuring of political ideology as it relates to curriculum theory. It suggests that no curriculum theorist can "objectify" an educational process without knowing the cultural variables of the learner.

I believe that it is legitimate to state that the schools have locked out cultural differences, which of course, are language and dialectical differences in the racial and language minority communities. Educators have not, until recently, looked at this process as a pedagogical force that may create new ways of looking at cognitive differences and how these differences relate to knowing more than one way of viewing the world. When I hear the Black community stating that most Blacks know how to "play the White man's game," or when I hear a bilingual state that she/he can move in and out of one culture to another, it suggests to me that many social science skills have been developed by the voices of minority groups. More importantly, it strongly suggests that curriculum makers have not seriously looked at cultural differences as a possible way to shape school programs around meaningful experiences for children. What has happened is that a political model has superceded the pedagogical/language model and, consequently, most schools were based on a monosocial-class paradigm that fits only the "selected few" in our society.

Bibliography

Apple, Michael W., and King, Nancy. "Economics and Control in Everyday School Life." *Ideology and Curriculum*. Boston: Routledge and Kegon Paul, 1979.

Ben-Zeev, S. "The Effect of Bilingualism in Childhood from Spanish-English Low Economic Neighborhoods on Cognitive Development and Cognitive Strategy." In Ontario Institute for Studies in Education (Ed.), *Working Papers on Bilingualism/Travaux de Recherches sur le Bilingualisme, No. 14*. Toronto, Ontario: author, October, 1977.

Berger, P. L. and Luckmann, T. *The Social Construction of Reality: A Treatise in the Sociology of Knowledge*. Garden City, N.Y.: Doubleday, 1967.

Chimombo, M. *A Study of Code-Mixing in Bilingual Language Acquisition*. December 1978. (ERIC Document Reproduction Service No. Ed. 169 786)

Christian, Chester C., Jr. "Social and Psychological Implications of Bilingual Literacy." In *The Bilingual Child* (A. Simões, ed.) New York: Academic Press, 1976.

Cummins, J. "The Influence of Bilingualism on Cognitive Growth: A Synthesis of Research Findings and Explanatory Hypotheses." *Working Papers on Bilingualism, No. 9*. Ontario, Canada: Ontario Institute for Studies in Education, April 1976. (ERIC Document Reproduction Service No. ED 125 311)

Doyle, A., Champagne, M., & Segalowitz, N. "Some Issues in the Assessment of Linguistic Consequences of Early Bilingualism." In Ontario Institute for Studies in Education (Ed.), *Working Papers on Bilingualism/Travaux de Recherches sur le Bilingualisme, No. 14*. Toronto, Ontario: author, 1977.

Ervin, S. & Osgood, C. "Language Learning and Bilingualism." *Journal of Abnormal & Social Psychology*, 1954, *49*, 139-146.

Haugen, Einar. "Norm and Deviation in Bilingual Communities." In *Bilingualism* (P. Hornby, ed.) New York: Academic Press, 1977.

Hoffmann, C. & Ariza, F. *Bilingualism in a Two-Year-Old Child*. Paper presented at the British Association for Applied Linguistics, Annual Meeting, 1978. (ERIC Document Reproduction Service No. ED 171 150).

Huerta, A. "The Acquisition of Bilingualism: A Code-Switching Approach." *Working Papers in Sociolinguistics*. Austin, Texas: Southwest Educational Development Lab, 1977. (ERIC Document Reproduction Services No. ED 155 416)

Kuhn, Thomas S. "The Structure of Scientific Revolution." Chicago. The University of Chicago Press, 1962.

Lambert, W. E.; Havelka, J. & Crasby, C. "The Influence of Language Acquisition Contexts in Bilingualism." *Journal of Abnormal & Social Psychology*, 1958, *56*.

MacNamara, J. "Bilingualism and Thought." In Alates, C. J. (Ed.), Report on the 21st *Annual Roundtable Meeting in Linguistics & Language Studies*. Washington, DC.: Georgetown Univ. Press, 1970.

Martinez-Boyd, D. "An Investigation of Preschool Monolingual and Bilingual Children's Understanding of English and Spanish Grammar." Unpublished Master's Thesis, Catholic University of America, 1978.

Martinez-Boyd, Diana. "Childhood Bilingualism, Cognition and Metalinguistics-A Review of Recent Research." Unpublished paper.

Redlinger, W. "Early Developmental Bilingualism: A Review of the Literature." *Bilingual Review*, 1979, *VI* (1), 11-30.

Sienkiewicz, L. "Phonological Evidence for Coordinate and Compound Bilingualism." *Papers in Linguistics, 1974-1977: A Collection of MA Papers from the Students in the Linguistics Dept. of Northeastern Illinois University*, 1974. (ERIC Document Reproduction Service No. ED 161 291)

Simões, Antonio, Jr., ed. *The Bilingual Child*, New York, Academic Press, 1976.

Simões, Antonio, Jr. "The System-Context Approach to Curriculum Theory in Bilingual Education." In *Bilingual Education Teachers Handbook*, (M. Montero, ed.) Cambridge, Mass.; National Assessment and Dissemination Center, 1979.

Weinreich, U. *Language in Contact*. New York: Linguistic Circle of New York, 1953.

THE PARADOX OF THE PROPER NORM FOR ESL EVALUATION

Gerald E. DeMauro
New Jersey State Department of Education

Abstract

The question of what are appropriate norm groups for English language proficiency tests used to evaluate English skills growth is critical. Because all students who are identified as limited in English proficiencies must be served by programs, control groups from whom treatment is withheld are usually not available. Typically, the test norms are used as proxy comparison groups to circumvent these problems. However, instruments of this type are generally normed on monolingual or mainstream English populations, who may not be appropriate standards to measure pretest to posttest growth. They have had greater exposure to English at the time of pretest. To achieve the same score, their rate of skills acquisition must therefore be different. A possible solution may be to norm these tests on limited-English-proficient students. However, since these students must receive supplemental English intervention, they cannot constitute an expectation of how much students would have gained without the program. A better solution is derived from the concept of discriminant validity. A totally discriminant test completely separates limited-English and English-proficient students, while a nondiscriminant test fails to separate these populations. Analyses of these two extremes suggest strategies for using English norms developed from mainstream students as proxy evaluation comparison groups. Examples of these strategies and data generated from their use in New Jersey studies are presented.

Overview

In New Jersey, English language proficiency is defined in normative terms. Students who speak other languages at home are placed into English as a Second Language (ESL) or bilingual education programs which have ESL components if their English language skills are poorer than those of their grade peers. This judgment is made by assessing student performance on placement tests normed on monolingual English students. Because ESL and full bilingual education programs receive state funds, they must be evaluated. All limited-English-proficient (LEP) students must be served by these programs. Therefore, evaluation designs cannot use control or comparison groups in the usual sense because treatment may not be withheld from any students with similar needs.

The instrument programs use for placement should be sensitive to differences in English skills and also to changes in students' English proficiencies. It might also serve to evaluate the impact of state programs on English skills by using the normative sample as a proxy comparison group. Such a solution would greatly economize testing in bilingual education programs.

This solution is similar to that employed by the Title I (now Chapter 1) Model A1 (Tallmadge and Wood, 1976). As in bilingual/ESL programs all students with identified needs receive treatment. Pretests and posttests are analyzed in norm-referenced scores, usually Normal Curve Equivalents (NCEs). When students score the same NCE on an April posttest as they had on an October pretest, they have gained approximately as much as students in the test norms over these six months. Increases in NCEs mean that program students have gained more during some specified period (usually a school year) than might have been expected from the scores of students in the normative samples (Tallmadge, 1976).

On the surface, this solution offers bilingual and ESL programs a means of using test norms as proxy comparison groups. Students could be administered an instrument normed on monolingual English students. Their NCE scores could be used both as one criterion for program entry,* and as a pretest for program evaluation. Such a model certainly economizes the testing program.

Unfortunately, this model assumes that growth in English skills made by the monolingual English normative sample is a reasonable expectation of growth for limited-English-proficient students.

For example, third grade Title I students are probably homogenous in both Title I and school experience. Students in the test's norm samples for that grade are similar in school experience and have had similar exposure as the Title I students to the area being measured. Bilingual education classes are not as homogenous. Moreover, their exposure to the skill being measured by an English proficiency test is different from the English exposure of the normative sample.

Clearly, use of monolingual English norms for evaluation purposes is not a simple issue. This paper explores various applications of these norms, considerations that must be made in their use and some alternate approaches to the evaluation issue. The first question a test user must ask is whether one instrument may be valid for placement and evaluation.

Validity of Measures

In this instance, the construct being measured is English proficiency in the academic setting. Placement

*If it were the sole criterion, it could not be used as a pretest. Student scores might regress to the mean and show spurious gains (see Tallmadge and Wood, 1976).

and evaluation are related because both are concerned with demonstrating differences in proficiency.

Tests which serve both functions should be sensitive at and below the cutoff (scoring criterion for program eligibility). This is insured during the item selection process. Less difficult items increase the test's sensitivity in this range. One means of determining if the test may serve both functions is reviewing item p-values to determine if they are not inordinately difficult for LEP students.

Validity and Cutoffs. Validity in placement instruments is determined by how well the cutoff is set. As it changes, the validity coefficient varies.

Crehan (1974) examines this relationship. He defines the validity of a mastery test as its ability to make correct decisions about students. In proficiency testing, the instrument should place LEP students into bilingual/ESL programs and English proficient students into the regular school curriculum. If two samples, LEP and English-proficient students, are tested, then validity is operationalized as the ratio of the number of students correctly placed to the total number of students tested.

Hambleton and Eignor (1980) provide two means of setting cutoffs which should maximize test validity. In the contrasting group method, plots of monolingual and LEP students' scores are made. The intersection of the scores is the first estimate of the cutoff. It is adjusted upward or downward to maximize the validity coefficient defined above.

The borderline group method uses teacher judgment to identify students who are at the borderline, in terms of being academically proficient in English. Their median score becomes the cutoff.

A similar method involves testing LEP and English-proficient students and analyzing group membership via discriminant analysis. The highest score for membership in the LEP group and the lowest score for membership in the English-proficient group should be averaged to yield the cutoff.

Ulibarri, Spencer and Riyas (1981) claim that variation in cutoffs causes disagreement between instruments on placement decisions. The cutoff determines if a child is proficient enough to be placed in a monolingual English classroom. Agreement between the test decision and the child's proficiency determines validity. Thus, the instruments vary in validity to the extent that they vary in how they would place a child.

Thorndike (1982) defines taxonomic validity in a similar manner. The validity to make placement decisions is the ratio of variance between groups to total variance. The square root of this is a point biserial coefficient. The coefficient is more informative than the proportion of correct placement decisions because it accounts for scoring variation on the test within groups.

Evaluation Test Validity. Validity of evaluation instruments depends on their sensitivity to the course curriculum. This is assessed by matching test objectives to course objectives.

The objective of ESL programs, as part of bilingual education or as independent programs, is to bring the English skills of participants to the levels of English-proficient age peers. Evaluation instruments should reflect initial deficits and increases in English skills.

Finally, the validity of both placement and evaluation instruments depends on their capacity to measure proficiencies relevant to English classroom performance. Test items should be embedded in academic materials, or what Cummins (1981) called "context-reduced," if they are to predict classroom proficiencies.

Homogeneity (Unidimensionality). The sensitivity of both types of instruments to differences in English skills can be maximized through the item-selection process. Samples of varying English skills, including LEP students (Angoff, 1971), should be tested.

Homogeneity of the instrument over the domain is central to the validity of both types of tests and to the validity of multiple uses of a test. Unfortunately, technical manuals rarely contain information related to this issue. Yet, under the common underlying proficiency model (Cummins, 1981), the English language skills of LEP students are predicted by an underlying linguistic construct (Cummins, 1981). This construct is manifest in language skills at all ranges of development. In placement tests, homogeneity allows comparison of the skills of English-proficient and LEP students because it assures the instrument is focused on the proficiency variable. (See Hambleton and Cook, 1977 for a review of this assumption.) In evaluation instruments, homogeneity guarantees that scoring differences are related to skills differences.

The test may be shown to be homogenous over the domain through factor-analytic or other scaling techniques. This also may be determined if the items retain their ranks in difficulty over populations of varying English skills. The hardest items for LEP students should also be hardest for English-proficient students. Variation in the difficulty ranks would occur if the test is measuring different skills in each group.

If the test is not homogenous, it may separate children on other skills besides proficiency. Moreover, if such a test measures different variables as students' proficiencies increase, it will be of little evaluation utility.

Reliability

Both placement and evaluation tests should be

internally consistent within areas (reading, speaking, etc.) and over time (test-retest reliability).

Tests that are particularly sensitive to surface skills (see Dieterich, Freeman and Crandall, 1979 for review) may show large student improvements over time by virtue of repeated exposure to its written and oral vocabulary. Tests like the 1982 edition of the New York *Language Assessment Battery* (LAB) offer alternate forms to avoid this familiarity. When multiple forms are used, e.g., one as a pretest and another as a posttest, the instrument should demonstrate alternate-form reliability. Horizontal (across form) equating is also needed to permit comparisons across levels.

Equating

Placement tests are used for program entrance and exit. During the intervening years, students change grades and, often, test level. Changes in scores may reflect variation in the objectives of these levels or in the abilities of the normative samples. Placement tests should be equated vertically, across levels.

Use of Norms

The instrument's ability to interpret English skills may be increased or reduced by the choice of scoring metrics. The current *LAB* (Riverside Publishing Company, 1976) is normed on monolingual English students. ESL evaluations using norm-referenced scores use NCEs or other standard scores which refer to these students.

Angoff (1971) and others argue that reference to a particular norms group imposes a meaning on results that is not only inaccurate at times, but is always becoming obsolete, since the norms are based on the distribution of skills at one point in time.

This argument applies to proficiency tests normed on monolingual English samples. Although the norm-referenced scores provide a comparative standard of skills of LEP students to monolingual English peers, this standard may be dated.

Moreover, test norms may not be as useful for ESL evaluation as they are for Title I. The Title I model assumes that, at the time of pretest, the normative sample for any grade had as much exposure to the skills being measured as the Title I group. This assumption permits an expectation for posttest scores to be derived from the performance of the normative sample. The expectation is compared to actual student posttest scores.

This assumption cannot be made in evaluating ESL performance because at the time of pretest, ESL students have had less exposure to English than their monolingual English grade peers. Consequently, a LEP student scoring the same on the pretest as monolingual English students in the test norms may well score higher on the posttest. The LEP student scored as well on pretests after less exposure. With an equal time of exposure between pretest and posttest and a structured curriculum, the LEP student should perform better.

It seems more appropriate to use students with similar exposure to English to norm evaluation tests. For example, fall performance of LEP students could be referred to LEP fall norms, and spring posttest performance might be compared to spring LEP norms in each grade.

Unfortunately, this model is controlled too well. Between the pretest and posttest, LEP students comprising the norms will have received ESL instruction (by law). The normative sample cannot serve as a proxy comparison group because its performance will not represent "no-intervention" achievement.

This problem can be avoided by norming tests on students who have not yet had ESL instruction. For most districts, indeed most states, this would require large constant flows of immigrants.

Even if this were possible, in states like New Jersey, such norms would be decidedly unrepresentative of the current LEP population in sociocultural factors. Angoff's (1971) comment on norms obsolescence would be most meaningful if immigration patterns continue to vary.

Also, the normative sample might not be an appropriate comparison group for second year, third year or more experienced ESL students. The problem is similar to that involved in using monolingual English norms. Such students have had more English exposure at the pretest than normative sample students. They should be gaining English more slowly than immigrants who have scored as well without as much exposure.

The final general consideration regarding the use of normative samples as proxy comparison groups is statistical power. LEP students should perform poorer than the monolingual normative sample, and the distribution of their scores should be confined to lower score ranges. However, unless the distribution is fairly normal, the statistical tests may lack sensitivity to student growth or to differences in group abilities.

For example, suppose every LEP student scores lower than the monolingual English normative group and received an NCE of 1. If the lowest ability student scores as well (in raw scores) on the posttest as the highest ability student did on the pretest, that student will still have an NCE of 1 on the posttest. The instrument would not have been sensitive to the student's growth. Such scores, in themselves, are meaningless for the teacher.

New Jersey Study

This difficulty is illustrated with scores from 731 New Jersey bilingual education students, who were cluster-sampled for a study of English language acquisition (DeMauro, 1981). They represented state

bilingual education students in socioeconomic status, program types, region of state and program experience. Students in grades 1-12 were tested. This sample is currently in its third year of assessment. (See Appendix A for a summary of the first year.)

The New York City *Language Assessment Battery* (Cumbo, O'Neill, Tilis and Weichun, 1976) was administered in October, 1979 and again in May, 1980. Scores were converted to NCEs based on spring norms available in the technical manual (Riverside Publishing, 1976) and fall norms available from the New York City Office of Testing. Both norms were developed from monolingual English samples.

The score distributions were compared to normal distributions via goodness of fit tests. Frequencies of scores were computed in intervals representing one tenth of the range of scores. Chi-squares were computed, for both raw scores and NCEs, for each year of program experience. Large numbers of first and second graders permitted separate statistical analyses in these grades. For other grades, the statistics were computed over levels of the test (grades 3-6 and 7-12), because fewer students were tested in each grade.

The results (Tables 1 and 2) reveal positive skews in the NCE distributions, except for second- and third-year second graders on the posttest. In these grades, the distributions are negatively skewed.

Table 1

Distribution of LEP Scores on the LAB Goodness of Fit Statistics and Standard Deviations

		Goodness of Fit Chi Squares[1]		Standard Deviations	
NCE Grade	Year	Pre	Post	Pre	Post
1	1	(5)71.841***	(5)37.963***	26.845	24.077
	2	(2)19.955**	(5)39.601***	26.000	23.719
2	1	—	—	15.298	20.364
	2	(4)13.953**	(4)15.920**	26.189	29.266
	3	(4) 9.867*	(4)22.447***	24.092	30.99
3-6	1	(1)16.920***	(2)15.027***	8.939	14.661
	2	(2)45.684***	(4)16.349***	17.134	17.184
	3	(3)20.524***	(3)30.538***	13.838	15.791
7-12	1	(1)47.806***	(1)30.638***	10.040	10.948
	2	(1)19.616***	(2)17.277***	13.594	12.699
	3	(1)28.910***	(3)11.900**	14.751	13.879
RAW					
1	1	(3) 5.650	(4)13.538**	8.060	8.217
	2	(2) 1.133	(4) 0.586	6.291	6.525
2	1	—	—	8.094	6.297
	2	(4)21.220***	(2)30.397***	8.475	5.640
	3	(4)86.321***	(1)31.920***	5.708	4.711
3-6	1	(2) 8.218*	(3) 6.282	17.352	17.340
	2	(3) 6.635	(3) 8.975*	19.207	17.505
	3	(4) 1.051	(3)10.031*	17.961	16.703
7-12	1	(3) 4.714	(3) 3.598	15.688	15.045
	2	(3) 4.287	(2) 1.766	15.039	13.657
	3	(2) 1.455	(2) 1.787	13.568	13.333

[1] Degrees of freedom are in parentheses

*p < 05
**p < 01
***p < 001

Table 2

Percentages of LEP Students Scoring Lowest (Floor) and Highest (Ceiling) Possible Scores on LAB

	Grade	Year	NCE Scores Percentage at Floor	RAW Scores Percentage at Floor	NCE Scores Percentage at Ceiling	RAW Scores Percentage at Ceiling
Pretest	1	1	34.85	3.03	1.52	1.52
		2	24.66	1.37	1.37	1.37
	2	1	63.64	4.55	9.09	9.09
		2	25.00	1.92	5.77	5.77
		3	12.00	2.00	6.00	6.00
	3-6	1	43.75	3.13	6.67	6.67
		2	27.85	1.27	1.27	1.27
		3	16.82	0.93	0.93	0.93
	7-12	1	62.33	1.30	1.30	1.30
		2	30.65	1.61	1.61	1.61
		3	30.99	1.41	1.41	1.41
Posttest	1	1	24.24	1.52	4.55	4.55
		2	16.44	1.37	2.74	2.74
	2	1	22.73	9.09	4.55	4.55
		2	7.69	1.92	15.39	17.31
		3	2.00	2.00	28.00	28.00
	3-6	1	28.56	1.56	3.13	3.13
		2	13.92	1.27	1.27	1.27
		3	7.48	0.93	0.93	1.87
	7-12	1	45.45	1.30	1.30	1.30
		2	32.26	1.61	1.61	1.61
		3	19.72	1.41	1.41	2.82

In fact, although there were too few students (22) to test the distributiuon, 22.7 percent of first-year second graders scored the two highest posttest scores. Among third-year second graders, 52 percent scored at least 39 on the 40-item test. Second graders' scores probably reflect the ease of this test level, which must also be sensitive to reading and writing skills of kindergarteners.

In a test that totally discriminates LEP from English-proficient students, all entering (first-year) LEP students would score an NCE of 1 because the highest LEP students would score lower than the lowest monolingual English student. This is an outcome of a very skilled norms group, rather than of an overly difficult test. An overly difficult test would result in positively skewed raw scores, but the raw scores of bilingual education students on the *LAB* are normally distributed, while the NCE scores are not. If the variation in raw scores of LEP students is meaningful in terms of abilities, then the conversion to NCEs sacrifices this information.

In the case of a totally discriminant test, a linear transformation of the raw scores based on LEP student scores would provide the evaluator with more information concerning student gains. With such a

transformation, a totally discriminant test would separate those LEP students whose scores were most similar to English-proficient students. Providing the instrument were measuring the same trait across ability levels, this transformation would enable programs to use the instrument for placement and evaluation purposes.

This condition would be met if the orders of item difficulties were the same for LEP and English-proficient students. If the order varies with increasing proficiencies, the test is not measuring the common underlying proficiency. Separate analyses would be necessary for various proficiency levels, and the instrument would be of questionable utility for evaluation.

Even more homogenous tests may be better analyzed by blocking (partitioning) groups according to program experience. As Myers (1972) points out, such a design: (a) makes groups more homogenous regarding a variable that affects scoring; (b) yields meaningful interactions (as we shall see); and (c) is more efficient than a one-factor design. Cummins (1979) agrees and argues that ignored (not partitioned) sociocultural and linguistic factors often reduce the power of evaluations of bilingual education.

Within a grade (or test level, if the student sample is small) and language group, three LEP groups might be designated. The first has one or less years of experience in ESL. The second has two and the third has three years experience. Two empirical results recommend partitioning years of experience: (a) the proportions of students on the test floor (Table 2) rise as a function of years in program; (b) in New Jersey, students who have more program experience outscore those with less experience in English with reference to monolingual English norms. All groups are tested in October and again in April.

Such a design may be analyzed via ANOVA or a general linear model performed on NCEs derived from monolingual English norms. If one uses the norms of a language proficiency test, the proxy control group is probably a monolingual English sample. A between-subjects variable is program experience, which we will call Factor A. The effect of ESL experience estimates discriminant validity of the test, in the sense of Thorndike's formulation of variance between groups defined by English proficiencies.

All possible outcomes of this analysis are considered in terms of significant effects. A discriminant test will yield significant differences among groups based on program experience. A nondiscriminant test will not be sensitive to group proficiency differences.

Two models of outcomes are investigated: results of the totally nondiscriminant test and results of the totally discriminant test. The utility of placement tests normed on monolingual English samples for evaluation purposes will be examined under these extreme models. Considerations for the evaluation utility of the instrument under both extremes will apply to the evaluation utility of placement tests currently used by programs.

The criterion for utility under the extreme cases is whether or not the outcomes are meaningful in terms of second-language acquisition theory. If the theoretical factors which determine these outcomes are consonant with the ESL program, then the test is probably sensitive to the intervention. If they are not, then the instrument may be measuring extraneous variables.

For example, students may regress in English skills, relative to the monolingual English norms, if intensive English is introduced before the students have developed first-language competencies (Cummins, 1981). If first-language skills are assessed and strengthened and precautions are made against this, then the test may be measuring other extraneous factors, and its utility for evaluation must be questioned.

Finally, the totally discriminant test is one in which there is no overlap in scores on the pretest between English-proficient and first-year ESL students. A totally nondiscriminant test cannot distinguish these two groups.

Definition of Possible Outcomes

Table 3 and Figure 1 show the possible outcomes of the evaluation design described above. The four conditions represent variations of significant effects for totally nondiscriminant and totally discriminant instruments.

Table 3

Significant Effects in
Illustrative Evaluation Model

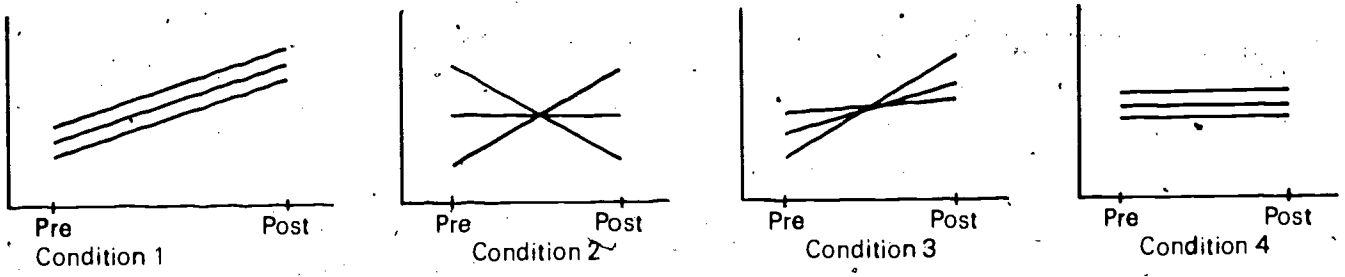
Condition	Test Type	Significant within Subjects Effects	Possible Explanation
1	Totally Nondiscriminant[1] Totally Discriminant[2]	Pre-post	Large between-subjects error Program effectiveness
2	Totally Nondiscriminant	Pre-post x Experience	First language deficiencies in younger group
	Totally Discriminant		Same
3	Totally Nondiscriminant	Pre-post, Pre-post x Experience	Individual differences in growth patterns offered by ESL experience
	Totally Discriminant		Course of curriculum
4	Totally Nondiscriminant Totally Discriminant	None None	Instrument failure Instrument failure

[1] Totally nondiscriminant tests do not distinguish groups by years of program experience.

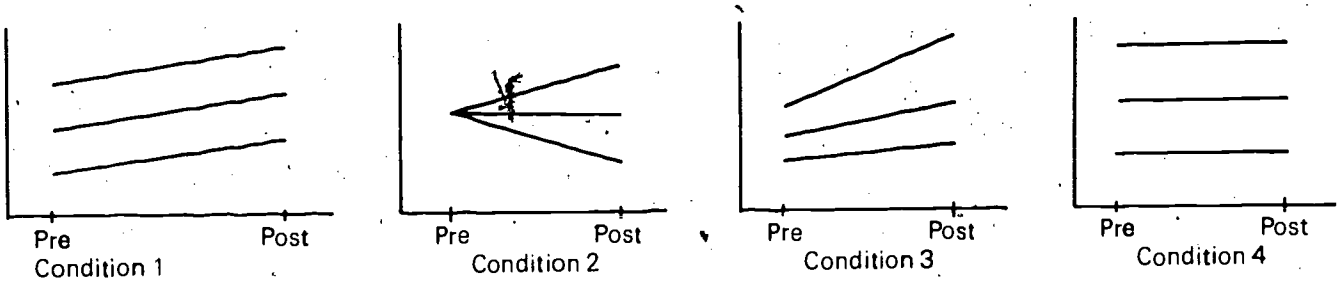
[2] Totally discriminant tests yield significant program experience effects.

Figure 1

Possible Evaluation Outcomes of
a Proficiency Test in a Mixed Design
(Pretest-Posttest X Program Experience).



Totally Nondiscriminant Test



Totally Discriminant Test

53

It is assumed that more experienced ESL students will score higher on highly discriminant tests. As students gain ESL experience, the distribution of their English proficiencies has greater overlap with the scores of monolingual English students.

Condition 1

Totally Nondiscriminant Case. Only pretest to posttest effects are significant. This would be an unlikely outcome for an instrument that could not distinguish students of different proficiencies. A large between-subjects error (S/A), but relatively small within-subjects error (SB/A), could account for this result.

Factors that are not partitioned could contribute to the between-subjects error. This agrees with Cummins' (1979) belief that subject factors influence language proficiency. For example, socioeconomic status significantly predicted English acquisition among students in New Jersey (DeMauro, 1981). Another source of unmeasured variance, especially among younger students is whether or not they have developed requisite linguistic skills in their first language. Such unmeasured sources may inflate error and cause underestimates of program effects.

Totally Discriminant Case. This occurrence is very plausible. Each group makes about equivalent pretest to posttest gains, and the more experienced groups score higher than the less experienced groups.

In the New Jersey study, Level II (grades 3-6) scores yielded results similar to these. Although other variables (language, socioeconomic status and grade level) were partitioned, there were significant pretest to posttest gains and students with more program experience outscored students with less program experience.

Condition 2

Totally Nondiscriminant Case. The amount of gain, here, depends upon program experience. A clear example of such an outcome is a crossover effect (See Figure 1). A loss in NCEs represents less growth than could be expected from the growth of the monolingual English normative sample. Such an unlikely interaction might occur if English skills development actually regresses relatively, given certain sociocultural or cognitive factors.

Cummins (1979) accounts for such cases in his developmental interdependence hypothesis. He believes that second-language competence depends on first-language competence at the time of intensive exposure to the second language.

If the group being evaluated is from one early grade, the pretest to posttest relative loss may result from intensive English exposure before the child had sufficient first-language competencies. (Third year children may have been exposed in preschool.) Snow and Hoefnagel-Hohle (1978) report that first-language fluency decreases in young subjects on intensive exposure to the second language. Competence in the second language would then also suffer if the two are interdependent.

If the group which regresses is the first-year group, the results might explain program effects. Perhaps this is a bilingual program which stresses initial strengthening of first-language skills, and the test measures English surface skills. There may be an initial relative loss in those skills (first year), followed by small increases or decrements during the second year, and a growth spurt during the third year, as students begin developing deep structure competencies. Cummins (1981) claims that language-minority students approach norms more easily in surface proficiencies than in "context-reduced" or academic proficiencies. Such an outcome would indicate sensitivity of the instrument to surface skills, but not necessarily academic proficiencies. The instrument should be used if these skills are of interest, but academic performance in English will probably not be predicted by these surface skills.

Totally Discriminant Case. Significant between-groups and interaction effects would appear as a fanning (See Figure 1). The developmental interdependence hypothesis might explain these. Students with poor first-language skills might continue to fall further behind monolingual English peers, while those with better first-language skills would grow quickly. If all students were in the same grade, those with more program experience may not have had the opportunity to develop requisite first-language skills.

One must be cautious that the instrument is measuring language proficiency over the full scoring range. For example, among higher ability students, the instrument should not measure another construct like creativity.

Condition 3

Totally Nondiscriminant Case. There are significant pretest-posttest and interaction effects. Although there is significant growth relative to monolingual English norms, there is no between-groups effect. Large between-subjects error variance would explain this. Snow and Hoefnagel-Hohle (1978) have reported the possibility that although second-language acquisition is predicted by patterns of first-language acquisition, there are large individual differences in these patterns. The interaction effect would indicate that such patterns are mediated by amount of ESL experience.

If there are growth types that characterize groups of children, then these groups should be analyzed separately. Results from tests that are sensitive to varied growth patterns may be difficult to analyze. A structured, sequenced ESL curriculum may reduce this individual variation. Caution should be taken that the instrument is sensitive to the curriculum.

Totally Discriminant Case. There are many explanations for all effects being significant. Some are appropriate to specific age groups.

Walberg, Hase and Rasher (1978) present results that would explain the circumstance in which the most experienced group is gaining least. They report that rapid gains in teacher ratings of the English skills of Japanese children often diminish over time. This occurs when the children are placed in a "homogenous" second-language environment, similar to how Cummins (1979) describes submersion programs.

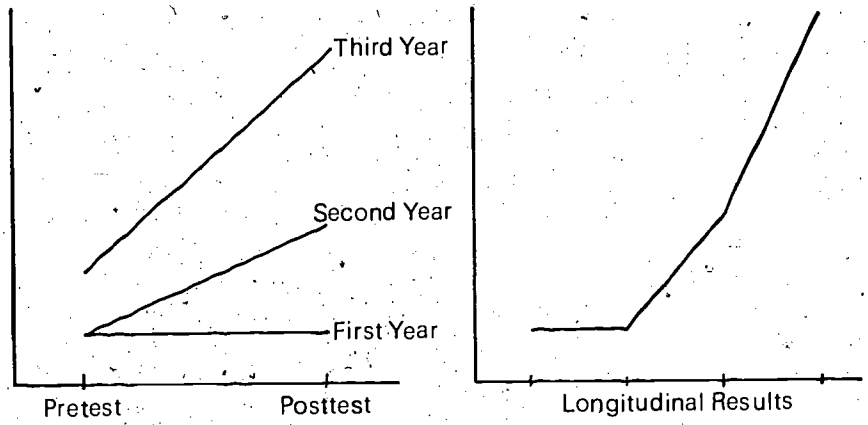
If the smallest gains are associated with less program experience, the outcome may reflect a curriculum which strengthens the native language skills of younger students in the first program year(s) before English is introduced to the classroom. More experienced children would appear to gain more during the same pretest-posttest interval.

If the largest gains are made by the middle experience group, it may be that the more experienced students have attained a ceiling on the test which has not yet been reached by the middle experienced students. Alternately, less experienced students may have initially shown English growth during a period of first-language strengthening. After a growth spurt in a later year, the students may begin to approach the test ceiling.

These explanations depend on English growth being curvilinear. Sampling students with different years of experience approximates sampling different points of the growth curve. Figure 2 illustrates this.

Figure 2

Hypothesized Curvilinear Functions of English Growth



One last explanation for the phenomena is that all program students in a grade received the same curriculum, regardless of program experience. This curriculum would have differential utility to the students, depending on their experience. An instrument sensitive to these effects would provide useful evaluation information and should be used.

Condition 4

Totally Nondiscriminant Case. Condition 4 could result from large individual variations in growth, which result in violation of the statistical additivity model for repeated measures (Myers, 1972). Perhaps the instrument measures skills that are unrelated to program exposure. If there are no systematic factors of interest to be partitioned (sociocultural factors, sex, etc.), the test should not be used for evaluation purposes. Alternately, the curriculum may not be helping students learn English.

Totally Discriminant Case. If the test is sensitive to differences related to program experience, it is not likely to be insensitive to growth within a year. Large within-subject error variance might account for this. Again, individual differences in growth rates may make the statistical model very conservative. If there are effects of interest that can be partitioned, this may reduce error variance. If not, the test is probably not very valuable for evaluation.

In the New Jersey study, significant effects were attributed to program experience in Levels I and III, without significant pretest to posttest effects. In Level I (grades K-2), grade level interacted with pretest to posttest growth. Second graders made significant gains relative to monolingual English norms, while first graders did not. In Level 3 (grades 7-12), the *LAB* did not reveal any patterns of interest in pretest to posttest growth.

Interpretation

The first decision to be made is whether the instrument's discriminant validity warrants its use for placement. Some manuals (for example, *Language Assessment Scales*, DeAvila and Duncan, 1975; *Miranda Test of English Proficiency*, Miranda and Associates, Inc., 1980), address this issue. Instruments that are valid for placement might then be considered for evaluation uses.

In the examples above, discriminant validity is estimated by how well more experienced ESL students are statistically separated from less experienced ESL students, relative to monolingual English norms. By current language acquisition theory, more experienced students would only score lower relative to these norms if they have been exposed to English before they were sufficiently competent in their native languages. When the level of linguistic competence is controlled, a test that is homogenous over the domain should, by the linguistic interdependence theory, yield higher scores for more experienced students.

One solution for using monolingual English norms to evaluate ESL programs is to partition groups by program experience. In this way, the standard for each group is not the performance of monolingual English peers, but rather, the relative performance of other program students, as they gain academic experience in English.

The evaluation hypothesis is that students in their last program year(s) will outscore students in their first program year(s), relative to monolingual English peers. By looking at pretest to posttest growth within experience cohorts, observations may be made about rates of growth, as well.

One final caution in using the monolingual English norms deals with the violation of the statistical assumption of normal distribution. These data will almost always be positively skewed. Table 1 shows the skews for pretest and posttests are homogenous within grades and experience levels. Such violations of the assumption slightly reduce the probability of computing large F-ratios (Myers, 1972).

Of course, the benefit of using NCEs based on monolingual English norms is that they have intrinsic meaning. These scores estimate student performance in English relative to that of mainstream students.

The following checklist provides a review sheet for programs to use in estimating whether a placement instrument will be useful for evaluation purposes.

Table 4

Checklist for Test Use in ESL Evaluation

Assurances	Related Condition
1. Content validity, homogeneity over domain	All
2. Measure and/or partition related sociocultural and other student factors	1, 2, 3
3. Measure first-language competencies	2
4. Measure "context-reduced" language skills	2

Totally Discriminant Test-Failure to Discriminate Experience Groups

The above conditions assume that totally discriminant tests would distinguish more-experienced from less-experienced groups. This assumption implies that there were monolingual English students in the normative sample who scored lower on the instrument than some of the LEP students. If there were no overlap in scores, LEP students of all levels of program experience would score an NCE of 1.

When floor effects are extreme, the monolingual English norms should be abandoned. A scale derived from LEP student scores would better serve evaluation. This should certainly be done if the high proportion of LEP students scoring a 1 does not improve with increases in program experience.

Raw pretest scores for each grade, within years of experience, should undergo normal transformation. Posttest scores could be converted to the same scale by subtracting each from the pretest mean and dividing by some function of the pretest standard deviation.

For each grade, a spring scale could be developed for each of three (or more) years of program experience. Transformations of scores for students entering in the fall could be interpolated. The interpolated value would fall between one transformation using the mean and standard deviation of first-year students in the students' current grade levels and another transformation using the mean and standard deviation of first-year students in the previous grade level.

This is a once-a-year testing model, except for two groups: 1) first year fall entrants who would be tested

on entrance and pretested in the spring; 2) students who change test levels and would have to be pretested in the fall of the year of change. The fall score would be an interpolated transformation similar to that used for fall entrants.

Each spring's scores would serve as a posttest for the concluding year and pretest for the coming year. These two functions could be served using two transformations. As a posttest, scores would be transformed based on the statistics of the previous year of experience, and as a pretest for the coming year, the transformation would use statistics for the coming year of experience. For example, spring pretest scores for first-year students would be transformed based on means and standard deviations for first-year students. Their posttest scores the following spring would be transformed using first-year mean and standard deviation for the next grade level. This same score could be used as a pretest for the coming year by transforming it based on the second-year mean and standard deviation in that grade level.

In this manner, the normalized scores would always control for grade progression without being confounded by identical program interventions (as with LEP norms) between treatment group and normative samples. Separate statistics for transformations need not end at three years of experience, but large enough samples with more experience may not be available to yield stable means and standard deviations. Table 5 shows this.

Table 5

Choice of Appropriate Scales
(Example)

Scale					
Grade	Experience	Group	Test	Experience	Grade
1	1	Group 1, pretest, first year—first grade			
	2	Group 2, pretest, second year—first grade			
	3	Group 3, pretest, third year—first grade			
2	1	Group 1, posttest, first year—second grade			
	2	Group 2, posttest, second year—second grade Group 1, pretest, second year—second grade			
	3	Group 3, posttest, third year—second grade Group 2, pretest, second year—second grade			

New Jersey data suggest that growth is a function of grade level. In high school the gains made in different grades are small and approximately equal. It may be more efficient to develop norms for whole grade clusters, e.g., 10-12, rather than within separate grades for these students.

Finally, the suggested scales are not without problems. They do not control for cultural exposure to English from the media, friends, etc.* However, this procedure, especially for more experienced students, may be the best approximation of English-skills growth as a function of program exposure.

Nonnormative Meanings of Test Results

Some available proficiency instruments, e.g., the *Bilingual Syntax Measure* (Burt, Dulay, and Hernandez, 1975) and the *IDEA Oral Language Proficiency Test* (Ballard and Tighe, Inc. 1979), provide results that are meaningful in terms of linguistic theory. These instruments hypothesize taxonomies of language skills and plot student progress along these theoretical structures (levels, etc.).

Meaning of results from normed tests usually comes from the norms, but other meanings might be given by relating student skills to the test results (Angoff, 1971). For example, teachers could group test scores (every five points) and record report card grades in ESL. For each student achieving the scores within every group, abilities associated with the modal course grades for each group of scores would be associated with the test scores.

Similarly, Findley and Nathan (1979) propose methods of evaluating ESL programs from a performance perspective. Teachers could rate whether or not students have mastered various performance objectives at the time of testing. Again, test scores might be blocked into five-point intervals, and an assessment may be made of whether or not most students scoring in each interval have mastered each objective. Lists may be made of mastered objectives associated with each scoring range.

Angoff also suggests that such methods may be reversed. Mastery of performance objectives and/or course grades could be standardized by showing the distribution of test scores associated with each objective mastered or different course grades. The modal test score (blocked in small intervals) would represent the score associated with each mastery or course level.

Hambleton and Cook (1977) explain that an ability scale could be determined using latent trait models. Scores on this scale are related to the probability of passing each test item and, therefore, have meaning in terms of test content. Programs with large populations might consider similar applications of latent trait models.

This option might also be used in reference to monolingual English norms. The test could be calibrated to describe proficiencies of monolingual

*Programs concerned with these factors might employ three-point models in which intervals spanning academic years are compared with summer intervals spanning vacation periods.

English students at each ability score. LEP students attaining various ability scores would have proficiencies similar to those of the normative student.

Finally, item calibration may be used both in criterion-reference testing and normative testing (Wright, 1977). Items may be drawn from banks that are tailored to students' proficiencies and referenced either to norm or performance criteria.

Summary and Conclusions

Placement tests and ESL evaluation tests both measure English proficiencies. The placement tests must determine how well students could perform in English in the academic setting. Since monolingual English students are normally instructed in this environment, placement instruments are reasonably normed on monolingual English school populations.

It is also reasonable that these instruments might be used to evaluate the ESL instruction, either as a component of bilingual education or as a program itself. Such use of these tests raises the question of whether test norms may be used as proxy comparison groups. The growth of monolingual English students, however, may not be an appropriate standard to gauge the growth of LEP students. Other norms based on LEP students are also problematic.

A proposal for using monolingual English norms, in which LEP students are grouped by program experience, is offered. The scores of each group relative to the monolingual English norms are compared. Such analyses are somewhat conservative. This use of monolingual English norms relies heavily on a control (or at least measurement) of first-language skills before intensive introduction to English and a review of the homogeneity of the test.

Finally, Angoff (1971) observes that norms acquire meaning through familiarity. In this sense, monolingual English norms are valuable because their meaning is immediate. They are used in placement to define how LEP students would perform in monolingual classrooms.

The placement and evaluation uses of these norms may not have different meanings. The placement test tells how far a student must go in the ESL program. The evaluation test marks his/her progress on the route. The placement test is merely one point along the route. Moreover, each test should have a meaning in terms of curriculum. The placement test should prescribe areas of particular linguistic need, and the evaluation test should mark student improvement in those areas. Both of these functions demand instrument sensitivity to the same student capacities.

Finally, we have discussed these instruments as though they yield a single score for normative comparisons. Actually instruments may provide separate reading, writing, speaking and listening norms. In this case, the arguments would apply to each of these areas separately. Each subtest should be homogenous, etc., placement could be accomplished through profiles across the four areas. This would help target curriculum to a particular student's needs. Separately normed subtests would also enable teachers to evaluate a specific student's improvements in the targeted needs areas.

Appendix A

Gains in NCEs on the Language Assessment Battery by Bilingual Education Students in New Jersey

Test Level	Years of Experience	Pretest Mean	Posttest Mean
1	1	25.5	29.2
(grades K-2)	2	31.6	35.3
	3	36.6	54.3
2	1	8.8	18.0
(grades 3-6)	2	17.2	24.2
	3	19.1	27.6
3	1	7.9	10.1
(grades 7-12)	2	14.3	14.5
	3	15.7	18.8

References

Angoff, W.H. "Scales, Norms and Equivalent Scores." In R.L. Thorndike, Ed., *Educational Measurement* (second edition). Washington, D.C.: American Council on Education, 1971.

Ballard, W.S. and Tighe, P.L. *IDEAL Oral Language Proficiency Test.* Brea, California: Ballard & Tighe, Inc., 1979.

Burt, M.K., Dulay, H.C. and Hernandez, E.C. *Bilingual Syntax Measure.* New York: Harcourt, Brace, Jovanovich, Inc., 1973.

Crehan, K.D. "Item Analysis for Teacher-Made Mastery Tests." *Journal of Educational Measurement, 11*(4), 1974, 225-262.

Cumbo, R.F.; O'Neill, F.J.; Tilis, H.S. and Weichun, W. The *Language Assessment Battery.* Iowa City: Riverside Publishing Company, 1976.

Cummins, J. "Linguistic Interdependence and the Educational Development of Bilingual Children." *Review of Educational Research, 49*(2), 1979, 225-251.

Cummins, J. "Four Misconceptions about Language Proficiency in Bilingual Education." *National Association of Bilingual Education Journal, 5*(3), 1981, 31-45.

DeAvila, E.A. and Duncan, S.E. *The Language Assessment Scales.* Corte Madera, Ca.: Linguametrics Group, 1975.

DeMauro, G.E. *The Impact of Bilingual Education on English Acquisition in New Jersey.* Trenton: New Jersey State Department of Education, 1981.

Dieterich, T.G., Freeman, C. and Crandall, J.A. "Linguisitic Analysis of Some English Proficiency Tests." *TESOL Quarterly, 13*(4), 1979, 535-550.

Findley, C.A. and Nathan, L.A. "Functional Language Objectives in a Competency-Based ESL Curriculum." *TESOL Quarterly, 14*(2), 1980, 221-231.

Hambleton, R.K. and Cook, L.L. "Latent Trait Models and Their Use in the Analysis of Educational Test Data." *Journal of Educational Measurement, 14*(2), 1977, 75-96.

Hambleton, R.K. and Eignor, D.R. "Competency Test Development, Validation and Standard Setting." In R.M. Jaeger and C.K. Tittle (Eds.), *Minimum Competency Achievement Testing: Motives, Models, Measures and Consequences.* Berkeley, Ca.: McCutchan Publishing Corp., 1980.

Jaeger, R.M. and Tittle, C.K. (Eds.) *Minimum Competency Achievement Testing: Motives, Models, Measures and Consequences.* Berkeley, Ca.: McCutchan Publishing Corp., 1980.

Miranda, L. *Miranda Test of English Proficiency.* Bethesda, Md.: L. Miranda and Associates, Inc., 1980.

Myers, J.L. *Fundamentals of Experimental Design* (Second Edition). Boston: Allyn and Bacon, Inc., 1972.

Snow, C.E. and Hoefnagel-Hohle, M. "The Critical Period for Language Acquisition: Evidence from Second-Language Learning." *Child Development, 49,* 1978, 1114-1128.

Tallmadge, G.K. *Interpreting NCEs: ESEA Title I Evaluation and Reporting System, Technical Paper No. 2.* Mountain View, CA: RMC Research, Corp., 1976.

Tallmadge, G.K. and Wood, C.T. *User's Guide: ESEA Title I Evaluation and Reporting System.* Mountain View, CA: RMC Research Corp., 1976.

Thorndike, R.L. (Ed.) *Educational Measurement.* (Second Edition), Washington, D.C.: American Council on Education, 1971.

Thorndike, R.L. *Applied Psychometrics.* Boston: Houghton Mifflin Co., 1982.

Ulibarri, D.M., Spencer, M.L. and Rivas, G.A. "Language Proficiency and Academic Achievement. A Study of Language Proficiency Tests and Their Relationship to School Ratings as Predictors of Academic Achievement." *National Association of Bilingual Education Journal, 5*(3), 1981, 47-79.

Walberg, H.J., Hase, K. and Rasher, S.D. "English Acquisition as a Diminishing Function of Experience Rather Than Age." *TESOL Quarterly, 12*(4), 1978, 427-437.

Wright, B.D. "Solving Measurement Problems with the Rasch Model." *Journal of Educational Measurement, 14*(2), 1977, 97-116.

LANGUAGE AND COGNITION: A VIEW OF THE CONTROVERSY FROM THE PERSPECTIVE OF BILINGUALISM

Neal Kirschenbaum, Ph.D.
Program Coordinator-Special Education
National College of Education

The multifaceted multidimensional fabric which comprises the individual and shared domains of language and cognition give rise to a complexity that beckons the researcher to investigation. Language and cognition, as well as their interaction, have been intensely examined by the research community. In spite of this vigorous investigation, many issues still warrant clarification. What role does language play in cognition? How do language and thought interact? These questions have created major controversies and resulted in voluminous reports, yet the controversies and questions still remain. The focus of this paper is on the interaction of language and thought, more specifically the role that language plays in the human thought process as seen from the perspective of bilingualism.

Theoretical Background

What role does language play in thinking? Is language a mandatory prerequisite for logical thought? These questions have been intensely debated and the opposing views of Bruner and Piaget regarding the interrelationship of language and thought have emerged and dominated the arena of psycholinguistics.

Piaget (1967) views language as an important facilitator of thought, but not as a foundation or a necessary prerequisite to cognition. Cognition, in general, and higher level logical thinking, in particular, do not have their roots in a linguistic base, but rather in the sensory interaction with the environment. As Inhelder and Piaget (1964, p.293) note,

> Whether a child understands words like "all" and "same" or any other form of words used to refer to the concept of class inclusion and similarity, whether he understands the sort of language we use to refer to the asymmetrical and transitive relation of a series, these are questions that are mainly dependent on the level which he has reached in the development of operational behavior — and that development is relatively independent of any other, because it is governed by its own laws of equilibration. We, therefore, say that language is not a sufficient and necessary reason for the process.

Thus, language is clearly not given the status of a mandatory requirement of thinking. Even abstract logical thinking developing in the formal operational stage of cognition, which is viewed by many as being synonymous with or at least embodied in language, is a view clearly rejected by Piaget. Piaget clearly states,

> . . . in spite of appearances and current opinion, the essential characteristic of propositional logic is not that it is verbal logic. First and foremost, it is a logic of all possible combinations, whether these combinations arise in relation to experimental problems of purely verbal questions . . . the real power propositional logic lies not in this (verbal) support, but rather in the combinational power which makes it possible for reality to be fed into the set of possible hypotheses compatible with the data (Inhelder and Piaget, 1958, p.253).

In contrast, Bruner (1964; 1966) opposes Piaget's model of the independence of thought and language. Bruner perceives language to be closely intertwined with cognition and as an underlying requisite to logical thinking. Bruner (1964) postulates that intellectual growth in man can only occur as a result of his representational and integrative abilities. In sustaining cognitive growth, one must be able to represent in an understandable fashion recurring events and regularities within his environment, as well as integrate past events with present. Bruner (1964) further hypothesizes three developmental stages of cognitive growth which are in fact developmental technologies or modes of internal representation of the complex environment which defines the existence and parameters of cognition. The first means of representation of which the child avails himself is the enactive representation which embodies a motoric representation of past events. Bruner (1964), p.2) cites the following example:

> We cannot, for example, give an adequate description of familiar sidewalks or floors over which we habitually walk, nor do we have much of an image of what they are like. Yet we get about them without tripping or even looking much. Such segments of our environment — bicycle riding, tying knots, aspects of driving — get represented in our muscles, so to speak.

This is followed by an iconic representation where the individual can represent events via perceptual imagery and the selective organization of these images. The most advanced representational technique is a symbolic representation of events. It is in this stage that Bruner postulates language to be the means enabling the formation of symbolic representations and to be the primary force propelling the child from the static iconic stage to the more flexible symbolic representation.

Thus, it is via language that we form symbols, and language and cognition become inseparably

intertwined, with the latter's very existence dependent on the former. Bruner (1964, p.4) claims,

> Once the child has succeeded in internalizing language as a cognitive instrument, it becomes possible for him to represent and systematically transform the regularities of experience with far greater flexibility and power than before.

He continues,

> In children between 4 and 12, language comes to play an increasingly powerful role as an implement of knowing. Through simple experiments, I have tried to show how language shapes, augments, and even supercedes the child's earlier modes of processing information. Translation of experience into symbolic form, with its attendant means of achieving remote reference, transformation, and combination, opens up realms of intellectual possibility that are orders of magnitude beyond the most powerful image-forming system. (Bruner, 1964, p.13)

Research evidence has been presented in support of both sides of this theoretical controversy. Bruner (1964) gives evidence that language can enable the transitional child to conserve if initially shielded from the perceptual distortion and allowed to verbally represent the situation prior to visually perceiving the event. Thus, the symbolic representation of language allows the child to surmount his being perceptually bound to the situation giving credence to the critical role of language in thinking. However, when Bruner's concept that thought has a linguistic base was extended to suggest that language training can develop conservation ability, the results of a study by Sinclair-de-Zwart did not support this hypothesis. It was determined that cognitively advanced children who demonstrated conservation ability did in fact possess a linguistic sophistication employing more relational terms, more highly differentiated terms and coordinated descriptions. However, training children to use these linguistic expressions did not improve their conservation ability, which bolsters Piaget's view of language as a reflection of thinking, rather than the basis for it.

Similarly, Furth (1966) compared the cognitive development of hearing children and deaf children who have no linguistic input and who develop no, or limited, linguistic ability. Bruner's theoretical perspective would dictate that deaf, linguistically limited children should display deficient cognitive growth, yet no substantive differences were noted supporting Piaget's contention that thought is not founded in language.

In this paper, similar to Furth (1966), the ongoing controversy of the role of language in cognition will be examined by investigating the cognitive development of a special population — bilinguals. The theoretical premise the reader is asked to accept is as follows: if Piaget's theoretical perspective regarding the independence of language and thought is correct, one would expect that bilinguals and monolinguals, despite their different linguistic development, should not display cognitive differences since language is not the basis for thought and even complex logical processes have their roots in environmental interaction as opposed to language. If, however, Bruner's theoretical stand on this issue is correct, and in fact, language and thought are synonymously inseparable with language being the core of the symbolic mode of representation, then the wider linguistic range and more diversified linguistic structures of the bilingual child should have an enhancing effect on his cognitive development. The bilingual child should display a qualitative or quantitative cognitive difference compared to monolinguals in the embedment of more advanced abstract or symbolic thought forms or in the display of a more rapid cognitive development.

Research Evidence

If intelligence tests are valid, research relating IQ scores and bilingualism can be brought to bear on this controversy. Piaget would maintain the expectation of no significant difference in IQ, while Bruner would expect bilinguals to attain higher IQ's because of their advanced linguistic and cognitive development. However, as was noted in a previous paper (Kirschenbaum, 1982) the results of correlational studies of bilingualism and IQ range the gamut from intellectual deficiencies to intellectual enhancement. The contradictory findings as well as the very limited generalizability of the results of most studies investigating IQ and bilingualism were largely attributed to gross methodological inadequacies which were outlined in a previous paper. (The reader desirous of a comprehensive critical analysis of this area should see Kirschenbaum, 1982).

The study by Peal and Lambert (1962) is one that does warrant further discussion because of its distinctive status of being methodologically rigorous employing balanced bilinguals, adequate controls, and appropriate standardized verbal and nonverbal intelligence tests in the proficiency language of the bilingual. (The only methodological criticism that is applicable is the lack of longitutional data). This study engaged 10-year-old monolingual and balanced bilingual French-Canadians and measured intelligence on both verbal and nonverbal IQ tests. Moreover, this study not only quantitatively computed the IQ's, it also qualitatively compared the intelligence of monolinguals and bilinguals finding several qualitative differences. In analyzing the performance subtests, it was determined that more than simple visual motor activity was involved in those subtests in which bilinguals surpassed monolinguals. These subtests required a symbolic reorganization involving abstract concept formation and an ability for quick cognitive reorganization and flexibility. Peal and Lambert (1962) hypothesized that because bilinguals have two words for each referent, i.e., two symbols for each object, bilinguals must conceptualize their environment in

terms of abstract, general qualities. This is especially true for the compound bilingual who learns both languages in the same setting and, therefore, with the same referents. Peal and Lambert (1962, p.14) state:

> This ability to think in terms of abstract concepts and relations, independent of the actual word, apparently is required in the symbolic reorganization type tests. The monolinguals may have been forced to form concepts or abstract ideas of things and may be more likely to think in terms of concretes. They could not be expected, therefore, to be as agile at concept formation as the bilinguals, and they might appear handicapped comparatively.

Additionally, the increased cognitive flexibility noted in bilinguals may be rooted in their consistant alternating from one language to the other. Similarly in the performance tests administered in this study, a cognitive flexibility is required since symbolic reorganization often calls for substituting one concept for the other.

A factor analysis of the bilingual and monolingual performance on the intelligence test also revealed that the bilinguals developed more independent abilities and a more diversified structure than monolinguals. This fact takes on extreme significance in light of Guilford's (1959; 1967) definition of intelligence as a general factor composed of smaller specific factors based on the person's experience. Thus, not only does the bilingual's cognitive structure differ, but this difference engenders an intellectual superiority when compared to the monolingual.

How does one explain these cognitive advantages observed in the bilingual child? Bruner can explain them easily citing language as underlying thought, and the wider linguistic range and structure of the bilingual enhances his use of the symbolic representational mode. However, how do Piagetians deal with these results? In attempting to reconcile these results with Piaget's theory one cannot quickly dismiss the criticism that this study is not longitudinal. Since this research is not longitudinal and represents correlation at a particular isolated moment, one cannot assume causality. Thus, we cannot determine if bilinguals are more intelligent because of their bilingual nature, or whether more intelligent, cognitively flexible people with a greater abstract concept formation ability tend to become bilingual.

There is, however, additional experimental evidence supporting the hypothesis that the cognitive development in the bilingual child is different from that of the monolingual child. Leopold (1971, p.14) found a "noticeable looseness of the link between the phonetic word and its meaning." Leopold (1939) conducted a longitudinal observation of his German-English bilingual daughter from the time she exhibited speech. Leopold observed that unlike monolingual children his daughter did not insist on stereotyped wording in stories, often substituting words in songs or rhymes. This represented a distinction between words and meanings, and thus, Leopold (1939; 1971) claimed that the bilingual developed an early, premature ability to form concepts and a realization that words are not fixed and concrete. Apparently bilinguals, as opposed to monolinguals, do not have a concrete attachment to words, but they do have an advanced ability to form abstract concepts, a finding supported in Peal and Lambert (1962).

This finding receives further support from Ianco-Worral (1970). This study attempted to test Leopold's hypothesis of an accelerated cognitive development in bilinguals, as specifically observed in the looseness between words and meaning. A total of 30 Afrikaan-English bilinguals (from a one-language home environment) divided in two age groups, 4-6 and 7-9, were tested in a series of two experiments. An equivalent group of English and Afrikaan monolinguals were also tested. A multifactor determination of bilingualism was conducted via mother interviews, a test of bilingualism, and teacher reports.

The first experiment tested the semantic-phonetic preference in the bilinguals and monolinguals. Each child was given a set of three words: a standard word and two choices, one which was semantically similar, and one which was phonetically similar (but semantically different). It was found that young bilinguals could semantically match words to a significantly higher degree than young (4-6) monolinguals. Both groups of older children (7-9) indicated semantic preference and chose to match words on a semantic rather than a phonetic basis.

The second experiment tested the bilinguals' and monolinguals' recognition of the arbitrary nature of *names* and their ability to separate names and meaning. In this experiment a questioning method of Vygotsky was used. The following are sample questions.

1. "Why is a cow called a 'cow?'" For all groups, bilingual and monolingual, the responses were largely in terms of attributed qualities. For example, "a 'cow is a cow' because you milk it."

2. "If we were making up names, could we call a dog a cow?" The results indicate that in both the younger and older groups, a greater number of bilingual children agreed in principle the names could be changed, while a majority of monolinguals said they could not.

3. "Let's call a dog, a cow — does this cow have horns?" The response to this question improved with age, but there was no significant difference between monolinguals and bilinguals.

It is concluded from the results of this study that Leopold's observation was valid, and the bilinguals' development indicates a two- to three-year advancement over monolinguals. This also points to a

superior ability to form abstract concepts and conceptual relationship because on a continuum of concrete — abstract, word meanings are more abstract than word sounds. The bilinguals' ability to substitute names in principle, as opposed to the monolinguals, also exemplifies their cognitive flexibility and advanced conceptual development.

Another study (Feldman & Shen, no date) supporting the hypothesis that bilinguals have a different cognitive development than monolinguals, tested five-year-old bilinguals' and monolinguals' ability in tasks involving object constancy, naming, and the use of names in sentences. It was assumed that since bilinguals have two codes for all referents, this would facilitate their awareness of the arbitrary nature of names. It was, therefore, hypothesized, that in the aforementioned tasks, a bilingual child would display advanced cognitive development, as opposed to a monolingual child.

The object constancy task entailed showing an object to the child and physically transforming it while viewed by him/her (e.g. a match burned, a cup crushed). It was then placed with an identical pre-transformed object, and the child was directed to pick up the object he was previously shown. The naming task consisted of the child's ability to call objects by their common names (call a cup, a "cup"); to learn nonsense names, (call a cup, a "wig"); and to switch common names (call a cup, a "plate"). In the sentence task, the child had to use these three labels (common, nonsense, switched common) in various sentences.

The results indicate that bilinguals had done significantly better on all three tasks, with the sharpest disparity in comprehension measures as opposed to production. (See Seidner, 1982.) A further analysis reveals that bilinguals were equivalent to monolinguals in their facility to acquire new names (nonsense labels) and their general knowledge of common names. However, bilinguals surpassed monolinguals in all other areas of naming, using names in sentences, as well as object constancy. Thus, it appears that bilinguals have a superior cognitive structure in these areas of development.

Further support for cognitive development differences between bilinguals and monolinguals is derived from the fact that bilinguals use more divergent thinking than monolinguals (Scott, 1973), which is considered an indication of creativity by several authors (Wallach & Kogan, 1965). In this study (Scott, 1973), divergent thinking was taken to be synonymous with a unique cognitive style, rather than a creativity predictor. Early divergent thinking was measured by the unusual uses subtests and late divergent thinking, by the lines and patterns subtests of the Wallach and Kogan tests.

The results of this study indicated more divergent thinking in bilingual subjects than monolinguals. Scott (1973, p.10) states, "The results of this study seem to indicate that bilingualism can both influence and be influenced by divergent thinking which operates as both cause and effect." The author discusses the results in terms of supporting the hypothesis that bilinguals have greater cognitive flexibility than monolinguals.

However, not all the evidence points to this differential development in bilinguals. Burling (1971), in observing his bilingual son, claims he did not find the looseness of linking words and meaning that Leopold (1971) discussed. In the longitudinal observation of his son, he claims:

> There can never have been a child more obsessional in this respect than Stephen. When we read to him he insistently protested the slightest alteration in any familiar text. This was true even though we read to him in his second language, English, where he would suppose the form and meaning to be the least rigidly identified. It must have been an idiosyncratic trait rather than bilingualism that freed Leopold's daughter from insistence upon stereotyped wording (Burling, 1971, p.185).

Another longitudinal study concerns the St. Lambert experimental school. This experimental school was designed to foster bilingualism by primarily instructing children in the second language. This experiment involves the experimental bilingual school, as well as English and French instruction schools used as controls. The control and experimental group were matched on a number of variables. However, the basic design of this study was to examine the feasibility of this experimental school in fostering bilingualism, not to examine cognitive differences of bilinguals and monolinguals. Therefore, there may have been some children who were to some extent bilingual in the control groups, thus limiting the applicability of this study. However, if bilingualism causes a differential development, one may expect some cognitive differences in children attending a school promoting bilingualism.

In examining the cognitive development of children after grades one and two, no cognitive inferiority or superiority over controls was reported. In a more comprehensive report (Bruck, Lambert, & Tucker, 1973) noting the scholastic and cognitive development over six years, no cognitive advantage or disadvantage was found after grades four, five, and six. (It should be noted that on a test of cognitive flexibility in grades three, five and six, the pilot experimental class performed significantly better than controls. However, the authors do not relate this superior performance to their bilingual schooling because they did not obtain similar results for follow-up classes.)

Conclusion, Summary and Implications

In conclusion, refocusing on the previous theoretical discussion, Bruner's theory is bolstered by those studies showing bilingualism to have an enhancing effect on intelligence. In addition, there are many studies that go beyond a mere correlation with intelligence and examine the qualitative cognitive development of the

bilingual child. The bulk of these studies indicates a differential cognitive development in the bilingual child as opposed to the monolingual. These studies were discussed in terms viewing the bilingual child as having greater cognitive flexibility, superior ability to form abstract concepts, a more highly diversified cognitive structure, greater divergent thinking ability, a less concrete attachment to words, an earlier object constancy development and less egocentricity. These findings were anticipated according to Bruner's theory, and therefore lend support to it.

The final remaining question is if the Piagetian can explain the recent evidence indicating bilingual children have numerous cognitive advantages? There are some indications, as noted, that these advantages do not exist, and the bilingual's and monolingual's cognitive development do not significantly vary. However, the bulk of the evidence presented in the latter part of this paper, seriously questions Piaget's view on language and cognition.

The final point which I wish to address to the reader is a practical implication of the theoretical promise presented in this paper. Since the nature of this paper was theoretical, I will not delve into this point in-depth, however, its importance bears mentioning. If the reader has at the very least accepted a possibility that bilingualism engenders an advanced cognition, then we are compelled to reexamine our educational approach in the American school system. Second language learning, in general, and bilingualism, in particular, appear to be approached from an extreme xenophobic and national chauvinistic vantage with the major educational thrust being to assimilate and accelerate English language learning and to simultaneously ignore and de-emphasize non-English language learning. The implication of this paper is that we are clearly charting an educationally counterproductive course in terms of the cognitive development of children. If bilingualism enhances cognition, then we should be promoting bilingualism and second language learning from the inception of a child's educational career.

References

Bruck, M.; Lambert, W. E.; Tucker, G. R. "Cognitive and Attitudinal Consequences of Bilingual Schooling: The St. Lambert Project through Grade Six." McGill University, 1973, unpublished paper.

Bruner, J. S. "The Course of Cognitive Growth." *American Psychologist*, 19, 1964, 1-15.

Bruner, J. S. *Toward a Theory of Instruction*. Cambridge, Mass.: Harvard University Press, 1966.

Burling, R. "Language Development of a Garo and English-speaking Child" In A. Bar-Adon, & W. F. Leopold (Eds), *Child Language: A Book of Readings*. Englewood Cliffs, N. J.: Prentice-Hall, 1971, 170-185.

Feldman, C. & Shen, M. "Some Language-Related Cognitive Advantages of Bilingual Five Year Olds." University of Chicago, unpublished paper.

Furth, H. *Thinking without Language*. New York: The Free Press, 1966.

Guilford, J. P. "Three Faces of Intellect." *American Psychologist*, 14, 1959, 469-479.

Guilford, J. P. *The Nature of Human Intelligence*. New York: McGraw-Hill, 1967.

Ianco-Worrall, A. D. "Bilingualism and Cognitive Development." *Child Development*, 43, 1972, 1390-1400.

Inhelder, B. & Piaget, J. *The Early Growth of Logic in the Child*. New York: Harper & Row, 1964.

Kirschenbaum, N. "The Effects of Bilingualism on Intelligence: A Critical Review." In S. Seidner (Ed.), *Issues of Language Assessment; Foundations and Research*. Springfield, Ill.: Illinois State Board of Education, 1982.

Leopold, W. F. *Speech Development of a Bilingual Child: A Linguistic Record*. Chicago: Northwestern University, 1939. Vol. 1.

Leopold, W. F. "Patterning in Children's Language Learning." In A. Bar-Adon & W. F. Leopold (Eds), *Child Language: A Book of Readings*. Englewood, N. J.: Prentice-Hall, 1971, p. 134-141.

Peal, E. & Lambert, W. E. "The Relation of Bilingualism to Intelligence." *Psychological Monographs*, 16, 1962, 1-23.

Piaget, J. "Language and Thought from the Genetic Point of View." In D. Elkind (Ed.) *Six Psychological Studies*. Translated by A. Tenzer. New York: Random House, 1960.

Scott, S. "The Relation of Divergent Thinking to Bilingualism: Cause and Effect." 1973, unpublished paper.

Seidner, Stanley S. *Ethnicity, Language and Power from a Psycholinguistic Perspective*. (Reprint) Bruxelles: Centre de recherche sur le plurilinguisme, 1982.

Wallach, A. & Kogan, N. *Modes of Thinking in Young Children*. New York: Holt Rinehart Winston, 1965.

Part II
Assessment Approaches

THE USE AND MISUSE OF INSTRUMENTS

Dr. Susan Durón
Illinois State Board of Education

Introduction

Rotberg (1982) states that there are 3.6 million school-aged children in the United States with a limited ability in English. It is reasonable to assume that within any given population there exists a normal distribution of both giftedness and handicapping situations. Assessment for giftedness, general achievement, behavior disorders, and language or learning handicaps is complicated when students come from homes in which a language other than English is spoken. Achieving an adequate assessment of bilingual students has been and continues to be an area of concern for educators. The limitations of existing instrumentation have been discussed in terms of cultural bias (James, 1981; Bernal, 1975), norming or standardization procedures (Hilliard, 1981; Pletcher, 1978) and validity (Law 1977; SWRL, 1980). Because the topic of this paper deals with the use and misuse of instruments, those particular issues of concern to educators and practitioners will be addressed.

This paper will discuss two areas: A) the purpose of student assessment in terms of the referral question based on the various kinds of testing including screening, identification, student/program evaluation, and exit or reclassification, and B) problems that exist resulting in test misuse. For the purposes of this paper, language proficiency testing will be emphasized because it is within the framework of language testing that most programmatic decisions are made.

Purpose of Student Assessment

Important to consider in any discussion of testing is the function of tests. Unless the end result of testing is to improve the quality of educational service, the overall value is questionable. Aran-Mendez (1975) states that "the most important function of the test is to obtain all possible information about the student so his/her learning could be guided and his/her growth promoted" (p. 86).

Educators often ask experts to identify which is the "best" test. There exists a fallacy regarding the inherent goodness or badness of a test. A test's goodness or badness can only be considered in relation to two overlapping domains: logical and ethical. The logical domain deals with test appropriateness. A "good" test is one that is considered appropriate as it relates to the assessment need juxtaposed with the referral question. Invariably, students are referred for testing without an adequate problem statement accompanying the referral. It is essential for the evaluator to ask the individual making the referral to articulate the variables that led to the request that an assessment be conducted. Such questions to ask the teacher are: Was there a question regarding language interference? Would testing facilitate program entry or exit decision making? Should the relationship be explored between the child's first and second language proficiency? Is testing necessary to aid in curriculum planning? These and other questions are logical extensions rising out of a clearly articulated need. There are literally hundreds of good tests that are available for use. Any test, however, is only as good as the information that it yields in response to a pre-specified referral question.

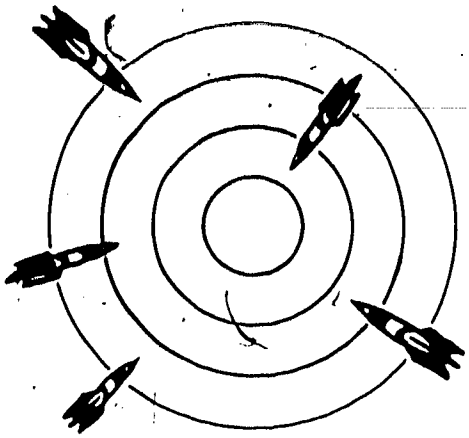
The second domain to consider regarding test goodness or badness deals with the technical aspects of tests. A deliberate simplification of these aspects follows in order to demystify certain concepts surrounding psychometrics. Figure 1 addresses test reliability and validity. A test is valid if it measures what it claims to measure. This concept is of particular interest to bilingual educators because of the potential hazard surrounding tests that are constructed specific to one single reference group. For example, a test of auditory phonemic discrimination in English would not be valid for students who speak little or no English.

A test may be considered reliable if the results are consistent, that is, if the same student under similar testing conditions receives the same score repeatedly. For example, a child who is a native English speaker should repeatedly receive a high score on an English language proficiency test. Figure 1 illustrates how a test may be reliable but not valid, valid but not reliable, and both valid and reliable. These are but a few of the technical aspects of test construction. For further discussion, see TenBrink (1974), Borg (1974), Glass and Stanley (1970). The necessity of the demystification of tests is paramount. Too often teachers ignore their intuitive judgment regarding student performance in deference to test results. Tests should be considered tools which, in the hands of a good craftsman, can be helpful in achieving the desired results.

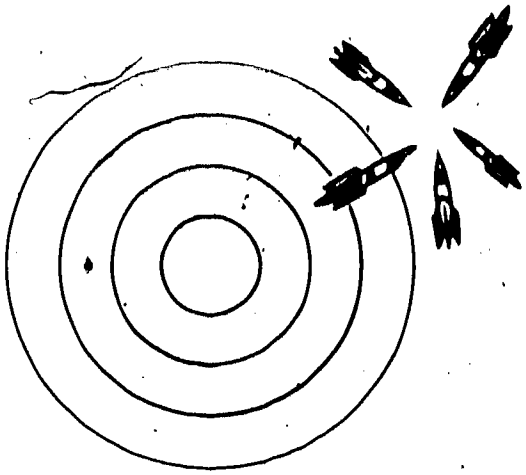
There are a variety of uses for testing which can be grouped into two interrelated categories: Student Assessment and Program Evaluation. One of the types of student assessment is identification and screening instruments targeting students for possible inclusion into bilingual programs. Language dominance tests, home language questionnaires, teacher observation surveys, etc., are useful for identification and screening. The key factor in this type of assessment is the focus on those variables which can determine program eligibility as quickly as possible.

The next type of student assessment is diagnostic testing whose purpose is for curriculum placement, program placement or prescriptive teaching. Diagnostic tests are instruments which yield analytical information

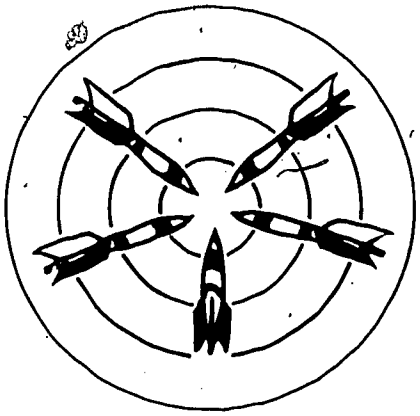
Figure 1



VALID BUT NOT RELIABLE



RELIABLE BUT NOT VALID



BOTH VALID *AND* RELIABLE

often identifying students' strengths and weaknesses. "The emphasis of diagnostic tests is on gathering specific information on what the child does or does not know and why, not just how many answers the child has correct" (Duron & Kerins 1982, p.49). Diagnostic tests can be language proficiency, norm- or criterion-referenced tests.

A final kind of student assessment instrument is used for exit/reclassification purposes. This type of test is also diagnostic in scope allowing evaluators to determine which students previously identified as being limited-English proficient would be able to successfully compete in an all-English classroom. Language proficiency and achievement tests are most often used as exit/reclassification tools to assess the listening, speaking, reading and writing skills of students enrolled in bilingual programs.

Program evaluation occurs in a response to a need to demonstrate program efficacy to parents and community members, local and state decision makers and those individuals interested in bilingual education. Often program evaluation, particularly in light of Title VII, is only designed to meet the reporting requirements set forth by the federal government. Program evaluation should encompass a variety of data sources including interviews with bilingual program and general staff; questionnaires to students, staff and parents; review of student test results, grades, evaluations; and other information such as proposals, class list, meeting minutes, etc.

Law (1977) in his article entitled "Evaluating Bilingual Programs" states that the nature of bilingual/bicultural education goes against evaluation using a vigorous experimental control. He goes on to say that the

measurement problems are acute because the population is unique linguistically and culturally. The fact that all students who are eligible for program participation should be served denies an experimental design with sampling and control groups. Considering that fact alone, it is no wonder that there is a paucity of experimental research in the field of bilingual education.

Problems and Misuse of Instruments

The problems and misuse of instruments will be divided into two areas: those concerns related to all formal testing situations and those concerns idiosyncratic to tests designed for a bilingual population. A discussion of the most common abuses in the administration of standardized tests is found in Duron and Kerins (1982). These abuses include nonstandardized administration, ignoring time limitations, incorrect scoring, and misinterpretation of the standard error of measurement.

Standardized test administration is necessary in order to make comparisons between students' test results and those of the norm group. Each test provides a technical manual containing standard instructions which need to be carefully followed to insure comparability among scores. An example of nonstandard test administration is when a bilingual teacher translates items from a standardized vocabulary test into the child's home language. Another error, ignoring time limitations, can greatly influence test results. Norms or other performance standards are invalid when time sequence has been modified. An excellent test can be rendered useless if it is scored incorrectly. Simple things such as careless arithmetic errors or incorrectly reading of the norm table can drastically change the total score of the test. Scores should be checked and rechecked to avoid these types of errors.

The final problem of standardized tests is related to the standard error of measurement. This refers to "an estimate of the range of scores that would be found for a specified person if that person were to tested again and again on the same test" (Duron & Kerins, 1982). The "confidence level" of the score depends upon the test reliability and standard deviation. Scores on tests with low reliability and a high standard deviation would be questionable and should be substantiated through further testing or other non-test data.

The problems and misuse of instruments designed for bilingual populations are acute. In the final report on the "Development of Entry/Exit Criteria and Associated Procedures for Bilingual Education" (1981), the authors are critical in their discussion of language proficiency instruments. Reporting on the discouraging experiences of teachers and other test users, they state that, generally, users felt that teacher judgment was more likely to be a valid measure of both language proficiency and capability of succeeding in an all-English-medium classroom than any test they had been using" (pp. 2-10).

A problem exists in that most language proficiency tests deal with oral language only. While linguists disagree on how language should be assessed, it is generally agreed that the instrumentation should consist of various components including receptive and expressive language. There is a need for global instruments that address the child's listening, speaking, reading and writing skills. Most language proficiency tests currently on the market are designed to measure receptive language — there are few that are diagnostic in scope and even fewer diagnostic/prescriptive tests that would be useful to classroom teachers in modifying their instructional strategies.

Because the field of bilingual education is relatively new, many instruments were poorly designed or constructed in order to capitalize on an emerging market. Normative data are lacking or poorly documented, and often test reliability is not established. These factors considered, it is still encouraging to note that great strides have been made in the area of language assessment as awareness and psychometric sophistication are continually augmented. As more and more language groups are represented, however, additional tests need to be developed to assess students' native language skills and to compare their relative proficiency between languages.

The reality that tests are not perfect measures of student achievement must frame any discussion of their use and misuse. As such, tests can be used constructively if they are considered as aides or tools to be interpreted carefully. The end results of a good testing program will be within reach if the following occurs:

(1) Testing needs and purposes are specified.
(2) School personnel are sensitized to interpret test results.
(3) Local norms are developed.
(4) Tests are used diagnostically.
(5) The reporting plan is specified in advance.

The importance of specification of testing needs and purposes was articulated in the first part of this paper. As stated, formulation of the referral question can be facilitated if the evaluator and school personnel work together. At this time, problems and shortcomings surrounding tests can be mutually considered. Test scores too often become "magic numbers" to teachers who do not look beyond the score and fail to interpret the results.

Essential to a good testing program is development of local norms. The following section on developing local norms is reprinted from the *Handbook on Special Education* (1982) with permission of the Illinois State Board of Education.

How to Establish Local Norms

The first step in establishing local norms is to identify the data base which will be used in the process. The data base might be: (1) a set of scores of all target students attained during the current academic year; (2) scores of students participating in the program during the past years; or (3) scores of students currently participating in the program, as well as those of students who will participate in the program in the future. In the latter case, the test is actually re-normed each year by including additional data. Changes in the norms might be moderate, with no more than five years of data collection needed to established state norms.

A decision about which of the three methods to use depends on the particular situation of the local district. A limited number of students participating in the program during the current year may weaken the validity of the norms used through the first method. The second method may be used only if the same instrument has been used in the past. Also, changes in curriculum or redistricting might not make this option feasible. The third method requires more time and work than the others.

Once the data base has been identified, raw scores are converted into metrics useful and neccessary for making comparisons among the scores. Percentiles, stanines or grade equivalents can be established with data from each age or grade level.

Percentiles

Percentiles may be used to compare the performance of individual students with the performance of other individuals or a group of students. Use of percentiles helps answer the often-asked question: How well is this student doing in comparsion with other students? Performance of any group of students may be also be compared to the norms by calculating the median (the raw score at the 50th percentile) of the group and determining the relationship between the raw score and the score correspondence to the 50th percentile on the already established norms.

Care must be taken when using percentile ranks because a scale of equal intervals is not obtained. For example, a percentile rank of 50 cannot be assumed to be twice as good as a percentile rank of 25 or the difference between percentiles of 20 and 10 (20-10 = 10) cannot be assumed to be equal to a difference between percentiles of 50 and 40 (50-40 = 10). Therefore, percentile ranks cannot be used to compute means, standard deviations, correlation coefficients or other statistics for which equal interval scales are required.

Stanines

After attaining percentile ranks for each score, it is possible to convert percentiles to stanines which are standard scores consisting of nine possible values from one (low) to nine (high). For each of the nine possible stanine scores, the percentage of scores is distributed as follows:

	Percentage of Scores
1	4
2	7
3	12
4	17
5	20
6	17
7	12
8	7
9	4
	100%

Therefore, the lowest 4% of the scores are assigned to stanine of 1; the next 7% of the scores are assigned to stanine 2, etc. While stanines provide a relatively simple way of explaining test results to parents or individuals not sophisticated about testing, stanine values do not allow for precise interpretation of a particular score.

Grade Equivalents

Grade-equivalent scores represent an estimated average (median) performance level achieved by a sample of students if they were tested each month of the school year. For purposes of computing grade equivalents, the school year is divided into 10 months. Grade equivalents are expressed in terms of a year and a decimal fraction of a year. For example, if a median raw score of a sample of third graders in the 5th month of the school year (grade 3.5) is 40, then the grade equivalent for a raw score of 40 is 3.5 (3rd grade, 5th month).

Grade equivalents should be interpreted with considerable caution. Grade equivalents are useful as rough indicators of the [illegible] performance of students at particular grade levels. For example, a grade-equivalent score of 8.6 obtained by a sixth grade student on a math test does not necessarily mean that this student knows math at the 8th or 9th grade level. The 8.6 grade equivalent means the sixth grade student performed as well on the test as a "typical" student in grade 8.6 would perform on the same test.

Another problem in interpreting grade-equivalent scores is that an equal interval scale is not obtained. For example, the difference between grade-equivalent scores of 8.6 and 7.6 is not equal to the difference between grade-equivalent scores of 3.6 and 2.6. For this reason, grade-equivalent scores cannot be averaged. It should be emphasized that interpretation of test scores using grade equivalents is useful if the user remembers that a grade-equivalent score simply indicates that a student performs as well as the typical student at the particular grade level indicated by the grade-eqivalent score.

Diagnostic Use of Tests

A good testing program should include the diagnostic use of tests. For example, certain language proficiency tests contain various subtests that can be compared to analyze the child's strengths and weaknesses. This information can be useful in structuring instruction based on pre-established objectives. In addition, if a language assessment instrument is given in both languages, the child's relative language proficiency can be estimated so that decisions are facilitated regarding the language of instruction. Finally, test-item analysis on tests like the *Boehm Test of Basic Concepts* can address specific conceptual deficiencies in the child. Categories or clusters of items can be identified in terms of strengths and weaknesses for each child.

The evaluator must ask him/herself questions regarding the diagnostic use of tests. These questions deal with 1) pattern interaction variables (Is there a pattern to the student's responses? Is there evidence that the errors occurring are particular to the home language? To English? To both languages?); 2) student comfort-level variables (Is this the level of performance that would be expected from this child? If not, has rapport been established appropriately?); and 3) linguistic/cultural variables (Do the items that are missed reflect linguistic bias? Cultural bias?).

The final consideration of a good testing program to be discussed in this paper is the reporting plan. Important to consider is how test results are going to be disseminated and to whom. Audiences such as teachers, parents, the school board, etc., are possible recipients. The kinds of reports that will result from data aggregation should be planned and decided upon. Details such as the type of graphics and narratives should be considered. Figure 2 represents a form that can be adapted for local test coordinators, teachers or evaluators.

In conclusion, knowledge of the use and misuse of tests is something that enlightened educators must address. Familiarity with sources of measurement bias, which exists in the testing situation itself, is necessary. Bernal (1975) contends that lower test scores for bilingual children result from those biases mentioned in this paper. He recommends the following procedures for improving test results:

1. Organizing students into small groups for testing and warmup;
2. Matching the examiner with the student for ethnicity;
3. Having the examiner spend a few minutes with the student speaking the home language;
4. Preparing similar items for the student to practice, giving him/her an opportunity to articulate why they answered in certain ways and giving them feedback on these items (p. 51).

Bernal's procedures and other testing considerations discussed in this paper should not be thought of as stacking the deck in favor of bilingual students. Conversely, these procedures should take educators one step closer to describing the child's realistic achievement level, often wrongly reflected by inadequate or misinterpreted test scores. Keeping this in mind, the assessment instruments can be useful tools in the hands of a thoughtful, astute evaluator.

Figure 2

Purpose of Testing	
Client or Student	
Test Date(s); Time(s)	
Test Criteria	
Test Dissemination Plan	

Bibliography

Aran-Mendez, Olga. *With Bias Toward None: A National Planning Conference on Nondiscriminatory Assessment for Handicapped Children.* Lexington, Kentucky: University of Kentucky Press, 1975. pp. 86-88.

Bernal, Ernest J. "Issues Related to the Assessment of Chicano Children." In, Proceedings from *With Bias Toward None: A National Planning Conference on Nondiscriminatory Assessment for Handicapped Children.* Lexington, Kentucky: University of Kentucky Press, 1975. pp. 5-6.

Borg, Walter R.; Gall, Meredith; and Bell, Norman T. *Educational Research.* New York: David McKay Company, Inc., 1974.

Duron, Susan and Kerins, Thomas. "The Use of Test and Non-Test Data in Evaluation." *Handbook for Evaluation and Special Education Effectiveness.* Springfield, Illinois: Illinois State Board of Education, 1982. pp. 48-54.

Glass, Gene V. and Stanley, Julian C. *Statistical Methods in Educational Psychology.* Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1970.

Hilliard, Asa G. III. "Standardized Testing and Non-Standard Populations." *The Generator of AERA,* Vol. 11, No. 2, spring, 1981. pp. 13-27.

James, Thomas; Navarro, Richard; Stone, Nancy, Editors. *Bilingual Education for Hispanics: Issues of Language Access and Equity.* Institute for Research on Educational Finance and Governance: Policy Notes. California: Stanford University Press. Vol. 2, No. 4, autumn, 1981. pp. 1-2.

Law, Alexander. *Evaluating Bilingual Programs.* TM Report 61. Trenton, NJ: ERIC Clearinghouse on Tests, Measurement, and Evaluation, April, 1977.

Pletcher, Barbara P. *A Guide to Assessment Instruments.* New York: Santillana Press, 1978.

Rotberg, Iris C. *Health Education Review.* Vol. 52, No. 2, May, 1982. pp. 149-68.

Southwest Educational Regional Laboratories. *Designers Manual: Resources for Developing a Student Placement System for Bilingual Programs.* Austin, Texas: Southwest Regional Laboratories for Research and Development in Education Press, 1980. pp. 2-10.

TenBrink, Terry D. *Evaluation: A Practical Guide for Teachers.* New York: McGraw-Hill Book Company, 1974.

HOW CAN SECOND-LANGUAGE ASSESSMENT CAPTURE COMMUNICATIVE COMPETENCE?

Margo H. Gottlieb
University of Illinois at Chicago Circle
School District U-46

Communicative competence is a multidimensional construct with broad social, cultural, and linguistic parameters. Hymes (1971) first recognized the effect of context on cognition, noting that specific social situations dictate the selection of grammatical structures in language production. Within the last decade, other second language theorists have expanded this concept to include the knowledge of the psychological and social customs of one's community (Weininger, 1978), expression of interlanguage (Selinker, 1974) or approximative systems (Nemser, 1974), and pragmatics (Chomsky, 1972; Oller, 1978).

If oral language proficiency is a manifestation of linguistic competence, then functional language use reflects communicative competence. Whereas linguistic competence is equated with the mastery of the phonological, morphological, and syntactic systems of a language, communicative competence assigns priority to semantics. The emphasis on meaning is not restricted, however, to the relationship of lexemes within an utterance; it encompasses all the supralinguistic features of discourse. Therefore, communicative competence can be considered an added dimension of linguistic competence; not an autonomous and distinct entity, but an integrated and homogenized expression of language in cultural contact. The adaptation of linguistic competence to the total informational input of the situation is the essence of communicative competence (Savignon, 1972). The critical adjunct of context involves all the macroenvironmental factors that affect the communication act (Burt, Dulay, and Krashen, 1982).

Second-language assessment of communicative competence should ideally reflect: 1) a theoretic framework, 2) a functional language system, and 3) a model of bilingual education (Erickson, 1981). The field of linguistics impacts second-language methodology which, in turn, influences test construction (Upshur, 1973). The linear trend, however, moves in a step-like progression. Although Carroll proposed the idea of integrative tests as early as 1952, second-language instructors, for the most part, have remained faithful to structuralistic teaching and testing techniques. Even in today's English as a Second Language (ESL) and Bilingual Education instructional settings, discrete-point tests are the most prominent (Day, 1981). If the development of communication skills is indeed a programmatic goal of ESL and Bilingual Education, then language instruments must be constructed that can validate this claim.

The integrative testing movement is directed towards this end. It is unified in its holistic view towards language and its attempt to evaluate the use of natural language in social/cultural interaction. It is divergent in its methods of measurement. Three methodological orientations have emerged within the last decade:

1. Direct — the replication or use of a naturalistic setting as a backdrop to language interaction as seen in the Ilyin Oral Interview (1976), the Savignon Communicative Competence Tests (1972), and the Foreign Service Institute Oral Interview.

2. Indirect — the use of simulated or quasi-realistic activities to promote language production as illustrated by the cloze procedure (Taylor, 1953) and the noise test (Spolsky, 1968).

3. Mixed — the use of spontaneous language sampling techniques coupled with a discrete-point scoring system as evidenced in the Basic Inventory of Natural Language (Herbert, 1977) and the Bilingual Syntax Measure (Burt, Dulay, and Hernandez-Chavez, 1976). McCollum and Day (1981) refer to this category as quasi-integrative.

Noa, Russell, and Silverman (1976) depict three domains of language assessment: the social, the linguistic, and the communicative. Each facet of a representational cube is quartered so that communication skills reflect listening, speaking, reading, and writing; linguistic structures denote phonology, vocabulary, syntax; and semantics; and social settings represent the home, school, peers, and community. An ideal model of communicative competence should embrace all these defined areas; an ideal language assessment instrument should explain the model.

The global nature of communicative competence makes language testing very cumbersome. No one measure can effectively capture all the intricacies of language nor should it try. Absent is a filtering mechanism that retains those elements critical to successful language learning. Language assessment should be based on specific, predesignated needs of the population. Communicative competence testing should demonstrate a student's ability to use language in a variety of commonplace settings (Morrow, 1979). The remaining section of this paper will be devoted to how an Illinois school district has faced this issue.

The Measure of English Communicative and Conceptual Achievement (MECCA) was developed to provide ESL and Bilingual Education administrators and

practitioners with reliable data for instructional decision making. It is a comprehensive assessment system designed for the K-12 Limited-English-Proficient (LEP) student. Communicative competence is one of four sections; the composite score yields a profile of English competencies.

For the sake of practicality and feasibility, the perspective of communicative competence had to be narrowed. Considering the social factors affecting second-language acquisition, it was decided that input from the home, peers, and community could be obtained through parent and student interviews. Thus, the sociolinguistic variables related to language usage have come to be identified under three major headings: student information, parent information, and interaction of student with others. A code sheet was devised to systematize the process of data collection.

The focus for assessing communicative competence thereby became the school environment. More specifically, it centered on collateral learning (Dewey, 1938) or the hidden curriculum (Jackson, 1968), those aspects of school life that are essential for survival within the institution, but are outside the realm of academic learning. This covert curriculum includes the rules, norms, and values that are internalized by the student in the culture-bound atmosphere of school (Apple, 1979).

Applying this notion to language learning, the hidden curriculum can be equated with the context-embedded situations (Cummins, 1981a, 1981b) with which a student must cope within the school milieu. Thus, for the native English speaker who has had no experience outside an American classroom, communicative competence would be cognitively undemanding; for the LEP student, whose language and culture are not consonant with that of the school's functional language use in English (L_2) would be cognitively demanding (Cummins, 1981a, 1981b).

To accomplish the goal of measuring communicative competence, a list was generated of all the situational contexts a student encounters during the course of a typical school day. Each idea was categorized as an example of school procedures, classroom practices, or interaction with school personnel. Objectives were then formulated for a task-based performance test of listening and speaking comprehension (see appendix).

As almost all language tests seem to consist of randomly selected items of nonspecified content (Farhady, 1982), planning for content validity was deliberately incorporated into every phase. Preparation of a specification table was necessary to ensure the proportionate representation of product and process objectives (Thorndike and Hagen, 1977). A 2X3 matrix was created whereby the columns represented receptive and expressive language, respectively, and the rows, the contextual categories. For each cell, six items were constructed: two for the primary elementary grades, two common to both the elementary and secondary levels, and two unique to the junior high and high school.

A pool of receptive and expressive language questions was formed, revised, edited, and checked against the original objectives. All items were then rank ordered in a logical sequence from the most simplistic to the most linguistically complex. Illustrations for the receptive subsection were incorporated as needed.

Second-language-acquisition research has confirmed that a silent period devoted to one-way communication precedes responsive oral language production (Burt, Dulay, and Krashen, 1982). In defining a developmental sequence for communicative competence, it is necessary, therefore, to proceed from receptive to expressive behavior. In the MECCA, for example, for receptive competence, a student is required to identify the illustration that shows "What do you use to write with?" For the expressive counterpart, he/she is requested to respond to this situation: "It's time to take a math test. You take out your pencil and it has no point. What would you ask your teacher?"

Learner verification was conducted with both LEP's and native English speakers. The Educational Products Information Exchange (EPIE) Institute's format for the analysis of instructional materials was adapted to assessment instruments. After taping and transcribing the session, the data was transposed onto this form. Based on the information gained during this phase, test items and illustrations were clarified.

To facilitate the ease of scoring and to guide interpretation, a multiple-choice format was adopted as the scoring method. A four-point scale was developed that would accurately reflect a student's competence for the given situation:

1. no response,
2. inappropriate or incorrect response,
3. partially acceptable or incomplete response,
4. appropriate and meaningful response.

As a varying range of potentially plausible responses was acceptable, students could draw from their personal linguistic and experiential repertoires in answering the questions. The problem posed in this item, for example, lends itself to a variety of possible alternative solutions.

> "Last week you didn't go to school. When you come back, your teacher wants to know your excuse. Why were you absent?"

The issue raised for discussion — How can second language assessment capture communicative competence? — can now be approached more definitively. Applying Erickson's (1981) criteria to the communicative competence section of the MECCA, it can be stated that:

1. Psycholinguistic and sociolinguistic theory underlie the development of the instrument.

2. The hidden agenda of the schools offers a rich array of interpersonal communication context for assessing functional language usage.

3. Modification of the bilingual curriculum and the establishment of entry/exit criteria for the bilingual program will evolve from the pilot and field-test findings of the composite system.

Provision for social and cultural adaptability is as necessary to the LEP student as the promotion of academic achievement. Communicative competence demonstrates the ability of a student to respond to a wide range of linguistically diverse situations. However, it is merely one aspect of the complex language development process.

Within the school domain, second-language assessment needs to address total language usage-communicative proficiency in conjunction with conceptual achievement. Communicative competence alone cannot adequately determine a student's academic performance in a second-language classroom. Comprehensive evaluation measures are requisite for the provision of appropriate and meaningful educational experiences for all LEP students.

Appendix

Language Contexts in the School Domain: Communicative Objectives for a Task-Based Performance Test of Listening and Speaking Comprehension

I. School procedures

1. evacuate the building during a fire drill
2. require an excused absence
3. ask for early dismissal
4. explain tardiness
5. relate playground (campus) rules
6. line up for entrance/exit
7. respond to schedule of in-service days and vacations
*8. open your locker
*9. use the telephone
*10. describe your schedule
*11. argue against study hall in favor of an open campus

II. Classroom practices

1. follow oral directions (simple to multiple)
2. sharpen a pencil
3. request supplies (paper, scissors, crayons, ruler)
4. ask for a pass to the washroom or for a drink of water
5. apologize to a teacher or peer
6. use polite phrases and greetings
7. paraphrase the principal's directive over the intercom
8. explain the reason for no homework
9. recall class behavior code
10. argue with a peer
11. salute the flag
*12. explain 'cutting class'

III. Interaction with school personnel

1. nurse— explain what's wrong (injury or illness)
*2. counselor— react to a job interview
3. librarian or resource teacher— tell about a lost book
4. office secretary— explain a fight or incident
5. lunchroom attendant— pay for a meal or obtain a ticket
6. custodian— relay a message from the teacher
7. bus driver— give your new address or the route to a friend's
8. principal— supply requested information
*9. dean— respond to a suspension for smoking
*10. P.E. teacher— react to taking a shower

References

Apple, M. W. *Ideology and Curriculum*. London: Routledge and Paul Kegan, 1979.

Burt, M.; Dulay, H. & Krashen, S. *Language Two*. New York: Oxford University Press, 1982.

Carroll, J. B. "An Appraisal of Language Tests from the Standpoint of the Psychology of Language." In *Yearbook of the National Council on Measurement Used in Education*, 1952.

Chomsky, N. *Language and Mind*. Cambridge: MIT Press. 1972.

Cummins, J. "The Role of Primary Language Development in Promoting Educational Success for Language Minority Students." In *Schooling and Language Minority Students: A Theoretic Framework*. Los Angeles: Evaluation, Dissemination and Assessment Center, California State University, 1981.

Cummins, J. "Four Misconceptions about Language Proficiency in Bilingual Education." *NABE Journal*, 1981, *5* (3), 31-45.

Day, E. C. "Assessing Communicative Competence." In J. G. Erickson and D. R. Omark (Eds.) *Communication Assessment of the Bilingual Bicultural Child*. Baltimore: University Park Press, 1981.

* indicates an item unique to junior high and high school

Dewey, J. *Experience and Education.* New York: Macmillan, 1938.

Erickson, J.G. & Omark, D.R., (Eds.) *Communication Assessment of the Bilingual Bicultural Child.* Baltimore: University Park Press, 1981.

Farhady, H. "Measures of Language Proficiency from the Learner's Perspective. *TESOL Quarterly,* 1982, *16* (1), 43-59.

Hymes, D. *Language Acquisition: Models and Methods.* New York: Academic Press, 1971.

Jackson, P. W. *Life in Classrooms.* New York: Holt, Rinehart, and Winston, 1968.

McCollum, P. A. & Day, E. C. "Discrete Point Scoring of Expressive Language Samples." In J. G. Erickson & D.R. Omark (Eds.) *Communication Assessment of the Bilingual Bicultural Child.* Baltimore: University Park Press, 1981.

Morrow, K. "Communicative Language Testing: Revolution or Evolution?" In C. J. Brumfit & K. Johnson (Eds.) *The Communicative Approach to Language Teaching.* Oxford, England: Oxford University Press, 1979.

Nemser, W. "Approximative Systems of Foreign Language Learners." In J. C. Richards (Ed.) *Error Analysis: Perspectives on Second Language Acquisition.* London: Longman Group Limited, 1974.

Noa, J. K.; Russell, R. H. & Silverman, R. J. *Oral Language Tests for Bilingual Students: An Evaluation of Language Dominance and Proficiency Instruments.* Washington, D.C.: U.S. Office of Education, 1976.

Oller, J. W. "Pragmatics and Language Testing." In B. Spolsky (Ed.) *Advances in Language Testing Series: Two Approaches to Language Testing.* Arlington, Virginia: Center for Applied Linguistics, 1978.

Savignon, S. J. *Communicative Competence: An Experiment in Foreign Language Teaching.* Philadelphia: The Center for Curriculum Development, 1972.

Selinker, L. "Interlanguage." In J. C. Richards (Ed.) *Error Analysis: Perspectives on Second Language Acquisition.* London: Longman Group Limited, 1974.

Spolsky, B. "Preliminary Studies in the Development of Techniques for Testing Language Learning." *Problems in Foreign Language Testing, Language Learning,* Special Issue No. 3, 1968, 79-101.

Taylor, W. L. "Cloze Procedure: A Tool for Measuring Readability." *Journalism Quarterly, 30,* 1953, 415-433.

Thorndike, R. L. & Hagen, E. P. *Measurement and Evaluation in Psychology and Education.* New York: John Wiley and Sons, 1977.

Upshur, J. A. "Productive Communication Testing: Progress Report." In J. W. Oller & J. C. Richards (Eds.) *Focus on the Learner: Perspectives for the Language Teacher.* Rowley, Massachusetts: Newbury House, 1973.

Weininger, J. C. G. "Communicative Strategy among Children in a Bilingual School Environment (K-3)." In J. E. Redden (Ed.) *Occasional Papers on Linguistics Number 3: Proceedings of the Second International Conference on Frontiers in Language Proficiency Dominance Testing.* Carbondale, Illinois: Department of Linguistics, Southern Illinois University, 1978.

A MODEL FOR ASSESSING COMMUNICATIVE PROFICIENCY IN A SECOND LANGUAGE

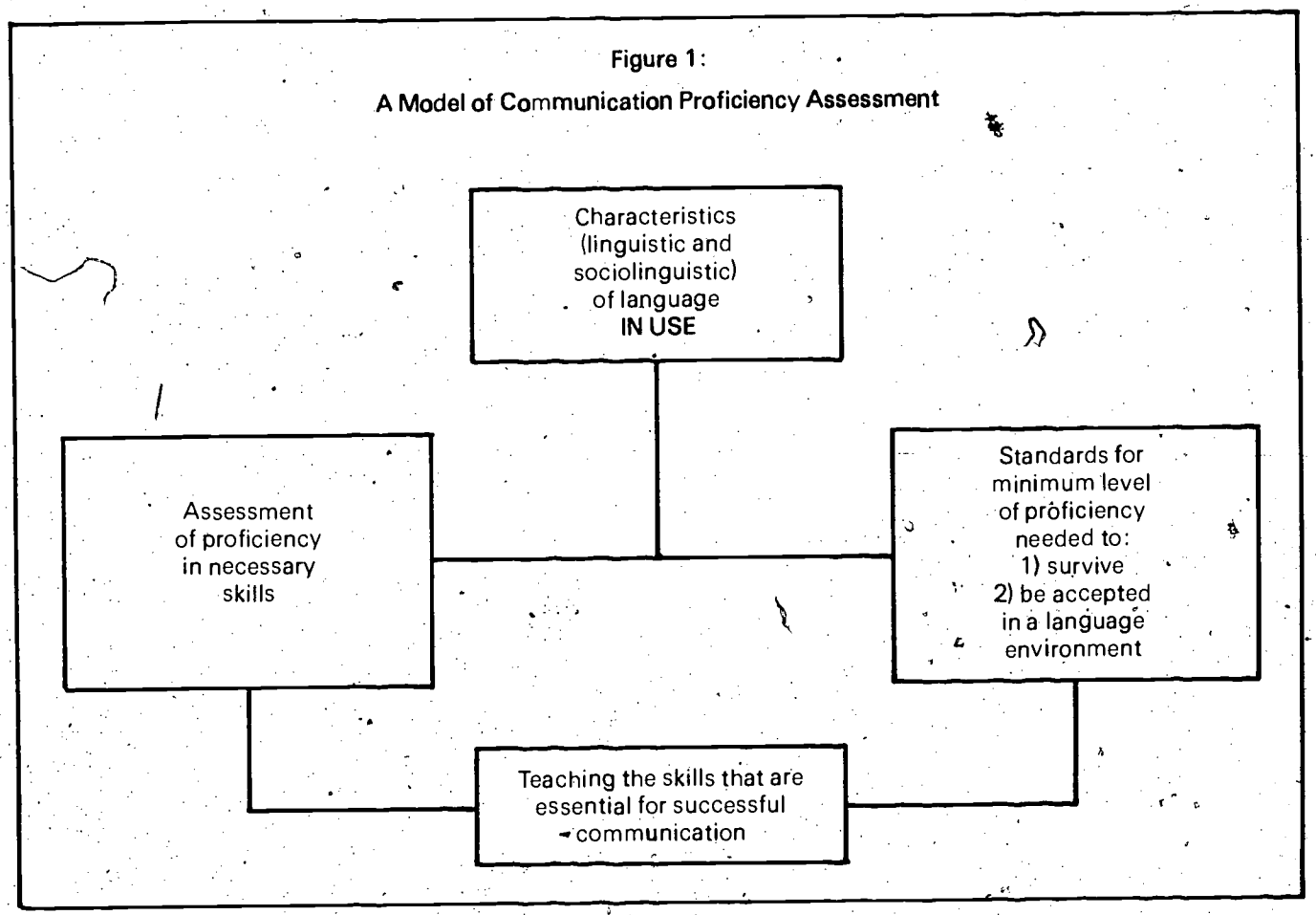
Else V. Hamayan
Bilingual Education Service Center
Arlington Heights, Illinois

The field of language assessment has witnessed much advancement in the last decade. Practitioners involved in teaching a second language are developing an increasingly accurate view of what constitutes proficiency or competence in a target language. Despite these developments, however, we lack consistency in both the approaches taken to second language assessment and in the actual tests that are available through the market. One of the reasons for this inconsistency in how second language proficiency is assessed may well be the lack of a conceptual framework that integrates a definition of language proficiency with pedagogy and assessment.

The above considerations necessitate a system of communicative proficiency assessment that takes into account the initial steps before assessment begins and the steps to be taken after assessment is completed. A system of communicative proficiency assessment should have as its basis a description of what constitutes successful communication in a given situation (see Figure 1). For example, what characterizes successful language use for a teacher in a bilingual classroom, for a student in a target-language classroom, or for an adult in a specific work situation? Questions such as these should form the basic premise of communicative proficiency assessment.

Figure 1:

A Model of Communication Proficiency Assessment



The first section of this paper is an attempt to identify characteristics of language in use (the top box in Figure 1). This will serve as the foundation on which an assessment system may be built.

CHARACTERISTICS OF LANGUAGE PROFICIENCY

To begin, we need a comprehensive description of language proficiency itself. As Brown (1980) indicates, development of a valid assessment system first requires a good definition of language proficiency and how it is acquired. The latter area has received ample attention in the field of second language research. Extensive research in first and second language learning has indicated that we learn language by a process of creative construction through which we formulate hypotheses about the language being learned and we apply and modify those hypotheses on the basis of interaction with the linguistic environment that provides the context for learning (Brown, 1980;

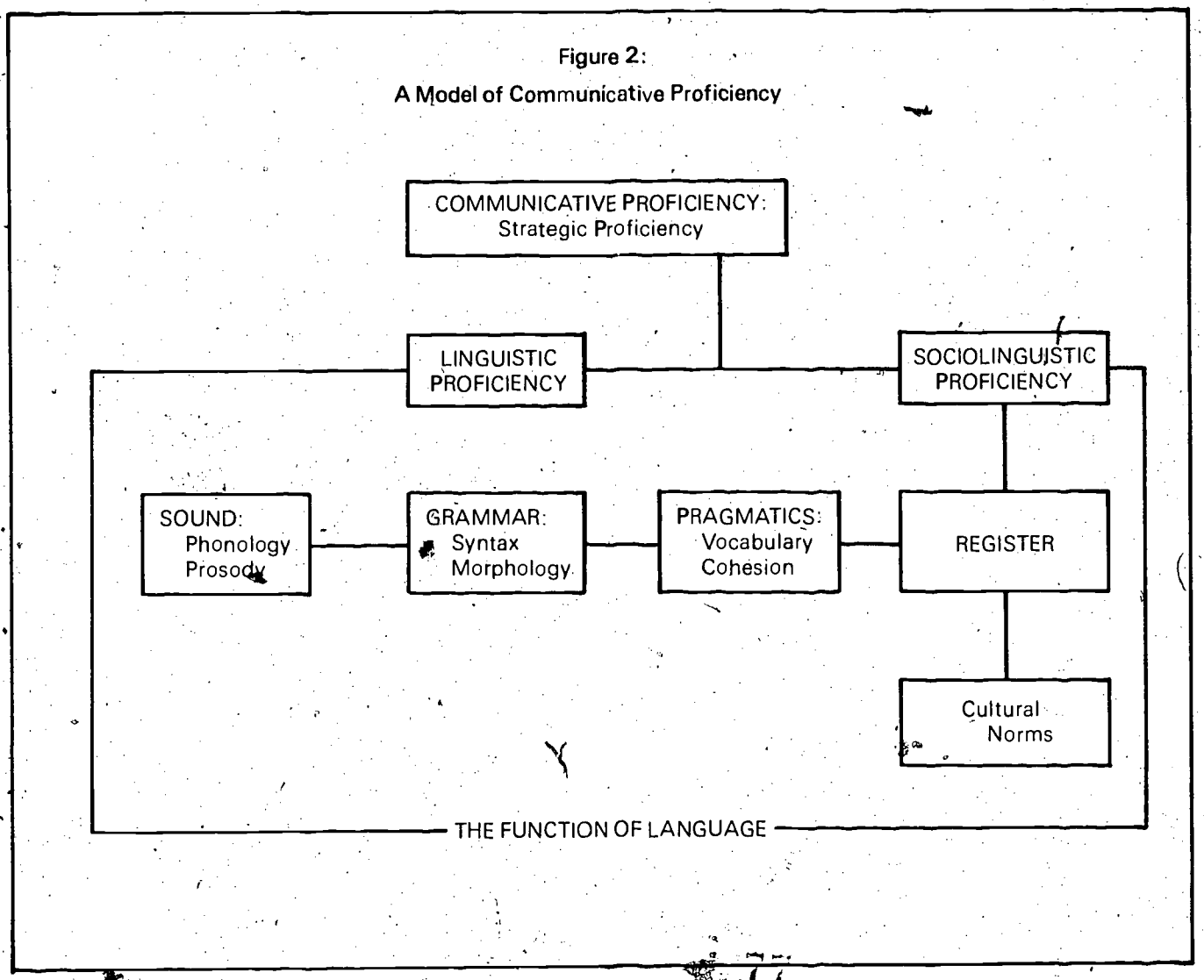
McLaughlin, 1978). By a series of hypothesis building and testing in the context of communicating in the target language, second language learners are said to pass through systematic stages, or interlanguages (Selinker, 1972), that progressively approximate fluent "native-speaker" language.

Although the second language learning process has been fairly elaborately described, it is not clear what constitutes *proficiency* in a language. Two major aspects of language proficiency have been identified and discussed in the literature. Although different labels have been used to refer to these two aspects, the basic notion of linguistic proficiency, on the one hand, and sociolinguistic proficiency, on the other, emerges. Many writers (See for example, Canale and Swain, 1980, Erickson, 1981) have referred to the two aspects as form and function. Others (Bachman and Palmer, Note 1) have referred to grammatical competence and sociolinguistic competence as two constituents of the larger concept of communicative competence. In this paper, I will adopt (with modifications) the terms used by Bachman and Palmer and refer to the general proficiency in language as communicative proficiency and to the two components as linguistic proficiency (form) and sociolinguistic proficiency (function).

The history of the development of a conception of communicative proficiency has been marked by the continuing debate between proponents of a model which emphasizes a general factor of language proficiency (see for example, Oller and Hinofotis, 1980) and those who maintain that communicative proficiency consists of a number of distinct components (such as Canale and Swain, 1980). While the dichotomy between the single-underlying factor model and the multiple-component model may exist at the theoretical level, one view need not preclude the other. The separate components of communicative proficiency may be seen as feeding into one general underlying language ability. In fact, there seems to be empirical evidence supporting both views: Oller maintains that one general factor accounts for the greatest proportion of reliable variance in language test scores, but in addition, other researchers (Carroll 1980; Bachman and Palmer, Note 1) have shown that more of the variance in language test data can be accounted for when information from specific language traits is taken into account.

Figure 2:

A Model of Communicative Proficiency



A framework of a general communicative proficiency, consisting of two components, is illustrated in Figure 2. The conceptual framework presented in this paper borrows from the Bachman and Palmer model, as well as that of Canale and Swain. In both models, it is hypothesized that communicative competence consists of several components which measure distinct underlying abilities. In an attempt to avoid any overlap between the components of communicative proficiency and in order to illustrate the relationship between those components, the framework in Figure 2 has been modified from the original presentation (Bachman and Palmer, Note 1). The term "proficiency" is used in the present paper instead of "competence" in order to avoid the possibly confusing distinction between competence and performance. Proficiency will henceforth be used to indicate mastery as it surfaces via performance on a given task. Secondly, the various components of communicative proficiency are categorized into the two major aspects, linguistic and sociolinguistic, corresponding to Erickson's form and function, respectively. This is a departure from Canale and Swain's model which includes: 1) grammatical competence (our linguistic proficiency), 2) sociolinguistic competence (our sociolinguistic proficiency), and 3) strategic competence. The last component has been excluded in the present paper as a separate component and is rather seen as part of a more general communicative ability. As Bachman and Palmer argue, strategic competence comprises an ability which affects all the components of communicative proficiency that are actually put to use. The present conceptualization of the two aspects of communicative proficiency also departs from Bachman and Palmer by classifying pragmatic ability as a component and not as one of the major traits.

Linguistic Proficiency

This aspect of communicative proficiency, which has been the focus of much of the earlier writings on language proficiency, includes mastery of the sound system, the grammar, and the pragmatics of a language.

The first component, the sound system, includes the mastery of the phonological system as well as the prosodic features of a language. The latter are important features of language that are frequently neglected. Non-fluent speech is often characterized by the use of inappropriate intonation or stress which may have a strong negative impact on communication. Bachman and Palmer exclude phonology from their model, arguing that it acts more as a channel than a component of competence. However, second language users clearly develop varying levels of proficiency in the sound system. Consequently, the effectiveness of a message — both connotatively and denotatively — would vary as a function of that level of proficiency.

The second component of linguistic proficiency is grammar, which includes mastery of the morphology as well as the syntax of a language. This component has been the focus of many tests of second language proficiency and has received ample attention in the field of language testing.

The third component is pragmatic competence which Bachman and Palmer associate with the ability to express and comprehend messages. Included in this component are mastery of vocabulary (hence, the placement of the component under the linguistic aspect) and the ability to process cohesively organized utterances. Although this set of abilities has been included in the linguistic component, it is nevertheless very closely related to the sociolinguistic component. This interrelation of linguistic components and the sociolinguistic aspects is to be emphasized and will be discussed later.

The relation among the three linguistic components also needs to be discussed. Proficiency levels in the three components — the sound system, the grammar, and pragmatics — are generally highly correlated. However, as Cziko (Note 2) has pointed out, most research in this area has used correlation coefficients between pairs of language measures; thus, individual performance is compared to a group as in a norm-referenced approach. A person's proficiency relative to meaningful "absolute" criteria has not been measured. Thus, a non-native speaker may have native-like proficiency in grammar, but may have a very low level of proficiency in the sound system of a target language. In fact, informal observation of many ethnic communities in the United States leads me to believe that the above description accurately fits a significant number of second language users. This could result in a high positive correlation between the two language measures, grammar and sound, which will not reflect in any way the (possibly significant) difference between the absolute levels of proficiency in the two measures.

Linguistic proficiency has received most of the attention in the area of language assessment. Most language tests on the market focus solely on components of linguistic proficiency, with disregard to the other, the sociolinguistic aspect of communicative proficiency. However, even within this restricted arena, the method of assessment varies dramatically. Two key issues in language assessment are whether language should be tested by discrete-point versus integrative tests and secondly, whether language should be tested directly versus indirectly (see Cohen, 1980).

Concerning the first issue, most marketed tests have traditionally taken a discrete-point approach, where language (or, in this case, linguistic proficiency) is broken down into its component parts and each one of those parts is tested separately. Thus, the ability to process subject-verb plural agreement (the sheep is jumping vs. the sheep are jumping) or the ability to discriminate between minimal pairs (ship-sheep) is tested. It has been suggested (Brown, 1980; Erickson, 1981) that the rationale for this approach is based on structuralist and behavioral psychological theories that view language as the sum of its parts. The discrete-point approach has met with criticism, particularly in light of the view that language is a synergistic phenomenon and that application of linguistic skill requires integration for its pragmatic use (Brown, 1980). In view of the latter formulations, we

need to refocus methods of assessment from the exclusive sampling of discrete-points within language to the inclusion of integration of different skills. We are witnessing such a change in current testing practices, such that integrative ways of assessing linguistic proficiency are being added to an increasing number of tests (for example, the Woodcock Language Proficiency Battery).

The second issue in language assessment method has concerned the dichotomy between the direct and indirect ways of testing linguistic proficiency. A direct test of linguistic proficiency samples the actual behavior being evaluated, while an indirect test samples behavior through a task that is different from a normal language-using task (Stoltz and Bruck, 1976). For example, if a student's ability to use the past tense in speech is to be assessed, a direct way of assessment would be to ask the student to give a talk in front of the class; an indirect way would be to give him/her a multiple-choice grammar test with specific items on the past tense. Traditionally, tests have taken an indirect approach to assessing oral language proficiency, usually, in multiple-choice formats. Possibly, this is due to "paper-and-pencil" tests being easier to control, administer, and validate. However, there may be a basic problem with assessing oral language through indirect means since oral language is a chameleon-like phenomenon, in constant change (Tarone, Note 3). Different types of speech elicitation methods produce different types of errors in second language users (LoCoco, 1976). For example, translation into the second language leads to a higher frequency of errors of interference from the native language than does spontaneous speech. There is an increasing tendency for language testers, then, to assess oral language proficiency in as direct a method as possible. Cohen (1980) has observed that tests lie on two continua, the first from the most discrete-point to the most integrative and the second from the most indirect to the most direct. Research suggests that test makers should adopt the latter extremes of each dichotomy.

Sociolinguistic Proficiency

This aspect of communicative proficiency concerns the social aspect of language use. Use of language, or communication, is very strongly affected by social constraints. Language is always used in a context, and consequently, oral language must be assessed within its appropriate contextual framework. Hence, this aspect of communication should be included within a model of language assessment.

There is ample evidence in the literature that language use varies as a function of various social and contextual factors. Ferguson (1959) identified twelve different speech functions related to the setting of the interaction, the topic under discussion, and the participants involved in the communication act. Fishman (1965) discussed the domains in which the form of language use varies. Bilingual persons often report an inability to convey messages effectively in certain topics (for example discussing psychology vs. discussing the weather), or to certain people (talking to one's parents vs. talking to one's colleagues) even in a language in which they are generally very fluent. Morrow (1977) defined receptive sociolinguistic proficiency as being the ability to gauge speech in terms of:

1) the setting to which an utterance is appropriate,
2) the topic being discussed,
3) the function of the utterance,
4) the attitudes conveyed by the speaker,
5) the presuppositions behind the utterance,
6) the role or the status that the speaker is adopting,
7) the level of formality of the utterance,
8) the mood of the speaker.

The same skills listed above for the reception of oral messages should also apply to the production of oral expressions. Basically, we are concerned with the use of the contextually appropriate register or code. These skills are probably very much affected by cultural constraints or cultural norms of acceptable behavior. The register that is used in Arabic, for example, to address the elderly is slightly more formal than the one used with younger peers. This reflects the deference that is accorded to the elderly in the Arabic culture (Berger, 1962). Bachman and Palmer (Note 1) include in the category of sociolinguistic competence the following sub-traits: the ability to distinguish registers, mastery of nonliteral figurative language, and use of relevant cultural allusions.

In sum, oral communicative proficiency consists of a conglomeration of skills, both linguistic and sociolinguistic in nature. These two components of communication have been discussed separately, and the skills that constitute each type of proficiency have also been singularly identified. Nevertheless, the interdependency of these separate skills cannot be sufficiently stressed. The potency of a message depends on the interlocutor's linguistic proficiency, which itself would determine and, in turn, be determined by the interlocutor's sociolinguistic proficiency. For example, a person's pragmatic skills, the ability to use target language vocabulary (a linguistic skill), would constrain the appropriateness of the register used (a sociolinguistic skill) and in turn, knowledge of the norms for appropriate language use in a specific social context (a sociolinguistic skill) would determine the choice of one set of words over another (vocabulary, a linguistic skill). The effectiveness of communication, then, may have to be judged relative to the function that language is supposed to serve in a particular context. An analogy comes to mind: if you are asked to leave on a long trip with only three items as luggage, where you are to go and what you are to do there would determine to a large extent what those items are to be. A fishing trip would require a very different set of clothing than, for example, a formal business trip. We need to know how the target language is to be used, in order to be able to, first, teach it, and second, to assess proficiency in it. Thus, we return to the question asked at the beginning of this

paper: what linguistic and sociolinguistic skills are needed in order for a person to survive, be accepted, and perhaps excel in a given situation?

SETTING STANDARDS FOR LEVELS OF PROFICIENCY

The question remains as to what characterizes successful use of a language. When a person does not possess native-like fluency, as may be the case for most second-language users, the need arises to identify those skills that are most important for successful communication; that is, we need to know how proficient is proficient enough. Further, since verbal interaction takes many forms and shapes, we need to know what it is that allows a second-language user to survive in a specific verbal interaction or a specific linguistic environment. The following questions seem appropriate.

1) How proficient must a limited-English-proficient student be to survive in an all-English classroom?

2) How proficient must a bilingual teacher be in order to teach in a classroom where the target language is used as the medium of instruction?

Questions similar to the above may be asked for the whole gamut of language environments and communication acts in which language users engage. For instance, how proficient must a limited-English-proficient person be to work at a hospital where English is the common language of communication, or to become an effective salesperson?

The answers to the above and other similar questions should then be used as guidelines for setting standards for minimum levels of proficiency in a language. This is the second step in establishing a language assessment system (right-hand-side box in Figure 1).

When standards or criteria for proficiency in a language are being established, the issue of levels or degrees of "success" needs to be addressed. At a minimal level, successful language use may refer to simply getting a message or an idea across. At a more exigent level, successful language use, in addition to getting the message across, may refer to being accepted or positively regarded by the interlocutor(s); this becomes especially pertinent for a non-native speaker from a linguistic minority interacting with a native speaker from the majority population.

To arrive at standards for a minimal level of proficiency, it would be helpful to examine the research that is currently being carried out on language as it is used in a specific context.

Standards for a Minimal Level of Proficiency

Although research is currently being done on language use in different situations, very few such studies have assessment as their stated goal. Research in second-language learning and use and research in second-language assessment have been developing independently from one another. Classroom-centered research is a good example of this autonomy: the questions being asked in that area of research — regarding the type of language that is used in the classroom or the type of teaching strategies used by teachers — have not been concerned with setting standards for survival in those particular environments, although as Allwright (Note 4) has pointed out, the view that the classroom is a setting for language acquisition has brought the field closer to second-language acquisition studies. The research findings in those areas can still be useful, and it would behoove researchers in assessment to examine that body of literature to obtain descriptions of language used by different types of interlocuters in different language environments. The following areas of research may provide useful sources of information for researchers in assessment.

1) For students who have a limited proficiency in the language of instruction, the question to be asked regards the type of language that is used in the classroom by the teacher as well as by native-speaker peers. It has been suggested (Cummins, 1980) that for limited-English-proficient students, some type of formal academic language skills — what Cummins has called Cognitive Academic Language Proficiency (CALP) — is needed in order for a minority child with limited proficiency in the majority language to survive in a majority classroom setting. CALP skills have been defined as being context-free or context-reduced. Empirical work is needed to identify, test, and further characterize what those particular skills might be. Standardized tests, normed on native speakers of English, are not suitable for limited-English-proficient students because we do not know for a fact that a non-native speaker needs to be as proficient as a native speaker in order to be an effective language user within the classroom. We return to the initial question set forth in this paper: what is the smallest amount of linguistic and sociolinguistic baggage that a child requires for a given "trip"? In effect, what type of language is used among peers in the classroom; what kind of student-teacher and teacher-student interactions take place; beyond the classroom, what language does a child need in the playground? Classroom-centered research provides a first step in answering these questions; Cathcart (Note 5), Lightbown (Note 6), and Schinke (Note 7) have examined the language that students are exposed to in a classroom where a language other than their dominant one is used.

2) For bilingual teachers who will be teaching in a nondominant language, it is essential to know what type of language and what type of teaching strategies are needed to convey information to students at different grade

levels. The research by Allwright (1980), Cathcart and Olsen (1976), Chaudron (1977), Hamayan and Tucker (1980), and Johnson (Note 8) should be helpful in delineating those skills. (For a review of this literature, see Long, 1980.)

3) For adults learning a second language, we must know the specific language skills that are needed to function in different language contexts. For the non-native speaker who will be attending a university in a nondominant language, the question is: what types of language skills are needed in order for a student to survive in a university setting where his/her nondominant language is used as the medium of instruction? The work of Allwright (1980), Bailey (Note 9), Gaies (1977), and Long and Sato (in press) would help provide answers to this question. For the non-native speaker who, for example, will be working in a hospital, the question regards the types of language skills that are needed to carry out tasks involved in the job. Educators involved in English for Special Purposes curricula and in notional syllabi took this approach many years before language assessment researchers got involved in it. (However, as Widdowson (1978) notes, while the notional syllabus claims to develop communicative competence within the design of the syllabus, this goal is not always achieved; the notional syllabus still takes the approach that language is a conglomeration of units.) It would be beneficial to those of us involved in language assessment to examine those syllabi and the tests that accompany them.

SOCIAL STANDARDS FOR LEVELS OF PROFICIENCY

Social aspects of language use are crucial for a comprehensive system of communicative proficiency assessment. Lambert and others (see for example, Gardner and Lambert, 1972) have pointed out the importance of language as a variable that interfaces with attitudes: attitudes are crucial in language learning processes; and language use, in turn, is important in the formation of attitudes. To date, there is ample evidence concerning the facilitative role that positive attitudes play in second-language learning (Tucker, Hamayan and Genesee, 1976). More recently, there is a growing body of evidence for speech, or language use, as a strong marker for group identity and for the formation of value judgments. The way you speak marks you as belonging to one ethnic group or another, as coming from one social class or another, and as having one educational background or another (Hymes, 1972). There is evidence (Giles, Bourhis and Taylor, 1977) that people use speech to form attitudes toward others and they depend on speech as a marker to such an extent that it may even supercede information regarding the other person's race (McKirnan, Smith and Hamayan, Note 10).

These findings are important to take into consideration both in teaching the various aspects of language that are deemed necessary for successful communication in a specific situation and in assessing the readiness of a non-native speaker to plunge into those situations where the target language is the medium of communication. These social issues must then be taken into account in establishing the criteria for successful use of a language. In this case, success may be defined in terms of getting a message across and in terms of being accepted by members of a group who are proficient in that target language. In a recent study, Bailey (Note 9) found that non-native-speaking university teaching assistants' accents, to a large extent, accounted for native-speaking students' evaluations of their effectiveness as teaching assistants. It is important, then, to train our language students to sound native enough and prestigious enough — whether that be linguistically (sound, grammer, pragmatics) or sociolinguistically — so that not only does the message get across, but the speaker is actually accepted and evaluated positively by others in his/her immediate environment.

SUMMARY

Thus far, language assessment procedures fall short of a comprehensive model that takes into account the nature of communicative proficiency, including both linguistic and sociolinguistic aspects. There seems to be a need to describe successful communication in different language environments and to set up criteria for needed linguistic and sociolinguistic skills, considering the social implications of speech that falls below the proficiency levels of a native-speaker.

Reference Notes

1. Bachman, L.F. and Palmer, A.S. "The Construct Validation of Some Tests of Communicative Competence." Paper presented at 1981 TESOL Convention, Detroit.

2. Cziko, G. "Developing Models of Communicative Competence: Conceptual, Statistical, and Methodological Considerations." Paper presented at 1982 TESOL Convention, Honolulu.

3. Tarone, E. "Interlanguage as Chameleon: A Problem of Observation." Paper presented at 1979 TESOL Convention, Boston.

4. Allwright, D. "Classroom-Centered Research on Language Teaching and Learning — A Brief Historical Overview." Paper presented at 1982 TESOL Convention, Honolulu.

5. Cathcart, R. "Will the Real Data Please Stand Out?" Paper presented at 1982 TESOL Convention, Honolulu.

6. Lightbown, P. "Classroom Language as Input to Second-Language Acquisition." Paper presented at 1982 TESOL Convention, Honolulu.

7 Schinke, L. "English Foreigner Talk in Content Classrooms." Unpublished Ph.D. dissertation, Northwestern University, Evanston, Illinois, 1981.

8 Johnson, N. "Teacher Student Interaction in Bilingual Classrooms: Four Approaches to Error Feedback." Paper presented at 1980 TESOL Convention, San Francisco.

9 Bailey, K. "Power of Language, Language of Power: The Teaching Behavior of Native and Non-native Speakers." Paper presented at 1982 TESOL Convention, Honolulu.

10 McKirnan, D.J., Smith, C.E. and Hamayan, E.V. "A Sociolinguistic Approach to the Belief-Similarity Model of Racial Attitudes." In review: *Journal of Experimental Social Psychology*

Bibliography

Allwright, D. "Turns, Topics and Tasks: Patterns of Participation in Language Learning and Teaching." In D. Larsen-Freeman (ed.), *Discourse Analysis in Second Language Research*. Rowley, Mass.: Newbury House Publishers, 1980.

Berger, M. *The Arab World Today*. New York, N.Y.: Doubleday and Co., 1962.

Brown, D.H. *Principles of Language Learning and Teaching*. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1980.

Canale, M. and Swain, M. "Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing." *Applied Linguistics*. 1, 1980, 1-47.

Carroll, B.J. *Testing Communicative Performance*. Oxford: Pergamon Press, 1980.

Cathcart, R.L. and Olsen, J.W.B. "Teachers' and Students' Preferences for Correction of Classroom Conversation Errors." In J. Fanselow and R.H. Crymes (eds.), *On TESOL '76*. Washington, D.C.: TESOL, 1976.

Chaudron, C. "A Descriptive Model of Discourse in the Corrective Treatment of Learners' Errors." *Language Learning*. 27, 1977, 29-46.

Cohen, A.D. *Testing Language Ability in the Classroom*. Rowley, Mass.: Newbury House Publishers, 1980.

Cummins, J. "The Entry and Exit Fallacy in Bilingual Education." *NABE Journal*. 4, 1980, 25-59.

Erickson, J.G. "Communication Assessment of the Bilingual Bicultural Child." In J.G. Erickson and D.R. Omark (eds.), *Communication Assessment of the Bilingual Bicultural Child*. Baltimore, MD: University Park Press, 1981.

Ferguson, C. "Diglossia." *Word*. 15, 1959, 325-340.

Fishman, J. "Who Speaks What Language to Whom and When?" *Linguistique*. 2, 1965, 67-88.

Gaies, S.J. "The Nature of Linguistic Input in Formal Second Language Learning: Linguistic and Communication Strategies in ESL Teachers' Classroom Language." In H.D. Brown, C.A. Yorio, and R.H. Crymes (eds.) *On TESOL '77*. Washington, D.C.: TESOL, 1977.

Gardner, R. and Lambert, W.E. *Attitudes and Motivation in Second Language Learning*. Rowley, Mass.: Newbury House Publishers, 1972.

Giles, H.; Bourhis, R.Y. and Taylor, D.M. "Towards a Theory of Language in Ethnic Group Relations." In H. Giles (ed.), *Language, Ethnicity and Intergroup Relations*. London: Academic Press, 1977.

Hamayan, E.V. and Tucker, G.R. "Language Input in the Bilingual Classroom and Its Relationship to Second Language Achievement." *TESOL Quarterly*. 4, 1980, 453-467.

Hymes, D. "Models of the Interaction of Language and Social Life." In J.J. Gumperz & D. Hymes (eds.), *Directions in Sociolinguistics*. New York, N.Y.: Holt, Rinehart & Winston, 1972.

LoCoco, V.G. "A Comparison of Three Methods for the Collection of L2 Data: Free Composition, Translation, and Picture Description." *Working Papers in Bilingualism*. 8, 1976, 59-86.

Long, M. "Inside the 'Black Box': Methodological Issues in Classroom Research on Language Learning." *Language Learning*. 30, 1980.

Long, M. and Sato, C. "Classroom Foreigner Talk Discourse: Forms and Functions of Teachers' Questions." In H. Seliger and M.H. Long (eds.), *Classroom Language Acquisition and Use: New Perspectives*. Rowley, Mass.: Newbury House Publishers, in press.

McLaughlin, B. *Second Language Acquisition in Childhood*. Hillsdale, N.J.: Erlbaum Press, 1978.

Morrow, K.E. *Techniques of Evaluation for a Notional Syllabus*. Reading: Center for Applied Language Studies, University of Reading, 1977.

Oller, J.W. and Hinofotis, F. "Two Mutually Exclusive Hypotheses about Second Language Ability: Indivisible or Partially Divisible Competence." In J.W. Oller and K. Perkins (eds.), *Research in Language Testing*. Rowley, Mass.: Newbury House Publishers, 1980.

Selinker, L. "Interlanguage." *International Review of Applied Linguistics, 10*, 1972, 201-231.

Stoltz, W. and Bruck, M. *A Project to Develop a Measure of English Language Proficiency. Final Report to the National Center for Education Statistics*. Arlington, VA: Center for Applied Linguistics, 1976.

Tucker, G.R., Hamayan, E.V. and Genesee, F. "Affective, Cognitive and Social Factors in Second Language Acquisition." *Canadian Modern Language Review, 23*, 1976, 214-226.

Widdowson, H. "National-Functional Syllabuses 1978: Part IV." In C.E. Blatchford and J. Schachter (eds.), *On TESOL '78*. Washington, D.C.: TESOL, 1978.

THE IDEA ORAL LANGUAGE PROFICIENCY TEST: A CRITICAL REVIEW

Pamela A. McCollum
The University of Texas at Austin

The purpose of this article is to analyze the IDEA Oral Language Proficiency Test (IPT) which is published by Ballard and Tighe, Incorporated of Whittier, California. The IPT is part of the IDEA program which consists of the IDEA Oral Language Proficiency Test and the IDEA Oral Language Management Program of Developmental English Activities. This program is an innovation in that it is the first language proficiency test which prescribes a specific language development program for Limited-English-Proficient (LEP) students based on its results.

Results from the IDEA Oral Language Proficiency Test are used to place LEP students into a language development program which is purported to correspond to their appropriate level of language proficiency and which will provide them with the skills which are necessary for reading. The IDEA program has been state-adopted in California and Texas and is widely used in Colorado, New York, Arizona, Oregon, Washington and Hawaii. While the appeal of such a self-contained program to teachers and administrators is obvious, a close examination of the program reveals that both its claims and the theoretical base upon which it is founded must call its use as an oral language proficiency measure for limited-English-proficient students into question.

Development

The authors state that the IDEA Oral Language Management Program was developed in response to a need for a developmentally sequenced approach to oral language skills for first and second language learners which would prepare them for reading. As a result, Wanda Ballard and Phyllis Tighe, two experienced California classroom teachers, developed a sequential program of English activities which would give students a strong base in oral language prior to reading. Ballard and Tighe found that L2 learners as well as some L1 students needed instruction in oral English activities prior to reading. Their program is said to address the needs of both groups of students by providing instruction in the areas of vocabulary, comprehension, syntax and verbal expression at seven language skill levels.

The IPT was developed after the IDEA Language Management Program was implemented. The authors, Wanda Ballard, Phyllis Tighe and Dr. Enrique F. Dalton, who acted as the technical advisor, state that the purpose of the IPT is "to determine the level of oral language mastery as it relates to the IDEA Oral Language Management Program and to accepted levels of NEP/LSP/FEP (non-English-proficient/limited-English-proficient/ and fully English-proficient) classifications" (Dalton, 1980, p. 5). The IPT is available in Spanish, Portuguese and English and is to be used in kindergarten through eighth grade.

Test Format

The English version of the test has two forms, A and B, which consist of eighty-three items which are said to measure proficiency in the areas of vocabulary, syntax, comprehension and verbal expression. Thirty-five of the test items require students to respond to questions related to cartoon drawings in the test booklet. The remainder of the items require students to perform such tasks as responding to commands to indicate comprehension, discriminating between minimal pairs, retelling a short story, describing an object and selecting the main ideas from a short passage which is read to them. The Spanish version of the IPT is not a translation equivalent of the English version and is only available in one form.

The IPT divides language proficiency into seven developmental levels A-F and M. The content knowledge and language skills which students are to possess at each level are specified and correspond to the material in the IDEA Oral Language Management Program levels (see Table 1). The test items at each level are said to measure the students' ability in vocabulary, comprehension, syntax and verbal expression.

The IPT differs from other language proficiency tests in that it divides language proficiency into seven developmental levels. It also differs in that it posits verbal expression as one of the dimensions of language proficiency. The authors never actually state which language skills comprise verbal expression although articulation skills are said to be included in it. An examination of the types of items which are included in the verbal expression category in the Rationale of Test Items (see Table 2) shows that this category contains a number of different types of items. They include sentence repetition, the production of the future and past tenses, discrimination of minimal pairs, descriptive skills (where one is required to describe a pencil), and the identification and retelling of main ideas from a spoken paragraph. It is not clear why these items were not placed in one of the three categories-vocabulary, comprehension or syntax or how the IPT scoring system reflects one's articulation skills.

The choice and assignment of test items by skill area and developmental level at times seems illogical (see Table 2). Test items across a skill area do not consistently reflect a progression from less to more complex in terms of the type of tasks which students are asked to perform. For example, in the verbal expression skill area, students have to perform sentence

repetition tasks at Level D while they are required to discriminate between minimal pairs at the following two higher levels of language proficiency. At Level F, which is the highest level which contains test items (Level M has no items but indicates a student responded correctly to 80% of the items on Levels B-F) in the verbal expression category, students are asked to discriminate between minimal pairs and select and retell the main ideas from an orally read paragraph.

Table 1

Summary of the IDEA Proficiency Test Content

A LEVEL A student knows less than 50% of skills listed in LEVEL B.

A LEVEL B student can:

1. tell his name and age.
2. identify family and common school personnel, classroom objects, basic body parts, common pets and fruits.
3. use present tense verb "to be".
4. use regular plurals.
5. answer simple "yes/no" questions appropriately.
6. follow simple directions involving basic positions in space.

A LEVEL C student can:

1. identify common occupations, clothing, farm animals, and foods.
2. express himself using the present progressive tense (he *is* work*ing*) of common verbs.
3. use conjunctions and negatives correctly.
4. follow the teacher's directions related to identifying positions on a page.
5. repeat simple sentences correctly.
6. comprehend and remember major facts of a simple story.

A LEVEL D student can:

1. identify common modes of transportation and household items.
2. name the days of the week.
3. describe common weather conditions.
4. use irregular plurals and possessive pronouns correctly.
5. ask simple future tense questions.
6. understand and express comparative and quantitative concepts.
7. understand and name opposites of key words.
8. follow directions of teacher involving movement in space.
9. repeat complex sentences correctly.
10. understand and identify moods in a simple story.
11. express himself using the present and future tense.

A LEVEL E student can:

1. identify money, marine animals, and common tools.
2. use superlatives and past tense correctly.
3. understand and name opposites of key words.
4. ask past tense questions.
5. discriminate differences in closely paired words.
6. describe and organize the main properties of common objects.

A LEVEL F student can:

1. identify the seasons and unusual occupations and animals.
2. use conditional and past perfect tenses of verbs.
3. discriminate fine differences in closely paired words.
4. express himself using past tense correctly.
5. comprehend and predict the outcome of a story.
6. recall and retell the main facts of a story.

Dalton, Enrique F. *The IPT Oral Language Proficiency Technical Manual.* Whittier, Calif.: Ballard & Tighe, Inc., 1979.

Table 2

IPT Test Items by Skill Area and Developmental Level

	LEVEL B*	LEVEL C	LEVEL D	LEVEL E	LEVEL F
VOCABULARY	3 People 4 School 5 Body Parts 6 Animals: Pets 7 Food: Fruit and Vegetables	13 People: Occupations 14 Clothing and Accessories 15 Animals: Farm 16 Food: Common	30 Transportation 31 Household Items 32 Adjectives: Weather 36 Calendar: Days of Week	49 Money 50 Animals: Marine 51 Miscellaneous Items 52 Tools	66 People: Occupations 67 Calendar: Holidays Festivals, Seasons 68 Animals
COMPREHENSION	8 Spatial Concepts: Space Prepositions 9 Commands	21 Spatial Concepts: Directionality on Page 22 Commands 26–29 Story: Factual Recall	34 Comparative Concepts 37–38 Opposite Concepts 39 Spatial Concepts: Directionality in Space 43–44 Story: Determining Mood	55–56 Opposite Concepts 57 Time Concepts	78–79 Story: Predicting Outcome
SYNTAX	10 Yes/No Response 11 Regular Plurals 12 Verb "to be"	17 Pronouns: Subject 18 Verbs: Present Progressive 19 Mass Nouns 20 Conjunctions 23 Negative Statements	33 Verbs: Habitual Present 35 Comparatives 40 Questions: Future Tense 45 Irregular Plurals 46 Pronouns: Possessive	53 Superlatives 54 Verbs: Irregular Past Tense 58 Questions: Past Tense	69 Verbs: Conditional 70 Verbs: Past Perfect
VERBAL EXPRESSION**	1–2 Survival Responses	24–25 Sentence Repetition	41–42 Sentence Repetition 47–48 Use of Future Tense	59–62 Auditory Discrimination 63–65 Descriptive Skills	71–74 Auditory Discrimination 75–77 Use of Past Tense 80–83 Story: Retelling Main Ideas

* LEVEL A Placement is indicated by less than 50% proficiency on LEVEL B.
** Includes Articulation Skills.

Dalton, Enrique F. *The IPT Oral Language Proficiency Test Technical Manual* Whittier, Calif: Ballard & Tighe, Inc., 1979.

The juxtaposition of test items through language skill areas at each language proficiency level is also inconsistent. At Level D proficiency, students are asked to repeat the days of the week to test vocabulary, to repeat sentences as a measure of verbal expression and to determine mood from an orally read paragraph to test comprehension. While the first two tasks may be of equal difficulty, determining the mood of a passage is a more difficult task and one which is of a different magnitude. In addition, in order to perform such a task, a student must understand and be able to make inferences from language. Inferential skills are generally acquired through instruction and are not related to one's ability to communicate in a language. Since some native speakers of English who are uneducated would have difficulty in performing such a task, it should not be expected of L2 learners. It is better suited for a reading achievement test, rather than a test of language proficiency.

There are also problems with the language which is used in the IPT test items. Many times, the directions or cues which accompany a test item are too difficult for the level of language proficiency which is being tested. For example, on Level C of form A, as the teacher points to a picture of cake and ice cream, she says, "See the cake and ice cream? These go together at a party. On my birthday I have cake *and* ice cream." Then the teacher points to a picture of a table and chair and says, "These go together at school. Tell me what you have at school. "I . . ." The language in many of the test items needs to be more direct, briefer and made more appropriate to the language-proficiency level which is being tested. A child at Level C proficiency could conceivably be able to correctly respond with "I have a table and a chair," but be unable to get past the language which was designed to elicit it.

Scoring

Responses to test items are scored for appropriateness and completeness. This means that with the exception of items which require the identification of lexical items, questions must be answered with full sentences in order to receive credit. In some cases, if the child fails to respond to an item, a paraphase of the item which appears on the examiner's score sheet should be given. If the child fails to respond to the initial item and its paraphrase, the item is scored as incorrect and the examiner is to proceed to the next item.

Testing continues through successive levels until the child fails to answer 50% of the items at a particular level correctly. When this occurs, testing is discontinued, and the child is diagnosed as having reached his/her instructional level for the Oral Language Management Program which provides appropriate English language activities which correspond to the child's level of English mastery. For example, if a child has progressed through levels B and C of the IPT by answering 80% of the items correctly, but fails to answer 50% of the items on Level D correctly, testing is stopped. She/he would be classified as having Level C mastery of English and would begin receiving instruction at Level D, his/her instructional level, in the language management program.

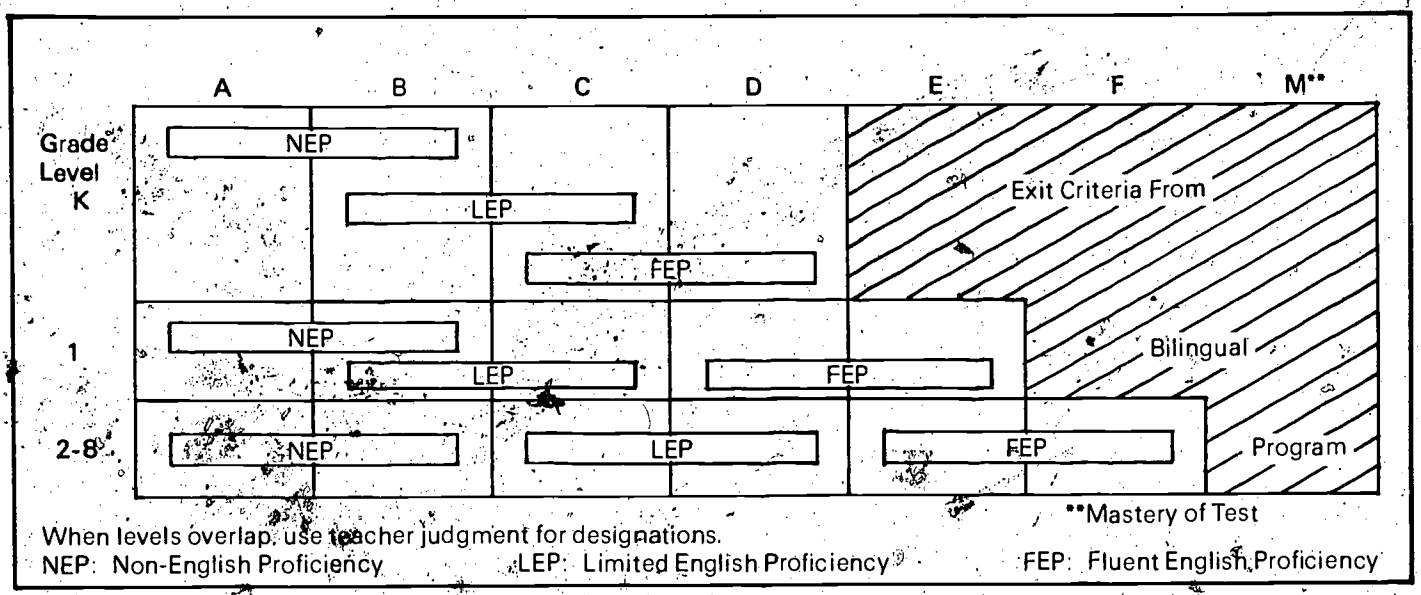
There are no items for Levels A or M. The former classification indicates that a student failed to respond correctly to 50% of the items on Level B. The Level M classification indicates that 80% of the items on the test (Levels B-F) were answered correctly. Students who place at Level M, the mastery level, are said to have achieved mastery of the test and are therefore not in need of instruction in the oral language management program.

Score Conversion

A conversion table is provided to derive the corresponding NEP/LEP/FEP classifications from the IPT scores to aid teachers in making decisions about placement or exit from the Bilingual program. The IPT score and its NEP/LEP/FEP equivalent are on a sliding scale which corresponds to the students' grade level. (See Figure 1).

Figure 1

Correlation between IPT Test Levels and NEP/LEP/FEP Designations*



When levels overlap, use teacher judgment for designations.
NEP: Non-English Proficiency LEP: Limited English Proficiency FEP: Fluent English Proficiency

*Dalton, Enrique F. *The IPT Technical Manual*. Whittier, California: Ballard and Tighe, Inc., 1979, p. 23.

Kindergarten students who score at Level C or D on the IPT are classified as fully English proficient, while first graders must score at Level D or E to be classified as FEP. Students in grades 2-8 must score at Level E or F to be classified as fully English proficient. Examination of Figure 1 shows that with the exception of Level A and Level F for children in grades 2-8, all of the language classifications overlap with another category. In these cases, teachers are advised to use teacher judgment to determine in which of the two categories a child should be placed.

The IPT Technical Manual states that these correlations were determined by a research study, but it is not described in the manual. There are several problems with this scoring system. First of all, correlating L2 learners' language proficiency with grade in school is unacceptable as year in school does not indicate the same amount of exposure to the language in the case of L2 learners. Secondly, due to the overlapping of categories within grade levels and reliance upon teacher judgment to classify students, one has to question the instrument's reliability. Finally, at all grade levels students are classified as fully English proficient, but are not eligible for exit from the bilingual program until they reach the next highest level of proficiency. In summary, the IPT scoring system, which assigns children to one of seven developmental levels of proficiency corresponding to grade level and then converts that score to an NEP/LEP/FEP classification upon which to make judgments about bilingual program placement, does not follow common procedures, and the rationale for this departure is not presented.

Field Testing

The IPT was normed on 2,061 identified NEP/LEP students in grades K-8 who were enrolled in bilingual classes in rural, urban and suburban schools in California. The authors do not state how the students were originally identified as NEP/LEP, nor do they provide information regarding their ethnicity or the variety of Spanish which was spoken by them. The examiners who conducted the field testing were all bilingual/bicultural certified teachers who received a ten-hour training session on the proper administration of the IPT. While the IPT Technical Manual reports high reliability and validity figures, close examination of the test's theoretical base, its explanation of what constitutes language performance and its failure to recognize some fundamental differences in L1 and L2 learners must call the use of this test as a language proficiency measure for L2 learners into question.

Perhaps the most glaring flaw in the IPT is its failure to distinguish between the L1 and L2 language acquisition process and characteristics of L1 and L2 learners. In the technical manual, Dalton describes the following stages of language development in language acquisition: babbling, echolalic, telegraphic and syntactic. He states,

> "The child finds that language is a valuable tool and uses it more extensively in daily social situations. Frequent opportunities to use the language tend to nurture and expand its growth in the child. This is to be encouraged in that oral language skills, both receptive and expressive, are the foundation of initial readings and, eventually, literacy." (Dalton, 1980, p. 2)

It is not clear what applicability stages of language development in first language acquisition have to L2 learners of English or even to native speakers of English who need further oral language development prior to reading. Children generally enter school already possessing nearly full linguistic competence in their native language. Crucial questions which were ignored prior to the development of this test, which purports to measure the language proficiency of both L1 and L2 learners of English, are: a) Does a child learning English as a second language progress through the same stages in the language acquisition process as an L1 learner? b) If the L2 learner does not pass through the same developmental stages as an L1 learner, what are the characteristics of the L2 acquisition process? and c) Which elements of the lexicon, syntax, phonology and functional language are acquired at each stage by both the L1 and the L2 learner?

There is a considerable body of research related to the similarities between the L1 and L2 language acquisition process. Brown's study (1973) was an observational, longitudinal study of English acquisition of three children. The study found that children learning English progress through a series of stages where they acquire a set of fourteen grammatical morphemes in roughly the same order. These results were further validated by a cross-sectional study on the order of morpheme acquisition in twenty-one preschool-age children who were learning English as a first language (de Villiers and de Villiers, 1973). The morpheme sequence which was obtained in the de Villiers' study was not identical, but corresponded closely to Brown's results.

Brown's work served as the impetus for many subsequent studies of morpheme sequencing in the area of second language acquisition. Dulay and Burt conducted a series of studies on children of different language backgrounds who were learning English as a second language (Dulay and Burt, 1973, 1974a, 1974b). In these studies, the Bilingual Syntax Measure was used to elicit speech from children ages 5-8 who were learning English as a second language. Analysis of the children's responses for the presence or absence of grammatical morphemes showed that L2 learners of English, regardless of their language background demonstrated the same order of morpheme acquisition in English. From this, Dulay and Burt conclude that "the strategies of second-language acquisition are universal" (1973, p. 256). It should be noted that the morpheme sequence which was found by Dulay and Burt was not identical to Brown's or to that of de Villiers and de Villiers. Dulay and Burt feel that the differences which appeared in their data are due to the greater linguistic and cognitive maturity of the L2 learner.

Other studies which examined the issue of the order of morpheme acquisition in L2 learners of English and which produced similar results were Fathman (1975) and Krashen, Sferlazza, Feldman & Fathman (1976). Although these studies used a different speech elicitation test, the SLOPE, the former found that both Spanish and Korean children exhibited a similar morpheme acquisition order while learning English. The latter replicated the study with adult L2 learners and obtained a similar acquisition sequence. These studies suggest that the order of morpheme acquisition in English by L2 learners is invariant regardless of one's native language or age. While this group of studies has been criticized on several grounds by Larsen-Freeman (1975), Rosansky (1976) and Hakuta (1976), the similarities among the data are impressive and bear further research.

In the presentation of the rationale for the IPT, Dalton mentions the original Brown study on L1 learners of English only in passing. He fails to explain how the study relates to the construction of the IPT or the items which were chosen to measure syntax at each of its levels, which contain items B-E. The issue of possible differences or similarities in the L1 and L2 acquisition process is never raised. Furthermore, the ordering of the items which measure one's control of syntax according to proficiency levels is illogical and does not follow a progression from simple structures at low proficiency levels to complex at higher levels. Examination of the IPT shows that the material tested at each proficiency level corresponds to the material presented in the IDEA Oral Language Management Program levels, rather than to levels of difficulty in language structure or to morpheme acquisition sequences in English by L1 and L2 learners.

Another indication that the IPT ignores characteristics of L2 learners is the fact that language proficiency levels are correlated to grade levels. While it is possible to correlate a child's L1 language development with his/her grade in school, such a correlation in the case of the L2 learner does not reflect the same amount of exposure to the language. A child who is a new arrival and who is beginning to learn English in the 6th grade can not be expected to have the same language skills as a native, English-speaking child who is in the same grade. Furthermore, such a system fails to accurately distinguish among L2 learners in the same grade who may have differing amounts of exposure to English due to their date of arrival in the United States in the case of older immigrant children, or due to sporadic school attendance in the case of migrant children.

Construct Validity

The IPT has even more serious problems of a theoretical nature in that the basic construct which it purports to measure is not clearly defined. While the IPT is labeled as an oral language *proficiency* test, the word *mastery* is used interchangeably with proficiency throughout. The point is made that linguistic competence must be inferred from linguistic performance data and that the constructs which underlie linguistic performance and upon which the IPT is based are the following: a) language is developmental; b) it is incremental; c) it is systematic; d) it is used to communicate in a social context; and e) it involves both receptive and productive skills (Dalton, 1979, p. A-3). It is implied that some of the constructs which underlie linguistic performance, as described by the author, are inherent in the IPT due to its format which places the child in a dyadic social situation with the examiner whereby the child must comprehend and produce progressively more difficult language. From the child's responses, his/her linguistic performance is evaluated and the level of language proficiency is determined. Dalton states, "Validation of the above constructs can be inferred if the data generated through the (construct validity) study demonstrate positive correlation with the age and grade of the student and the IPT level results" (Dalton, 1979, p. A-4).

The construct validity study was conducted on 364 students in a suburban school district in California. An analysis was made of the correlation between age and IPT results on form A and B of the English version of the test. The data were further analyzed after excluding the scores of NEP/LEP students which resulted in a sample which was comprised of native speakers of English or L2 learners who were fluent speakers of English. The same procedures were followed to derive a correlation between IPT results and the students' grade level. Not surprisingly, the study obtained high positive correlations between IPT results and age and grade level with high levels of significance on both forms of the test.

Having shown that a strong positive correlation existed between the age and grade of native speakers of English and language proficiency in their own language, Dalton concludes that these data validate the constructs which state that language is developmental, incremental and systematic. In addition, the data are also said to validate the use of the IPT as a method of assessing oral language proficiency through linguistic performance.

The fact that there is a strong positive correlation between the age and grade in school of native speakers of English hardly seems to be a worthy subject for research. Furthermore, one has to wonder why the construct validity study was conducted and what its applicability as a rationale for a language-proficiency measure for L2 learners is. An even more basic problem with the study is its definition of linguistic performance which is composed of vague generalities which are not accurate. While language may be said to be systematic and incremental in its development, one's linguistic performance is not. Linguistic performance is a speaker's use of his/her linguistic competence (underlying knowledge of the system of a language) at any one time. In a native speaker of a language, linguistic performance may vary from one occasion to the next due to such factors as memory, distractions, shifts of attention, etc. (Maclay & Osgood, 1959, p. 24)

Language proficiency measures attempt to accurately measure one's linguistic competence by inference from linguistic performance data which is variable. Due to this fact, a more germane construct upon which to measure a test for L2 learners is, what is the nature of communicative competence? What knowledge must one possess in order to produce grammatically correct speech in situationally appropriate contexts? Such an approach requires the articulation and validation of the learner's stages of language acquisition in the areas of syntax, morphology, lexicon, phonology, functional language and sociolinguistic skills. A complete theory of communicative competence would also have to order and weight items to indicate which are the most crucial for achieving intelligibility in a language. While such a theory has not been fully developed to date, researchers in the area of language testing are working toward that end. Federal and state regulations which require language proficiency testing for children in bilingual programs have caused great strides to occur in the area in recent years. Unfortunately, the IPT does not reflect recent theory and suggested practices in the field.

Content Validity

Content validity is the extent to which a test covers a representative sample of behavior in the domain to be measured. The IPT purports to measure language mastery in the areas of syntax, lexicon, phonology, morphology, comprehension and oral expression and reports extremely high percentages of items which correspond to and measure each one. This stems from the fact that test items were scored for all of the above categories which could possibly pertain to a response and not to those which are reflected in the IPT scoring system. For example, the item, "Who is this"? which requires the student to supply the lexical item, *teacher*,

is said to measure lexicon, phonology (/r/), morphology (/er/), comprehension and oral expression. In reality, the item only measures the production of a lexical item, as the IPT scoring system does not provide for the scoring of pronunciation, comprehension or oral expression. The technical manual reports that 100% of the items measure comprehension, but this is not so. In the above example, the student could have understood the question perfectly well, but might not have known the word *teacher* and supplied another word. From such a situation, one would be mistaken in concluding that the child did not comprehend the language in the question.

Within proficiency levels, the IPT has a very restricted range of items which measure each area of language. Form A of the English version of the test has an average of 16 items in Levels B-F. An analysis of the test items showed that over half of the items at each level (with the exception of Level F) call for the student to produce isolated vocabulary words in response to questions about a picture (Level B = 58%; Level C = 31%; Level D = 33%; Level F = 18%). Of the twelve items on Level B, only three measure the child's knowledge of syntax and each tests a different structure. Due to the manner in which test items were classified by the areas of language they purport to measure and also the restricted range of items tested at each level, the IPT's content validity is suspect.

Criterion Validity

The IPT test results were checked against three criteria to determine their criterion validity. Cross-tabulations of students' scores on the English version of the IPT, forms A and B, were made with a) teachers' prediction of students' IPT scores, b) students' previous scores on state-adopted language proficiency tests and c) teachers' predictions of students' NEP/LEP/FEP classifications. The technical manual reports high validity figures and levels of significance for all three analyses and concludes that the IPT has criterion and predictive validity as a measure of language proficiency.[1]

The most interesting result was the correlation of students' IPT scores with their NEP/LEP/FEP classifications as determined by one of four California state-approved oral language proficiency tests: the LAB, the LAS, the BINL, and the BSM.[2] The statistical analysis correlated the students' scores on the IPT, which could be one of the seven possible levels A-F and M, with their NEP/LEP/FEP scores as determined by one of the state-approved tests. High positive correlations between IPT scores and the classifications reported by the state-approved tests were obtained (Form A, Pearson's R = .75 at the .00001 level of significance, n = 721; Form B, Pearson's R = .59 at the .00001 significance level, n = 492).

These data are interesting when viewed in light of a previous study conducted by Ulibarri, Spencer and Rivas (1981) which examined the comparability of language proficiency classifications produced by the LAS, the BSM and the BINL. The test scores of over 1,100 students in first, third and fifth grades were compared. The study showed that the proportions of students who were classified as NEP/LEP/FEP by each test were highly disparate. They concluded that since the three tests did not assign approximately the same percentages of students to the three language proficiency classifications, they were not comparable. They also warned that if different tests were used for reclassification of students, it would be possible for a student to enter and exit a bilingual program on the same day (p. 79).

Given the fact that three of the language proficiency tests used in the IPT criterion validity study were those examined in the Ulibarri, et al. study and found to be incomparable in assigning students to the same language proficiency classifications, the criterion validity figures which were stated for the IPT seem implausible. If three of the tests which were used as criteria against which to measure the predictive validity of the IPT do not show agreement among themselves in assigning students to language-proficiency levels, how could they show a high positive correlation with the IPT when a fourth test, the LAB, was added?

While the IPT reports high construct, content and criterion validity, close examination of the studies casts shadows on its validity as an accurate measure of oral language proficiency for L2 learners. For that reason, the question of its reliability will not be addressed.

Conclusion

There has been a renewed interest in the subject of language-proficiency testing in recent years largely due to mandated language-proficiency testing of students participating in bilingual education programs. The original impetus for mandated testing resulted from the Aspira Consent Decree of 1973. The majority of state and federally funded bilingual programs are transitional in scope and are aimed at bringing the limited-English-proficient (LEP) child to the point where she/he can exit the bilingual program and begin instruction in the all-English classroom. Although language-proficiency testing in bilingual programs was mandated in order to identify both students who were eligible for program participation and those who had acquired sufficient English language proficiency to

[1] Correlation of *Teachers' Prediction of Students' IPT Scores and IPT Results* — Form A, Pearson's R = .79 at the .00001 significance level; Form B, Pearson's R = .66 at the .00001 significance level. *Students' Previous Scores on State-Adopted Language Proficiency Tests and IPT Results* — Form A, Pearson's R = .75 at the .00001 significance level; Form B, Pearson's R = .59 at the .00001 significance level; *Teachers' Predictions of Students' NEP/LEP/FEP Classifications with IPT Results* — Form A, Pearson's R = .71 at the .00001 significance level; Form B, Pearson's R = .63 at the .00001 significance level

[2] The abbreviations refer to the following commonly used proficiency tests: LAB — Language Assessment Battery. LAS — Language Assessment Scale. BINL — Basic Inventory of Natural Language. BSM — Bilingual Syntax Measure.

warrant their exit from the program, virtually none existed at the time.

Previous language proficiency testing was mostly done within the area of foreign language pedagogy. With the exception of the Foreign Service Interview (FSI) which was used by the State Department, the language-proficiency tests which were used in foreign language classrooms were paper and pencil tests which emphasized receptive, rather than productive, language skills. Furthermore, due to the influence of structural linguistics, they were discrete point in nature and tested only one specific language skill in one component of language at a time, e.g., the ability to comprehend elements of the phonological system, etc.

These types of measures were unsatisfactory for they did not always positively correlate with one's ability to communicate in a language and did not satisfy the requirement of the Aspira Consent Decree which stipulated that "the placement of children in educational programs using English or Spanish as the medium of instruction be determined by their ability to 'effectively' participate in instruction" (Shuy, 1977, p. 79). When viewed in the light of prevailing linguistic theory, their definition of language proficiency as control of the syntax, lexicon and phonology of a language is too narrow and ignores a native speaker's knowledge regarding functional language and the rules which govern language use.

Shuy (1977) defines functional language competence as "the underlying knowledge that allows people to use their language to make utterances to others in terms of their goals. It includes a knowledge of what kinds of goals language can accomplish (the functions of language) and of what are permissible utterances to accomplish each function" (language strategies) (p. 81). Examples of language functions would be giving an order, making promises or apologies, etc. Operating from this position, more integrative or global language proficiency measures are called for which more closely approximate the process which speakers of a language go through when communicating in a language.

While proficiency measures today do stress productive skills, they are still for the most part discrete-point or at best quasi-integrative (tests which employ discrete-point scoring of speech samples). They still focus mainly on measuring those elements of language which are the most testable and most easily quantified — phonology and lexicon. Their focus is narrowed further in that the language used in the tests is generally restricted to the school domain.

Language-proficiency testing is poorly understood and approached with dread by teachers and administrators who feel that language-proficiency test results provide no answers for them at the instructional level. They see it as a time-consuming activity which impinges upon instructional time and yields no useful information regarding how one should instruct the child subsequently. They want to know, given the NEP/LEP/FEP classification of a student, what kind of instructional program should be provided for the child to help him/her achieve English proficiency.

The IPT Oral Language Proficiency Test is a response to this very real concern of teachers and administrators in that it provides an oral language management program of developmental activities purporting to enhance a student's language proficiency based on its results. It also claims to prepare L2 learners, as well as L1 learners who have language which is not fully developed, for reading instruction. Close examination of the IPT has shown that it has such theoretical and technical problems that one must question whether it actually measures language proficiency or simply a child's progression through the IDEA Oral Language Management Program. As it is doubtful that the theoretical construct — language proficiency — the IPT purports to measure really is being measured, both its claims and the advisability of its use for L2 learners are questioned.

References

Brown, R. *A First Language*. Cambridge, Mass.: Harvard University Press, 1973.

Dalton, E. F. *IPT Technical Manual*. Whittier, Calif.: Ballard and Tighe, Inc., 1979.

_____. *IPT Examiner's Manual*. Whittier, Calif.: Ballard and Tighe, Inc., 1980.

de Villiers, J. and de Villiers, P. "A Cross-sectional Study of the Acquisition of Grammatical Morphemes in Child Speech." *Journal of Psycholinguistic Research*. 1973, 2. 267-278.

Dulay, H. and Burt, M. "Should We Teach Children Syntax?" *Language Learning*. 1973, *23*(2), 245-257.

_____. "Errors and Strategies in Child Second Language Acquisition." *TESOL Quarterly*. 1974a. *8* (2). 129-136.

_____. "Natural Sequences in Child Second Language Acquisition." *Language Learning*. 1974b. 24 (1). 37-54.

Fathman, A. "Language, Background, Age and the Order of Acquisition of English Structures." In M. Burt & H. Dulay. (Eds.), *On TESOL '75*. Washington, D.C.: TESOL, 1975, 33-43.

Hakuta, K. "A Case Stud[illegible] Child Learning English as a Second L[illegible]*uage Learning*. 1976. *26*(2). 321-351.

Maclay, H.S. & Osgood, C. E. "Hesitation Phenomena in Spontaneous English Speech." *Word*, 1959, *15*, 19-44.

Rosansky, E. "Methods and Morphemes in Second Language Acquisition Research." *Language Learning*, 1976, *26*(2), 409-425.

Shuy, R. W. "Quantitative Language Data: A Case For and Some Warnings Against." *Anthropology & Education Quarterly*, May 1977, 7(2), 73-82.

Ulibarri, D. M., Spencer, M. L. and Rivas, G. A. "Language Proficiency and Academic Achievement: Relationship to School Ratings as Predictors of Academic Achievement." *NABE Journal*, Spring 1981, *5*(3), 47-80.

A SUGGESTED CUBIC RUBRIC FOR JUDGING THE ADEQUACY OF LANGUAGE ASSESSMENT PROCEDURES

Calvin K. Claus
National College of Education

Introduction

Over the past quarter century various classification schemes, taxonomies or rubrics have been proposed for use in analyzing the adequacy of educational tests. The *Taxonomy of Educational Objectives: Cognitive Domain*, edited by Bloom (1956), had its origins primarily in general, college-level, achievement testing. Neither this classification scheme nor the subsequent one by Krathwohl, Bloom and Massia (1964) focused specifically on language assessment. However, the applications in this area, particularly of the *Cognitive Domain Handbook*, are many. Of more direct relevance, though, is a rubric suggested by Stansfield (1982). He has proposed a three-dimensional design which incorporates linguistic components (phonology, syntax, lexicon), the traditional communication skills (listening, speaking, reading, writing) and a sociolinguistic domain (home, neighborhood, school). This 3 x 3 x 4 matrix of thirty-six cells deals with such a wide array of language assessment possibilities that it would be beyond practicality to tap all of these aspects with one test instrument. "Nonetheless, the matrix is useful in judging the validity of an instrument. Thus, as a general rule, the more cells we find included in a test, the greater its validity" (Standsfield, 1981, p. 236).

With validity as a major focal point then, it would be well to cite the *Standards for Educational and Psychological Tests* (1974) as another approach to analyzing test adequacy. Under the heading of *Content Validity*, standard E12 states: "If test performance is to be interpreted as a representative sample of performance in a universe of situations, the test manual should give a clear definition of the universe represented . . ." It comments further: "The definition should be operational, rather than theoretical, containing specifications regarding classes of stimuli, tasks to be performed and observations to be scored" (p. 45). An exploration of how some test makers have defined *specifications regarding classes of stimuli* and *tasks to be performed* has led this writer to the conclusion that another supplementary rubric is needed to judge the adequacy of language assessment procedures.

Basis for Another Rubric

Before proceeding further, a few definitions may be in order. A remark by Foster (1974) can serve as a basis for clarifying what is meant by *correct Standard English* or *correct Standard Polish* or *correct Standard (any language you can name)*. "It is unfortunate that school personnel too often are rigidly bound and limited by their concept of correct Standard English. In the reality of the life outside the school, however, the proper language is the language that brings people together and succeeds in accomplishing an objective" (p. 163).

Another term requiring clarification is *assessment*. A definition which this writer has found useful runs as follows: assessment is the systematic recording and reporting of a behavior change which is considered desirable in relation to a previously stated objective.

An obvious common thread running through both definitions is the necessity for having an objective. But how can one get a handle on the vast number of educational objectives that would constitute a "well rounded" curriculum? The National College Verb List shown in Figure 1 has been of some help in this regard. An earlier edition of this list was published by Gronlund (1970). The list as shown here appears in Pucel and Knaak (1975).

The 455 verbs have been found to be useful in generating statements of educational objectives and for developing assessment procedures to measure whether or not a learner has accomplished them. Clearly, it would not be possible to assess a person in all these behaviors at one time. However, as shown in Figure 1, the verbs and associated behaviors can be grouped in a number of ways. Further examination of the list reveals that additional classifications are possible.

One of these other ways of classification is related to the traditional dichotomy of testing techniques: Recognition test items vs. Recall (Mitchell, no date). A more functional description of these two could be, respectively: *Selected Response* items of the multiple-choice type, and *Supplied Response* items which involve single oral or written answers, as well as extended discourse both in writing and speaking along with all other remaining types of responding. Another dichotomous grouping which follows from this, reflects *2* Dimensional, silent, pencil-and-paper behavior vs. *3* Dimensional behavior involving oral-verbal production or interactions between people and between people and things. Thus, the 2 x 2 scheme or rubric shown in Figure 2 can be generated.

When we attempt to apply this 2 x 2 rubric in the specific area of language assessment, however, something seems to be missing. The missing facet emerges when referring to the admonition in the *Standards for Educational and Psychological Tests* about "specifications regarding classes of stimuli." What has been left out is the medium or channel through which the stimuli are presented to an examinee. This, then, leads to a third dichotomy: oral stimuli presentations vs. printed stimuli. Figure 3 shows what this cubic rubric contains when all three dichotomies are combined. Altogether, eight cells appear in the

Figure 1

Index Verborum Permissorum

The Functional, Forceful Four Hundred Fifty-Five

1. "Creative" Behaviors

Alter	Generalize	Question	Regroup	Rephrase	Rewrite
Ask	Modify	Rearrange	Rename	Restate	Simplify
Change	Paraphrase	Recombine	Reorder	Restructure	Synthesize
Design	Predict	Reconstruct	Reorganize	Retell	Systematize
				Revise	Vary

2. Complex, Logical, Judgmental Behaviors

Analyze	Combine	Contrast	Designate	Formulate	Plan
Appraise	Compare	Criticize	Determine	Generate	Structure
Assess	Conclude	Deduce	Discover	Induce	Suggest
		Defend	Evaluate	Infer	Substitute

3. General Discriminative Behaviors

Choose	Describe	Discriminate	Indicate	Match	Place
Collect	Detect	Distinguish	Isolate	Omit	Point
Define	Differentiate	Identify	List	Order	Select
				Pick	Separate

4. Social Behaviors

Accept	Answer	Cooperate	Forgive	Laugh	Reply
Admit	Argue	Dance	Greet	Meet	Smile
Agree	Communicate	Disagree	Help	Participate	Talk
Aid	Compliment	Discuss	Interact	Permit	Thank
Allow	Contribute	Excuse	Invite	Praise	Visit
			Join	React	Volunteer

5. Language Behaviors

Abbreviate	Call	Indent	Punctuate	Speak	*Tell*
Accent	Capitalize	Outline	Read	Spell	Translate
Alphabetize	Edit	Print	Recite	State	Verbalize
Articulate	Hyphenate	Pronounce	Say	Summarize	Whisper
			Sign	Syllabicate	Write

6. "Study" Behaviors

Arrange	Circle	Diagram	Itemize	Mark	Record
Categorize	Classify	Find	Label	Name	Reproduce
Chart	Compile	Follow	Locate	Note	Search
Cite	Copy	Gather	Look	Organize	Sort
			Map	Quote	Underline

7. Music Behaviors

Blow	Clap	Finger	Hum	Pluck	Strum
Bow	Compose	Harmonize	Mute	Practice	Tap
			Play	Sing	Whistle

8. Physical Behaviors

Arch	Climb	Hit	March	Ski	Swim
Bat	Flex	Hop	Pitch	Skip	Swing
Bend	Float	Jump	Pull	Somersault	Throw
Carry	Grab	Kick	Push	Stand	Toss
Catch	Grasp	Knock	Run	Step	Walk
Chase	Grip	Lift	Skate	Stretch	

9. Arts Behavior

Assemble	Cut	Frame	Mold	Roll	Stamp
Blend	Dab	Hammer	Nail	Rub	Stick
Brush	Dot	Handle	Paint	Sand	Stir
Build	Draw	Heat	Paste	Saw	Trace
Carve	Drill	Illustrate	Pat	Sculpt	Trim
Color	Fold	Melt	Polish	Shake	Varnish
Construct	Form	Mix	Pour	Sketch	Wipe
			Press	Smooth	Wrap

10. Drama Behaviors

Act	Direct	Enter	Imitate	Pantomime	Respond
Clasp	Display	Exit	Leave	Pass	*Show*
Cross	Emit	Express	Move	Perform	Sit
				Proceed	Turn

11. Mathematical Behaviors

Add	Compute	Estimate	Integrate	Plot	Substract
Bisect	Count	Extrapolate	Interpolate	Prove	Sum
Calculate	Cumulate	Extract	Measure	Reduce	Tabulate
Check	Derive	Graph	Multiply	Solve	Tally
Circumscribe	Divide	Group	Number	Square	Verify

12. Laboratory Science Behaviors

Align	Conduct	Dissect	Keep	Plant	Set
Apply	Connect	Feed	Lengthen	Prepare	Specify
Attach	Convert	Grow	Limit	Remove	Straighten
Balance	Decrease	Increase	Manipulate	Replace	Time
Calibrate	Demonstrate	Insert	Operate	Report	Transfer
				Reset	Weigh

13. General Appearance, Health and Safety Behaviors

Button	Comb	Eat	Fill	Taste	Unzip
Clean	Cover	Eliminate	Go	Tie	Wait
Clear	Dress	Empty	Lace	Unbutton	Wash
Close	Drink	Fasten	Stack	Uncover	Wear
			Stop	Untie	Zip

14. Miscellaneous

Aim	Erase	Hunt	Peel	Scratch	Store
Attempt	Expand	Include	Pin	Send	Strike
Attend	Extend	Inform	Position	Serve	Supply
Begin	Feel	Kneel	Present	Sew	Support
Bring	Finish	Lay	Produce	Share	Switch
Buy	Fit	Lead	Propose	Sharpen	Take
Come	Fix	Lend	Provide	Shoot	Tear
Complete	Flip	Let	Put	Shorten	Touch
Correct	Get	Light	Raise	Shovel	Try
Crease	Give	Make	Relate	Shut	Twist
Crush	Grind	Mend	Repair	Signify	Type
Develop	Guide	Miss	Repeat	Slip	Use
Distribute	Hand	Offer	Return	Slide	Vote
Do	Hang	Open	Ride	Spread	Watch
Drop	Hold	Pack	Rip	Stake	Weave
End	Hook	Pay	Save	Start	Work

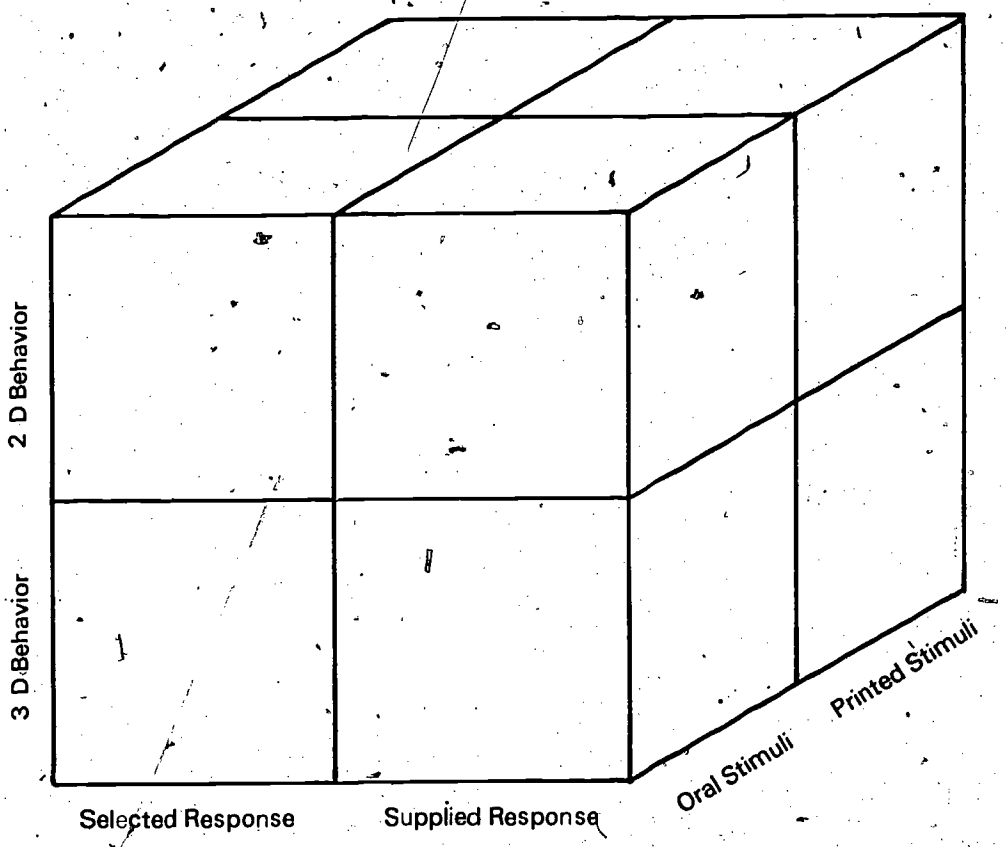
Note that all of the verbs in this entire list could be subsumed under either *Show* or *Tell*.

Figure 2

Examples of Verbs Classified According to A 2 x 2 Scheme

	Selected Response	Supplied Response
2D Behavior	Choose Detect Differentiate Discriminate Distinguish Identify Indicate Match Omit Order Pick Select Arrange Categorize Circle Classify Find Locate Mark	Define, in writing, . . . Describe, in writing, . . . List (all but two or three of the "creative" behaviors) (all of the complex, logical, judgmental behaviors) (almost all the mathematical behaviors) Chart Cite Compile Copy Diagram Follow Itemize Label Map Name Note Organize Quote
3D Behavior	Choose Detect Differentiate Discriminate Distinguish Identify Indicate Isolate Match Order Pick Place Point Select Separate Arrange Categorize Classify Find Gather Locate	Define, orally, . . . Describe, orally, . . . List (all but two or three of the "creative" behaviors) (all of the complex, logical, judgmental behaviors) (all of the social behaviors) (all of the music behaviors) (all of the physical behaviors) (almost all of the arts behaviors) (all of the drama behaviors) (almost all the laboratory science behaviors) (all of the general appearance, health, and safety behaviors) Follow Label, orally . . . Name, orally . . . Organize Quote

Figure 3



Application of the Cubic Rubric to a Sample of Language Tests

Figure 4 and Figure 5 illustrate how well two language tests fill the cells in the proposed classification scheme. A preliminary step was to phrase a statement, an educational objective, for each identifiable part of the test being analyzed. In Figure 4 the parts of the Illinois Test of Psycholinguistic Abilities (Kirk, McCarthy and Kirk, 1968) are checked against the rubric. All sections of the Short Tests of Linguistic Skills (Frederickson, 1976) are the ones shown in Figure 5.

In Figure 4 the ITPA touches all but three cells, with heavy representation in the 2D, Selected Response, Printed Stimuli and the 3D, Supplied Response, Oral Stimuli categories. A better balance and more complete coverage of cells is evident in Figure 5. Here the STLS missed only in two categories. When gaps appear it now is possible to generate a set of statements and related test situations to increase the coverage. For example, on the STLS in the 3D, Selected Response, Printed Stimuli Cell, an objective such as the following might fit: Pick out, from each of six groups of four objects, the one specified by a set of printed directions. For the missing 3D, Supplied Response, Printed Stimuli Cell, a statement such as this might suffice: Ask someone for help as if being in the position of a person in three different situations described in writing.

In addition to the two tests analyzed here in some detail, a quick, summary check was made on seventeen other tests taken in a sampling of the *ETS Test Collection: Tests in Microfiche* (1975 through 1981). Figure 6 shows how broadly or narrowly each test matches up to the facets of the cubic rubric. It becomes quite clear, here, that several tests barely tap into the total number of possibilities for assessing language competence. This serves as a caution, rather than a value judgment for test users as well as test authors. If factors of time, money and energy loom large, then some aspects of linguisitic performance will not be assessed. Stansfield (1982), after developing a forty-eight cell, 4 x 4 x 3 matrix, said ". . . test authors must make choices as to which cells will be included and which will be left out" (p. 237).

Figure 4

		Selected Response	Supplied Response
Printed Stimuli	2D Behavior	Select, from 4 different pictures, the one which is like another picture previously exposed for only 3 secs. Select, from 4 different pictures, the one which has a functional relationship to another picture. Select, from 4 different pictures, the one which has a relationship to a fifth picture which is analogous to the relationship between two other separate pictures. Locate, in 30 seconds, as many designated, pictured objects as possible which have been embedded or camouflaged within each of 4 designs.	
	3D Behavior	Rearrange three or more pictured forms according to a pre-arranged, model sequence after a 5-second exposure of the model to be copied.	Pantomime how each of ten pictured common objects is used.
Oral Stimuli	2D Behavior		
	3D Behavior	Agree or disagree, by word or gesture, with assertions that particular named animals, plants, people or things respond in certain ways.	Complete, with one spoken word in each case, short sentences presented orally involving analogies some of which incorporate opposites. Repeat, orally, spoken digit sequences involving from 2 to 9 digits, presented at a rate of two digits per second. Describe, orally, each of five real common objects in terms of multiple characteristics such as name, color, shape, composition, function, parts, numerosity, simile, metaphor, or relationship to other persons, places or things. Complete short spoken and illustrated sentences, orally, with one word which, in each case, follows a rule of grammatical usage, number, tense, or a comparative-superlative relationship. Say the complete word for each of 30, 2 to 5 syllable words when presented orally with some syllables missing. Pronounce real and nonsense words composed of two to seven sounds when presented orally in separated phonetic units.

Figure 5

		Selected Response	Supplied Response
Printed Stimuli	2D Behavior	Choose, among given printed statements, the best response to questions about the content in each of two written paragraphs. Identify which of four printed words best completes each of ten statements. Differentiate which of three worlds, if any, is misspelled in each of five examples.	Complete, in writing, each of five sentences containing a missing word suggested by a pictorial prompt. Write-in an appropriate word or short phrase that completes each of five printed sentences in answer to a written question. Answer, in writing, five printed questions about a given picture.
	3D Behavior		
Oral Stimuli	2D Behavior	Pick, from three printed words, the one which is spoken in each of five examples. Select a printed YES or NO answer to each of five spoken questions about everyday objects or events.	Write, from dictation, each of five examples containing from one to five words. Respond, in writing, to five simple, spoken imperatives involving common, everyday, words and relationships.
	3D Behavior	Verbalize which step on a four-point scale represents personal competence in four types of communication described orally.	Reply, orally, to five simple spoken questions about everyday events. Respond, orally, to questions about simple things or actions which are shown.

Figure 6

Test Reviewed	Printed stimuli					Oral stimuli			
	2D, selected response	2D, supplied response	3D, selected response	3D, supplied response		2D, selected response	2D, supplied response	3D, selected response	3D, supplied response
Short Tests of Linguistic Skills (Frederickson, 1976)	★	★				★	★	★	★
Illinois Test of Psycholinguistic Abilities (Kirk, et al., 1968)	★		★	★				★	★
Spanish-English Dominance Assessment Scale (Spolsky, 1972)								★	★
Intermediate Reading Comprehension Test in Modern Chinese (Chou, et al., 1964)	★								
Pictorial Auditory Comprehension Test (Carroll and Ho, 1959)						★			
Flexibility Test to Measure Dominance in Spanish-English Bilinguals (Keller, 1974)	★								
Ambiguous Verbal Stimulus Test to Measure Dominance in Spanish-English Bilinguals (Keller, 1974)				★					
Navajo-English Language Dominance Interview (Spolsky, 1974)								★	★
Test of Language Judgement (Kahn, 1971)	★	★							
Test of Grammatically Correct Spanish and English (de Mestan, 1976)		★							★
Ott Test of Oral Language: English and Spanish (Ott, 1970)				★					★
French Achievement Test: Language Arts, K through 5 (Delahoussaye, 1973)						★			★
French FLES Test. (Barry, et al., no date)	★					★			
West Hartford Public Schools Spanish FLES Test (West Hartford, 1976)	★					★			
English as a Second Language Assessment Battery (ESLAB) (Rivera & Lombardo, 1979)	★	★				★	★		★
Language Dominance Survey: Grades K-1 (Language Dominance & Survey, 1974)				★		★			
Language Dominance Survey: Grades 2-4	★							★	★
Language Dominance Survey: Grades 5-7	★						★	★	★
Language Dominance Survey: Grades 8-12	★						★	★	★

*Discrete classification within the rubric is complicated in these two instances; the primary stimuli are printed pictures, but the directions on how to respond to them are presented orally.

Deciding How Many Cells of the Cubic Rubric Should Be Filled

Since the STLS has been shown to assess 75% of the types of performances outlined in the proposed cubic rubric, some might say this is sufficient. Certainly, it is the most complete of those analyzed here. With the addition of two more sets of tasks, it would give 100% coverage. While this might be an ideal goal to be reached, there are empirical studies that remain to be done before a definitive statement can be made. There is some evidence that if we rely on competence measures in one cell to predict competence in another without directly measuring it, gross errors can be made. Moss, Cole and Khampaljkit (1982) found a correlation of only .12 between a 2D, Selected Response, Printed Stimuli type test and a 2D, Supplied Response, Printed Stimuli Test. In more traditional terms, this is a correlation study between scores on an essay test and scores on a multiple-choice language usage test. This low correlation was for fourth graders. At the seventh grade, the correlation went to .39 and to .47 at the tenth grade.

The ease with which a multiple-choice language test can be administered and scored is frequently the basis for a decision that only this 2D, Selected Response, Printed Stimuli type examination be given. However, to make sure that a just judgment is made of a person's language competence, more, much more is needed. At the moment it might be best to aim at constructing a test which touches most if not all eight cells of the cubic rubric.

Conclusion

Several classification schemes, taxonomies or rubrics have been proposed for analyzing the adequacy of language assessment procedures. The proposed cubic rubric presented in this paper is not meant to supplant, but rather to augment, those other schemes which deal with linguistic specifics such as phonology, syntax, lexicon, pragmatics, and sociolinguistic variables. When a test is placed against the eight-fold matrix suggested here, a few more questions about adequacy of the test can be answered. Since this 2 x 2 x 2 classification scheme is based only upon a logical analysis of test performances, much empirical work remains to be done. Some data are available on correlations between a couple of the facets of the cubic rubric. Other intercorrelations are as yet unknown. As for integrating the cubic rubric with other classification schemes, it seems that the most likely application to start with is in the area of expanding and clarifying the traditional variables of listening, speaking, reading and writing.

References

Barry, D.; Mendall, L.; Singer, L. & Werfelman, Mr. *French FLES Test*. Princeton, New Jersey: ETS Test Collection, (no date).

Bloom, B. (Ed.). *Taxonomy of Educational Objectives: Cognitive Domain*. New York: Longmans, Green and Co., 1956.

Carroll, J. & Ho, W. *Pictorial Auditory Comprehension Test*. New York: Modern Language Association, 1959.

Chou, K.; de Francis, J.; Kao, Y.; Mills, H.; Pian, R. & Wrenn, J. *Intermediate Reading Comprehension Test in Modern Chinese*. New York: Modern Language Association, 1964.

Delahoussaye, H. *French Achievement Test: Language Arts, Kindergarten Through Grade 5*. Princeton, New Jersey: ETS Test Collection, 1973.

deMestan, M. (Ed.). *Test of Grammatically Correct Spanish and English*. Princeton, New Jersey: ETS Test Collection, 1976.

ETS Test Collection: Tests in Microfiche. Princeton, New Jersey: Educational Testing Service, 1975 through 1981.

Foster, H. *Ribbin', Jivin' and Playin' the Dozens*. Cambridge, Massachusetts: Ballinger Publishing Co., 1974.

Frederickson, C. *Short Tests of Linguistic Skills*. Chicago: Board of Education, Department of Research and Evaluation, 1976.

Kahn, A. *Test of Language Judgment*. Princeton, New Jersey: ETS Test Collection, 1971.

Keller, G. *Flexibility Test to Measure Language Dominance in Spanish-English Bilinguals*. Princeton, New Jersey: ETS Test Collection, 1974.

Keller, G. *Ambiguous Verbal Stimuli Test to Measure Dominance in Spanish-English Bilinguals*. Princeton, New Jersey: ETS Test Collection, 1974.

Kirk, S.; McCarthy, J. & Kirk, W. *Illinois Test of Psycholinguistic Abilities*. Urbana, Illinois: University of Illinois Press, 1968.

Krathwohl, D.; Bloom, B. & Masia, B. *Taxonomy of Educational Objectives: Affective Domain*. New York: David McKay Company, Inc., 1964.

Language Dominance Survey: Kindergarten Through Grade 7. San Bernadino, California: San Bernadino Unified School District, 1974.

Mitchell, B. *A Glossary of Measurement Terms*. New York: The Psychological Corporation, (no date).

Moss, P., Cole, N. & Khampalikit, C. "A Comparison of Procedures to Assess Written Language Skills at Grades 4, 7 and 10." *Journal of Educational Measurement*, 19, 1982, 37-47.

Ott, E. *Ott Test of Oral Language: English and Spanish*. Austin, Texas: Southwest Educational Development Corporation, 1970.

Pucel, D. & Knaak, W. *Individualizing Vocational and Technical Instruction*. Columbus, Ohio: Charles E. Merrill Publishing Company, 1975.

Rivera, C. & Lombardo, M. *English as a Second Language Assessment Battery (ESLAB)*. Boston: Boston University Press, 1979.

Spolsky, B. *Spanish-English Dominance Assessment Scale*. Princeton, New Jersey: ETS Test Collection, 1972.

Spolsky, B. & Holm, W. *Navajo-English Language Dominance Interview*. Princeton, New Jersey: ETS Test Collection, 1975.

Standards for Educational & Psychological Tests. Washington, D.C.: American Psychological Association, 1974.

Stansfield, C. "The Assessment of Language Proficiency in Bilingual Children: An Analysis of Theories and Instrumentation." In R. Padilla (Ed.) *Bilingual Education Technology*. Ypsilanti, Michigan: Eastern Michigan University Press, 1982.

West Hartford Public Schools Spanish FLES Test. Princeton, New Jersey: ETS Test Collection, 1976.

FIFTH INTERNATIONAL SYMPOSIUM ON EDUCATIONAL TESTING
UNIVERSITY OF STIRLING, SCOTLAND, JULY 2, 1982

Protase E. Woodford
Educational Testing Service

At ETS, since our founding in 1947, we have developed foreign language tests. It was not until the 1960's, however, that we attempted to assess speaking ability. The Modern Language Association of America received a grant from the U.S. Office of Education for development of listening, speaking, reading and writing tests in French, German, Italian, Russian and Spanish. These monies become available through the NDEA. Sputnik had frightened legislators into providing funds for the improvement of instruction in mathematics, science and foreign languages. Foreign language teaching methodologies and materials underwent a radical transformation. The model for much of this change was the old Army Language School in Monterey, California (now DLI). Among the consultants for the speaking test were representatives of the Foreign Service Institute of the U.S. Department of State.

FSI, in the late 1950's, was faced with the problem of both describing and evaluating the face-to-face communication skills of foreign services officers in dozens of languages. It did not enjoy the internal accountability system of the academic foreign language teaching community.

The quality of its program was judged by ambassadors, spies, heads of missions and others. If the "product" was not of sufficient quality, congress would soon hear of it and the marvelous money well would soon dry up.

The FSI Proficiency Rating Scale

Level 1

Able to satisfy routine travel needs and minimum courtesy requirements. Can ask and answer questions on topics very familiar to him within the scope of his very limited language experience; can understand simple questions and statements, allowing for slowed speech, repetition or paraphrase; speaking vocabulary inadequate to express anything but the most elementary needs; errors in pronunciation and grammar are frequent, but can be understood by a native speaker used to dealing with foreigners attempting to speak his language. While elementary needs vary considerably from individual to individual, any person at level 1 should be able to order a simple meal, ask for shelter or lodging, ask and give simple directions, make purchases, and tell time.

Level 2

Able to satisfy routine social demands and limited work requirements. Can handle with confidence but not with facility most social situations including introductions and casual conversations about current events, as well as work, family, and autobiographical information; can handle limited work requirements, needing help in handling any complications or difficulties; can get the gist of most conversations on nontechnical subjects (i.e. topics which require no specialized knowledge) and has a speaking vocabulary sufficient to express himself simply with some circumlocutions; accent, though often quite faulty, is intelligible; can usually handle elementary constructions quite accurately, but does not have thorough or confident control of the grammar.

Level 3

Able to speak the language with sufficient structural accuracy and vocabulary to participate effectively in most formal and informal conversations on practical, social, and professional topics. Can discuss particular interests and special fields of competence with reasonable ease; comprehension is quite complete for a normal rate of speech; vocabulary is broad enough that he rarely has to grope for a word; accent may be obviously foreign; control of grammar good; errors never interfere with understanding and rarely disturb the native speaker.

Level 4

Able to use the language fluently and accurately on all levels normally pertinent to professional needs. Can understand and participate in any conversation within the range of his experience with a high degree of fluency and precision of vocabulary; would rarely be taken for a native speaker, but can respond appropriately even in unfamiliar situations; errors of pronunciation and grammar quite rare; can handle informal interpreting from and into the language.

Level 5

Speaking proficiency equivalent to that of an educated native speaker. Has complete fluency in the language such that his speech on all levels is fully accepted by educated native speakers in all of its features, including breadth of vocabulary and idiom, colloquialisms, and pertinent cultural references.

The foreign language tests that we had been developing at ETS over the years were limited to skills of reception — listening comprehension and reading. They were objective tests — multiple choice — and very reliable.

We know that productive skills correlate highly with their corresponding receptive skills, speaking-listening, reading-writing, for relatively large groups, but there are always exceptions. A tongueless examinee might understand Portuguese perfectly.

The objective, multiple choice tests might be extremely reliable, but they were obviously not valid measures of face-to-face communication skills. The Foreign Sercive would sacrifice high reliability for validity. The first task was to develop a scale or series of descriptors. The FSI Scale ranged from 0 — no functional ability in the language, to 5 — ability equivalent to that of an educated native speaker.

The problem with a numerical scale, particularly for pedagogs, is the tendency to consider the points equidistant. A 3 should represent the same improvement over a 2 as a 2 over a 1. A cone, however, illustrates better than a line, the properties of the scale.

The Foreign Service developed expectations for attainment of different levels, by language, according to hours of instruction.

Once the Foreign Service had the scale in place, the next task was to develop a procedure whereby people could be put on the scale. The LPI, or language proficiency interview, was the result. The LPI is, in a sense, an endless series of parallel tests, each unique to the examinee. The interview lasts from 3 to 20 minutes, approximately.

Structure of the Interview

Every oral interview follows the same general structure. This general structure guides the interviewer by directing his or her attention to certain mandatory aspects of the test. An interview may be divided into *four phases*: warm-up, level check, probes, and wind-up. The level check and the probes take more time than the warm-up and the wind-up. At the very lowest levels, the limitations of the candidate's language may be such that the four phases will be indistinguishable from each other. At the very highest levels, neither warm-up nor wind-up will be necessary unless the candidate has not been speaking the language recently.

The Warm-up. The warm-up consists of social amenities and simple conversation at a level that is very easy for the candidate. (At the lowest levels, this may not be possible.) There are three purposes to this phase of the interview: (1) putting the candidate at ease: (2) reacquainting the candidate with the language if necessary: and (3) giving the interviewer a preliminary indication of the candidate's level. For candidates, the main purpose of the warm-up is to put them at ease with the testing situation and to reintroduce them to the language. The length of the warm-up will depend on the circumstances; candidates who have not spoken the language for some time may need to get back into it gradually, while others may themselves immediately shift the conversation to a higher level. Testers should never skip this phase, but they may shorten it considerably if the candidate does not seem to need it.

One good way to begin the warm-up is for the tester to introduce himself or herself to the candidate in the target language. Since introductions are usually learned early in foreign language classes, it is easy for most candidates to respond, opening the way for further conversational exchanges.

For the interviewer, the warm-up serves the important function of giving a preliminary indication of the candidate's level. This preliminary indication must be confirmed because about a third of candidates answer questions at the level and in the style in which they are asked. The best approach is for a tester to assume that the preliminary indication is to be checked in the next phase, the level check. In fact, the rest of the interview will be devoted to ascertaining whether this preliminary indication is accurate or not.

The Level Check. The purpose of this phase is to find the highest level at which the candidate can sustain a speaking performance. To find the level, the interviewer must test the breadth and depth of the candidate's ability in the language.

How fluent is the candidate? How well does he or she pronounce the language? How accurate is the grammar? How wide is the vocabulary? How correct is the syntax? How native is the expression of ideas and concepts in the language?

Sometimes the level indication given by the warm-up is misleading, and the interviewer can begin the level check too low or too high. If the test begins at too low a level, the interviewer can simply raise the level of the questions and begin the level check over again. If the test begins at too high a level (This problem can be caused by either the tester's questions or the candidate's answers), the interviewer must bring the level down. Starting at too high a level is to be avoided, since bringing the level of an interview down is difficult to do without giving the candidate a sense of failure.

In the level check, testers should check a number of topics (both interest and noninterest areas) to see if the candidate can perform consistently at the level in question. Can the candidate accomplish the functions with suitable content and accuracy? When the candidate successfully passes the level check, his or her performance provides a *floor* to the rating. The next phase aims at finding the *ceiling*.

The Probes. The purpose of this phase is to make sure that the level the interviewer has been checking is the candidate's highest sustained level. To probe, the tester should take the candidate above the previous level *several times* in different ways: an involved question, a

situation, a conversation between two testers into which the candidate is then drawn, etc. The interviewer may also want the candidate to ask some questions. If this phase has been successful, no candidate should leave the testing room feeling that he or she has not been tested to the limit of his or her ability.

This phase is purposely in the plural because there should be several probes, at least three or four. Probes should furnish clear examples of linguistic breakdown. Sometimes the candidate actually tells the interviewer that the limit has been reached by saying, "I don't know how to say that in your language," or "I know what I want to say but I can't say it." In other cases, a sharp drop in fluency, a sudden groping for words, or a dramatic increase in grammatical errors give evidence of the linguistic breakdown.

If the interviewer has carried out the level check at too low a level, the candidate will probably be able to respond to the probes consistently well. If this happens, then the interviewer must recommence the process of level check and probes and continue until the ceiling of the candidate's proficiency is found.

While the level check gives evidence of what candidates can do, the probes show candidates what they cannot do. Without this phase of the interview, candidates may appear to be more proficient than they really are. The probes allow a tester to explain why a candidate's speech is not at a higher level, providing diagnostic information with specific examples.

Experienced testers learn how to interweave the level check and the probes, so that the candidate is allowed to return to a level where performance can be sustained before being asked another higher-level question.

The Wind-up. The purpose of this phase is to leave candidates with a feeling of accomplishment after stretching their speaking ability to the limit. It is also the tester's last chance to check out any aspect of the candidate's speaking ability that may still be unclear. Normally, the wind-up should return to the highest level that the candidate was able to sustain during the interview. It may even be helpful, particularly at the lowest levels, to end the test by returning briefly to a topic discussed previously. It is, of course, always appropriate to close by thanking candidates for the interview.

ETS was asked by the Peace Corps in 1970 to take over Peace Corps Foreign Language Testing from FSI because the Peace Corps was overtaxing FSI resources. ETS was requested to consider alternatives to the FSI interview. After two years of exploration, we decided that what they had, with some refinements, suited them very well.

The Common Yardstick

As an outgrowth of the long-standing involvement of Educational Testing Service's language staff with the Foreign Service Institute (FSI) proficiency rating scale and in response to expressions of interest from the British Council, the English Speaking Union, and German and Japanese agencies, ETS, in June 1979, sponsored a small conference to discuss the possibility and desirability of establishing a "common yardstick" (or yardsticks) to describe performance in one or more language skills.

At the conference, descriptive scales of language proficiency developed in various countries and by international agencies such as the Council of Europe were distributed. Both theoretical and practical issues in the development and use of a single set of descriptive scales on an international basis were discussed. Background papers developed from prior work in this area were presented by ETS and by British Council staff, and other participants contributed information from the perspective of their own organizations.

There was unanimous agreement among the participants that development of descriptive scales for all language skill areas should be attempted. It was also recommended that a small working group from among the conference participants be designated to begin work on the scale development.

In November, 1979, a working group consisting of John Clark and Protase Woodford of ETS; Brendan J. Carroll, British Council; David Hicks, English Speaking Union; and Anthony Fitzpatrick, Deutscher Volkshockschul Verband met in London. The outcome of this meeting was the preparation, in rough draft form, of descriptive scales for oral interaction and writing and a general outlining of scale characteristics for listening comprehension and reading. Following the November meeting, draft scale descriptions for "passive" listening comprehension (excluding oral interaction) and for reading comprehension were prepared.

On the basis of these initial activities and the positive general response obtained, ETS requested and received funding from the foreign language area studies research program (U.S.O.E./D.E.) for the current project to continue work on a common metric for language proficiency.

Brief Description of Project Tasks

Proposed further activities for this project included the following three tasks:

(1) Distribution of the draft scales to recognized foreign language measurement specialists for their critique, commentary, and any suggestions for revision.

(2) Convening of a small group of measurement specialists to synthesize the recommendations of the reviewers and collaborate in the revision of the scales. The senior British member of the international work group would be invited to attend this meeting to provide a summary of similar inputs by measurement specialists in Europe.

(3) Assuming a generally positive outcome for activities (1) and (2), presentation and discussion of the language assessment scales and recommendations for future development activities to implement the use of these scales to executive officers of foreign language associations, government agency representatives, and representatives of the international business community.

Project Outcomes

Task (1): *Review by Measurement Specialists*

In December, 1980, selected foreign language measurement experts were sent project information and draft scales for their review and comment. Individuals requested to participate in this review included:

Dr. Lyle Bachman
University of Illinois

Dr. Michael Canale
Ontario Institute of Studies in Education

Dr. James Child
National Security Agency

Dr. Ray Clifford
Central Intelligence Agency and Defense Language Institute

Dr. James R. Frith
Foreign Service Institute

Dr. Helen Jorstad
University of Minnesota

Dr. Pardee Lowe
Central Intelligence Agency

Dr. Adrian Palmer
University of Utah

Dr. Howard Nostrand
University of Washington

Dr. G. Richard Tucker
Center for Applied Linguistics

Dr. Barbara Freed
University of Pennsylvania

Reviewers were asked to give (1) their appraisal of the overall merit of the project from both psychometric and practical standpoints, and (2) specific suggestions for the revisions of the draft scales with the rationale for such revision.

Task (2): *Measurement Specialist Working Group Meeting*

From the original group of reviewers listed above, a smaller working group was selected to participate in an intensive two-day meeting at ETS to consider the comments of all reviewers and to collaborate with ETS staff on the revision of each of the four language scales. The following individuals participated in the February 24-25, 1981 meeting at ETS:

Protase Woodford, Project Director
John Clark, Principal Investigator
Judith Liskin-Gasparro, Associate Examiner
Marianne Adams, Foreign Service Institute
Lyle Bachman, University of Illinois
Michael Canale, Ontario Institute for Studies in Education
James Child, National Security Agency
Ray Clifford, Defense Language Institute
Barbara Freed, University of Pennsylvania
Pardee Lowe, Central Intelligence Agency
Howard Nostrand, University of Washington

At the February 24-25 meeting, the participants discussed in detail the following issues:

1. **Skills Represented by Scales**

It was suggested that the traditional four skills — listening, speaking, reading and writing — be modified to include reading, writing and "pure" listening comprehension as discrete, measurable skills, and that "oral interaction" be used to replace "speaking" because of the listening skill required in real-life speech contexts.

2. **The Number of Scale Divisions**

The scales reviewed in connection with the project came from a variety of U.S., British, and European sources. The 0-5 Foreign Service scale was the most familiar to the participants. The utility of the various scales was discussed as well as a proposed 8-level scale (0-7) for oral interaction.

3. **FSI Scale/Oral Interaction Scale Comparison**

The relationship of the proposed 0-7 oral interaction scale to the FSI 0-5 (with +'s) scale was considered.

4. **Scale Progression**

Participants were also asked to consider whether the proposed scales provided for a smooth progression from one level to the next and whether any pair of level descriptions was too close to allow for a meaningful distinction between levels.

5. **Intra-level Consistency**

Participants were asked to consider whether the descriptive statements within each level would apply to most persons within the ability group, i.e. to make sure that there would be no descriptions of tasks or behaviors "too easy" or "too difficult" for people within the level.

6. **Interscale Comparability**

Discussion centered on the degree to which the scales for the four skills were consistent with regard to detail of description.

7. **Individual Scale Aspects**

Participants were asked to rank each of the four draft scales on the following criteria:

A. "understandability,"

B. "real-life referencing,"

C. ease and straightforwardness of use for rating examinee performance,

D. priority for development.

The participants were also asked to consider the scales presented to them in light of the "ideal" scale, that is, one that would include all of the features of oral interaction that they considered important.

As the meeting developed, it became apparent that the task at hand was extremely complex. Consequently, it was decided that a major part of the effort would be devoted to oral interaction.

The following were major outcomes of the meeting:

(1) A commitment to some form of the 0-5 government scale. The deliberations of the group demonstrated that all of the members were in some sense basing their reactions to the draft scales on the relationships of these newer scales to the government scale developed by the FSI. Since the government scale is relatively better known and since it has a long and respected history, it seemed most reasonable to begin with it as a base, making adjustments to it that would not alter the accepted understanding of the significance of Level 1 proficiency, Level 2 proficiency, etc.

(2) The realization that no scale currently in existence or under consideration does as complete a job of evaluating oral proficiency as the participants in the February meeting would like. Particular concern was focused on such aspects of language ability as register, cultural sensitivity, and, in general, the relationship between linguistic ability and the larger area of interpersonal communication. While these issues arise mostly at the upper proficiency levels, there are some languages for which they emerge as low on the government scale as Level 2. Time was also spent discussing and coming to a common understanding of the term "fluency."

(3) The decision that further work is most essential at the 0-2 range. This is the area in which most second-language speakers can expect to fall after taking advantage of the range of academic courses and extracurricular activities usually offered in secondary schools and colleges. Level 3 proficiency is usually attained only after extended residence in a country where the target language is spoken and/or through intensive or immersion-type language study. It was recommended that Levels 0, 1, and perhaps 2 be further subdivided to provide finer distinctions.

Given the outcomes of the February meeting, it became apparent that further refinement of the scale descriptions was needed before proceeding to the expected next step of the project, the convening of a meeting with executive officers of foreign language associations, government agency representatives, and representatives of the international business community. The intended focus of a meeting with these "user groups" was planned to be a presentation of the revised scales and a discussion of whether and to what extent the scales met their specific needs in the area of language proficiency evaluation.

Since the group at the February 24-25 meeting had recommended that further work be done on the lower levels of the oral interaction scale and, further, that work on the other skills be postponed while efforts were focused on the oral interaction scale, it was decided not to hold the meeting for language association officers, government agency representatives, and representatives of the international business community as planned. Instead, further work was done at ETS on the expansion of the lower end of the oral interaction scale.

Task (3): *Final Scale Revisions*

On October 6, 1981, the final meeting of the Common Yardstick Project was held at the CIA Language School, hosted by Dr. Pardee Lowe. The purpose of the meeting was to discuss and, if possible, reach consensus on the revisions to the oral interaction scale and to agree on future plans. The participants at the meeting were as follows: Protase E. Woodford, Judith E. Liskin-Gasparro, and Ihor Vynnytsky from ETS; Professor Barbara Freed, Assistant Dean for Languages, University of Pennsylvania; Dr. Ray Clifford, Academic Dean, Defense Language Institute; Dr. Pardee Lowe, CIA Language School; Dr. Yvonne Escola, teacher of French in Montgomery County, Maryland and program officer, National Endowment for the Humanities; and Dr. John L.D. Clark, Center for Applied Linguistics.

The discussion at the meeting focused on the needs of the government language schools and the academic community in the area of the evaluation of oral proficiency. The expanded lower end of the government scale was presented for discussion, and both Dr. Clifford and Dr. Escola agreed that it would provide valuable information for students as well as teachers of language programs. At the CIA Language School, language testers often offer finer descriptive distinctions, beyond the official ratings, in their evaluations of student examinees. Dr. Lowe reported that the informal descriptions were very similar to the expanded "intra-level" descriptions prepared by ETS. This congruence in the independently developed descriptions was an encouraging sign, and it was agreed that further work in this area would benefit from the experience of the CIA Language School.

The group discovered a second area of congruence between academic and government language assessment needs in the discussion of the value of a bilevel system of oral proficiency assessment. It was agreed that for other than a few very specialized uses, there is little need to discriminate between levels of proficiency at the 3 to 5 range. Most academic and professional needs will be satisfied by proficiency at the 3 level or lower, so it is at this lower end of the scale that most attention needs to be focused. The group agreed that the following labels for ranges of proficiency represent realistic descriptions:

0 elementary
1 intermediate
2 advanced
3-5 superior

In summary, the major outcomes of the meeting were as follows:

(1) Consensus on the usefulness of the expanded definitions at Levels 0 and 1. It was agreed that these descriptions were definitely "on the right track" and that further work on them would be a logical and valuable next step.

(2) Agreement on the usefulness of a bilevel system according to which further development efforts would be concentrated in the 0-3 range. Individuals above Level 3 would be designated as "superior." If a precise level were desired, it could be provided via the traditional face-to-face interview.

(3) Agreement that coordination of efforts among the various agencies concerned with language proficiency testing is a major concern that must be addressed. As of this date, a major cooperative venture, stemming from the Common Yardstick Project, has been launched that includes ACTFL, ETS, and the CIA Language School.

In addition, an invitational conference on language proficiency testing was hosted by Ray Clifford and the Defense Language Institute November 30-December 1, 1981 in order to discuss the current needs and projects of agencies inside and outside the government and to decide on areas of future development.

Summary and Further Planned Development Work

Although the final stage of the project, i.e. the meeting with representatives of language associations, government agencies, and the international business community did not take place as proposed, the new direction undertaken by the project will serve to build a stronger foundation to serve these constituencies better in the long run. The contributions of the language professionals and linguists were valuable in defining areas of strength and weakness in the existing scales, and especially in recommending that new efforts be based on the government definitions and scale. The decision to concentrate on the lower-end of the oral interaction scale resulted in the creation of intermediate working definitions. For several years, language professionals in academe have recognized that the absence of these expanded descriptions has severely limited the applicability of the government oral interaction scale to college and high school students.

Further development work, stemming from the Common Yardstick project, is already underway by ACTFL, ETS, and the CIA Language School. After the October 6 meeting, Mr. Woodford turned over to ACTFL the expanded descriptions of oral proficiency for level 0 and 1. ACTFL, which is working on the development of proficiency levels as goals of instruction under a grant from the International Research and Studies of the U.S. Department of Education, in turn, asked Dr. Lowe of the CIA Language School to investigate the validity and accuracy of the ETS descriptions. Dr. Lowe, assisted by funds and professional collaboration from ETS, designed a research project to determine (1) whether the expanded intra-level descriptions correspond to real-life language use and (2) whether the intra-level descriptions and independent raters will rank a group of tapes known to be within a given level in the same order. The scale with the expanded lower end will be taught to college faculty members in Spanish and French at the workshop sponsored by ACTFL and conducted by ETS under a grant to ACTFL by the U.S. Department of Education.

For the long range, development work similar to that which has been accomplished for oral interaction might be undertaken for the other language skills. Although further development work beyond the scale-definition and review stage would require additional financial support and would, also, of course, depend on the psychometric appropriateness and anticipated practical utility of the final scale descriptions, a fairly large-scale test development/validation project could be envisioned as a possible outcome of the initial work. This larger study could include each of the following activities:

(1) Development of comprehensive measures of language skills encompassing and operationally defining the descriptive scales. These would be very exhaustive and lengthy direct measures of each of the skills in question, requiring perhaps two full days of testing on the part of each examinee. It is recognized that these criterion tests would not be practical for regular measurement purposes, but would be used as comprehensive "benchmark" instruments exemplifying the scale descriptions and against which presently available, more easily administered tests (or smaller-scope tests yet to be developed) could be compared and validated.

(2) Development of validation measures external to both the large-scale "benchmark" tests and any smaller-scope tests. These external measures would be expected to include both examinee self-appraisal and "second-party" (e.g., classroom teacher, work supervisor) evaluation of the examinee's proficiency in the language skill areas at issue. These evaluations could take the form both of (a) direct utilization of the common yardstick scales (i.e., examinees and second-party observers would be asked to rate the performance vis-a-vis the common yardstick descriptors) and (b) use of more detailed and more "atomistic" descriptions of particular, language-use functions (e.g., "say the days of the week," "buy clothes in a department store," "talk about your favorite hobby at some length, using appropriate vocabulary"), which would be rated on a dichotomous (can do/cannot do) basis.

(3) Large scale administration of the comprehensive "benchmark" measures, smaller-scope measures, and external criterion measures to a large and varied group of examinees, for purposes of both construct/concurrent validation of the instruments in question and establishment of equating data relating examinee performance on the smaller scope tests to both the "benchmark" test results and to the common yardstick descriptors.

The exact nature and operational details of the activities outlined in 1-3 above, would, of course, have to be spelled out much more comprehensively at a later date. The intent at this point is simply to give a general overview of the kinds of development work that would seem to be logical and, we hope, practically useful in extensions of the initial development of the common yardstick descriptors.

Conclusion

At the time that the current study was proposed, the idea of a "common yardstick" or uniform descriptors of language proficiency was being considered only within a restricted population of measurement specialists and government connected linguists. The "yardstick" activities themselves and the reports on the yardstick to major foreign language education constituencies* have created extraordinary interest in the project across all academic levels. It is obvious now that the original scope of this project was far too broad. A result of the deliberations at the February meeting of the working group was a narrowed focus on a scale for *one* of the skill areas, the one considered of highest priority — oral interaction. The well-known and respected scale used by the federal government has seen limited use in the academic context primarily because if provides too little discrimination at the lower end 0.0 to 2.0. It is precisely at the lower end of the scale where there is greatest need for evaluation of language skills in schools and colleges.

The proposed expanded scale is a product of the working group's efforts subsequent to the February 1981 meeting and during the October 1981 meeting in Arlington, Virginia to refine and plan next steps.

The work accomplished through this project has served and will continue to serve a number of related projects.

The working group members are actively involved in continued dissemination of the draft scale to various foreign language constituencies. The expanded oral interaction scale is currently undergoing validation under an ACTFL-sponsored project and will — if proved valid — likely become the "yardstick" for describing the ability of American students to function in a real-life communication situation. Further work on the "yardstick" including further development and refinement of the existing draft scales for "pure" listening comprehension, reading and writing is planned. Support will be sought from a variety of sources. Training programs for high school and college foreign language teachers are scheduled for 1982. These training programs will focus on the evaluation of students' ability to understand and speak in a foreign language in a real-life context. The scale considered for use in the expanded, revised oral-interaction scale developed under the current Common Yardstick project. Among the sponsors of the training are:

The American Council on the Teaching of Foreign Languages (ACTFL),
The Northeast Conference on the Training of Foreign Languages,
Educational Testing Service,
Vassar College.

*Clark, Freed, Liskin-Gasparro, Lowe, Woodford have reported on the "yardstick" to such groups as Southern Conference on Language Teaching, Modern Language Association of America, Pennsylvania MLA, Flordia Foreign Language Teacher's Association.

BEYOND MEASUREMENT: USING EVALUATION TO IMPROVE FOREIGN LANGUAGE EDUCATION

Phyllis L. Hagel Jacobson

Present Crisis of Confidence

Background. Foreign language programs have increasingly come under attack in recent years. The following statements indicate the gravity of the present situation facing the profession: "Americans' incompetence in foreign languages is nothing short of scandalous, and it is becoming worse"; "Our schools graduate a large majority of students whose knowledge and vision stops at the American shoreline, whose approach to international affairs is provincial, and whose heads have been filled with astonishing misinformation"; and "All of this underscores that, crucial to any improvement of an intolerable situation, there is an urgent need for better trained teachers and for extensive retraining of those already serving in the nation's classrooms, particularly in view of widespread expert agreement that the decline of foreign language enrollments is in large measure a response to poor instruction."[1]

What conditions have prompted public commissions and other responsible critics to level these charges of lack of competence on the part of teachers and lack of achievement on the part of students? One answer is that the foreign language profession lacks convincing evidence on the effectiveness of foreign language programs as a whole, and of commonly used teaching methodologies in particular. The evidence of student achievement generally provided by foreign language educators is not expressed in the measurable, standardized terms commonly deemed acceptable to researchers, professional commissions and/or other measurement professionals. As Dale Lange noted in a letter to Representative Paul Simon of Illinois on 30 June 1980, ". . . the effectiveness of second language training at the college and university level is as questionable as that of secondary school language training. There is no evidence in the literature of second language learning of such effectiveness; it is only assumed by those in higher education."[2] Few methodologically rigorous studies have been done on the effectiveness of foreign language programs as a whole, or on the effectiveness of particular methodologies. Those studies which have appeared, such as those done in Albuquerque on intensive elementary school Spanish language instruction, and in Plattsburgh (NY) on the French immersion program at the elementary school level, generally pertain to high-concentration, intensive programs, rather than to the typical one-period-per-day elective type of program most commonly found in the nation's schools.

A second answer is that empirical evidence of the lack of student success and of teachers' abilities is readily obtainable by the public. One example of this type of evaluative assessment is cited by Paul Simon:

> It seemed obvious at the outset that the B.A. holder in foreign languages would be our prime candidate. However, it soon became apparent that most foreign language majors with a B.A. degree are not fluent enough . . . They could not sit at a negotiation table to discuss contractual arrangements or the purchase or sale of a given merchandise . . . The other day I had a student enroll in our MBA International Program. Her major was French. When I learned this, I asked her a very simple question in French, namely, that surely she must speak French well. The answer coming back to me in English was, 'I see I have to pursue my study of French some more.'"[3]

Incidents of this type provide critics with empirical evidence of lack of student achievement, and, by implication, of lack of teacher competence.

Some critics also accept hearsay as empirical "evidence" of the lack of demonstrable teacher competence, as illustrated by this correspondence between Paul Simon and Harry Reinert: "Responding to my letter of 19 August 1980, in which I objected that 'teachers at these levels do not need additional training...foreign language teachers have been trained and retrained for more than two decades,' Simon wrote in a personal letter to me on 25 August 1980: 'many foreign language professionals tell me that language teachers do need training in more innovative methods than those in which they were classically trained'" (Reinert, p.249).

Availability of Evaluative Information. What factors have constrained the availability of valid and convincing program evaluative information to forestall or counter these criticisms? First, according to the second chapter of Simon's book, programs have vacillated in their orientation in response to external social and political pressures. Lack of consistent goals has resulted in a lack of reliable program focus and hence a lack of

---

[1] *Strength Through Wisdom: A Critique of U.S. Capability*. Washington, D.C., 1980. pp. 5-8.

[2] Cited in Harry Reinert, "Caveat Emptor: The President's Commission." *Modern Language Journal* 65(1981), p.250.

[3] Paul Simon, *The Tongue-Tied American* (New York: Continuum, 1980), p.124.

standards/goals against which program implementation and student achievement can be measured.

Second, for many years, "evaluation" in the language field has been synonymous with "testing." As Omaggio states; "Evaluation is often thought of in the very narrow sense of 'tests' and 'grades'."[4] She also notes that teachers and administrators have not fully understood the broader focus of program evaluation beyond the measurement of achievement via testing: "TenBrink asked preservice and inservice teachers to answer the question, 'what comes to your mind when you hear the word 'evaluation'? The most frequently occurring responses were: Preservice teachers: test, grades, achievement, unfair, judgment; Inservice teachers: tests, measurements, grades, accountability, invasion of privacy" (p. 236). While the periodical literature carries numerous articles on focused testing strategies, few articles deal with applying broader program evaluation concepts to foregin language programs. Recently, however, monographs have begun to treat this topic.[5]

Even within the narrower domain of testing, there is still no consensus within the profession as to suitable, reliable and valid testing methods. For example, each of the suggested standard teacher proficiency measures in the field has drawbacks serious enough to make it unacceptable as a valid measure to a significant segment of the profession (Valette and Linder, p.202). As Clark points out, ". . . Even the FSI interview falls somewhat short of being an absolutely face-valid measure of communicative proficiency."[6]

Third, foreign language educators have not been in the forefront of the "back to basics" movement which has forced a closer look at the major objectives, the instructional sequences and the evaluation methodology of programs as a whole. Although there have been recent publications on this topic,[7] the profession as a whole has not organized to work either at the local or the state level to define the priority goals to be achieved within a program, even at a "minimum competency" level. Indeed, as perhaps a continuing reaction to events which occurred several decades ago, foreign language educators still resist efforts to develop national or state criteria for basic skills in second language education. As Reinert comments:

> Such unequivocal imperatives regarding pedagogical priorities in foreign language study should alert members of the profession to some possible undesirable consequences of the Commission's recommendations. The monolithic approach found in the report has the familiar ring of the strict audiolingualism of twenty years ago; veterans within the profession should be able to remember vividly the enormous pressure which was applied to assure national conformity to a single approach to teaching foreign languages (p. 250).

Concerted efforts have not been made to provide local or state needs assessment data which would allow professionals to determine themselves the objectives against which their programs could be measured.

Fourth, the profession has seen a proliferation of new programs and approaches (e.g., FLES, mini-courses, vocationally oriented courses, courses for "communicative competence," etc.) which have, in many cases, been hastily adopted and just as quickly discarded before reliable information could be obtained as to the worth and/or effectiveness of such programs and methods.

Fifth, most foreign language educators lack professional training in the area of evaluation. Few evaluators have the linguistic background which would allow them to work effectively with language educators, particularly in the crucial area of test development. The lack of trained personnel is reflected in the dearth of high-quality program evaluation studies. As Omaggio confirms, "Although there is an abundance of literature pertaining to program evaluation in general, the foreign language community has been strangely quiet about the problem, which leads one to suspect that evaluations are not being carried out with any regularity or in any systematic fashion" (p.240). Indeed, a survey conducted in 1978 under the auspices of ERIC-CAL on program evaluation in foreign language instruction concluded that "with some notable exceptions, there appears to be a low level of interest and activity in program evaluation of foreign language instruction in the United States, with the single general exception of the periodic (and too often pro forma) visits of the regional accrediation associations" (Omaggio, p. 243).

---

[4] Alice Omaggio, "Looking at the Results," *Building on Experience-Building for Success*, ed. June Phillips (Skokie, IL: National Textbook Co., 1979), p.237.

[5] For example, see Gilbert Jarvis & Shirley Adams, *Evaluating a Second Language Program* (Arlington, VA: Center for Applied Linquistics, 1979); and Rebecca Valette & Cathy Linder, "Measuring the Variables and Testing the Outcomes," *Building on Experience-Building for Success*, ed. June Phillips (Skokie, IL: National Textbook Co., 1979).

[6] John L.D. Clark, "Measurement Implications of Recent Trends in Foreign Language Teaching," *Foreign Language Education: A Reappraisal*, ed. Dale Lange (Skokie, IL: National Textbook Co., 1972), p. 223.

[7] For example, see Renate A. Schultz, "Back to Basics in the Foreign Language Classroom?" *Foreign Language Annals* 11 (1978), pp.647-55; the report on the Central States Conference, "Alternatives, Basics, and Competencies in Foreign Language Instruction," *Modern Language Journal* 63 (1979), pp. 364-66; and *Teaching the Basics in the Foreign Language Classroom: Options and Strategies*, ed. David P. Benseler (Skokie, IL: National Textbook Co., 1979).

Finally, the profession itself continues to be divided about the ultimate goals of language study. Some members hold to a rationale which rests mainly on aesthetic and humanistic grounds, a rationale influenced by university- and college-level educators for whom the study of literature is viewed as the ultimate purpose of study at the beginning language-learning level. Others, generally at the elementary and secondary school level, view language learning primarily as a means for communicating with and understanding other cultures. Still other educators view beginning language learning as a tool to be used to further international business and/or diplomacy. As Reinert observes:

> . . . if the Commission did give us everything the professional associations requested, this should properly be seen as an indication of the terrible price the profession must pay for its lack of unity and direction . . . Parochial concerns of the component members of JNCL have prevented it from speaking with authority. One result of this was that the Commission at its regional hearings received mixed and even competing messages from different segments of the foreign language profession. We must conclude, therefore, that the shortcomings with respect to foreign languages in the Commission's report are at least in part due to the profession's unwillingness to settle internal differences in order to advance the welfare of the profession as a whole (p. 252).[8]

These divisions within the profession further result in a core of educators who resist being held accountable for student achievement based on goals which they personally do not support.

Ultimate Effects. What are the ultimate effects of this lack of evaluative information on the profession and on foreign language programs? First and foremost, it has led to the downgrading of language study as a national priority in and of itself. As Reinert's analysis of the President's Commission report points out, ". . . Equally clear is that far from being the 'best thing that has ever happened to foreign languages,' the report's recommendations actually threaten the profession with loss of identity by reducing language learning to a minor supportive role within international studies" (p. 252).

It has also taken the impetus and the responsibility for professional investigation, evaluation and improvement out of the hands of the profession itself. House Bill HR 7580, "introduced in the House of Representatives on 13 June 1980, by Representative Paul Simon . . . would provide for a 'survey of the effectiveness of FL training in grades 7 through 12, and in community colleges, with an emphasis on . . . the desirability and costs of additional training for FL teachers in grades 7 through 12, and in community colleges' (Sect.3;2C)" (p.252). Furthermore, the President's Commission report proposes as one of its recommendations the creation and funding of a "National Criteria and Assessment Program" to "develop foreign language proficiency tests, and report on, monitor, and assess foreign language teaching in the U.S. (p. 13) . . . This assessment program should be administered through competitive contract by a nationally recognized professional association or agency with special competence in foreign language studies" (p.42). And, finally, the Commission further recommended the creation of an Advisory Council on Foreign Languages and International Studies in each state: "These councils should be comprised of approximately 10 persons, including teachers from all levels, representatives from business and labor and from major ethnic communities, lay people and at least one student . . . The Councils should advise the governor, the chief education officer and the state legislative body on Foreign Language and International studies in the state, and make recommendations towards their improvement" (p. 43). It is significant of the current state of affairs that this policy-recommending body has *no* foreign language teacher representation specified by name; nor must the "National Criteria and Assessment Program" necessarily involve a recognized foreign language organization.[9]

Finally, it has led to the further erosion of the public credibility both of the profession and of language programs in general. Even as distinguished an educator as James B. Conant has suggested that anything less than four years of a foreign language before college is worse than nothing (Simon, p. 94). Simon also quotes Wall Street Journalist Vermont Royster as observing that "students going to college are deficient in language skills. Some still are when they leave. That is indeed appalling, and its long past time we raised some hell about it" (p. 97).

Utility of Evaluation for Foreign Language Education

How can evaluation aid the profession in countering these negative perceptions and in returning control of the future directions of foreign language study to the hands of professional educators?

The term "evaluation" has had many definitions within the growing body of evaluation literature.[10] Current emphasis in the field focuses on evaluation as a process

---

[8] See also David P. Benseler, "The American Language Association: Toward New Strength, Visibility, and Effectiveness as a Profession," *Our Profession: Present Status and Future Directions*, ed. Thomas H. Geno (Middlebury, VT: Northeast Conference, 1980), pp. 143-56.

[9] At the present time, ACTFL has been taking the beginning steps towards a national criteria and assessment program: see also Regina H. Paul, "Needed: Stepladders of Foreign Language Learning," *Foreign Language Annals* 14 (1981), p. 379-84.

[10] Both Omaggio and Jarvis & Adams provide illustrative discussions of this point.

which provides information useful for decision making. Alkin makes a needed distinction between "evaluation" and "research": "On the one hand, there are studies designed primarily to add to the body of knowledge (research), on the other, those studies designed primarily to provide information for decision making (evaluation). And these two functions are separate and distinct."[11] Decisions about foreign language programs and staff training are being formulated today by those outside the profession — the urgent first step in regaining control for the profession is to be able to provide valid data to the decision makers.

The literature of evaluation provides several prescriptive models for the techniques of program assessment, for the collection of data and for its subsequent analysis; the reader needing training in these areas is encouraged to consult these references.[12] There are, however, certain elements common to most evaluation designs which, if carried out properly, will provide useful and reliable information about programs and staff. Each of these general evaluation activities will be discussed in relation to how they can help *foreign language* programs reestablish their credibility and provide convincing data to decision makers.

Evaluability Assessment. Evaluability assessment is defined as a process for determining in advance the likelihood of an evaluation's success.[13] It consists of two stages: first, an examination of program characteristics and second, an evaluation feasibility analysis. Many foreign language "programs" are in fact not comprehensive enough to support a full-scale program evaluation and need to modify their approach to evaluation to reflect local conditions. Other programs lack qualified staff, as seen in this example provided by Simon: "Some years ago I received a letter from a mother whose son had taken two years of high school Spanish and received straight A's. When he entered the University of Illinois, he was not admitted to an advanced class. They advised him that he had learned almost nothing in those two years of Spanish. When I chatted with the principal of that small high school, he told me that the teacher had only limited knowledge of Spanish. He was primarily a mathematics teacher" (p. 123). Such programs, whose lack of quality and effectiveness is evident, are unlikely candidates for full-scale evaluations and would be weeded out by an evaluability assessment. On the other hand, a program which is deemed feasible for further study can be assured of the utility of continuing with ongoing evaluation activities.

Needs Assessment. Needs assessment is defined as the determination of the difference between the desired status (of program, of learner, etc.) and the current status (Popham, p. 65). Quality needs assessment data can be used by foreign language programs and staff to:

- provide for community input as to what is seen as important in language study;
- develop defensible goals for language study which are appropriate to local conditions, philosophy and resources;
- determine the present status of student achievement relative to program goals;
- determine the educational needs of students in order to achieve program goals;
- design instructional sequences based directly on student needs;
- determine priorities for test development;
- determine requisite teacher competencies;
- provide evidence of what training needs of teachers are based on these identified teacher competencies;
- provide valid data to the public as to the importance of language study to the nation;
- provide valid data to the public as to the need for making language study an integral part of, rather than an optional adjunct to, international studies.

Needs assessment data, when categorized and aggregated on a national basis, can provide a defensible focal point for the joint efforts of the professional organizations. Unfortunately, needs assessment data has, to date, been notably lacking and thus, has allowed critics to level charges unsubstantiated by hard data such as those quoted in both the President's Commission report and Simon's book. Needs assessment data can also provide a defensible basis for making decisions on the design of a program which will be based on the expressed wants of the local community, the needs of the students and the competencies of the staff.

---

[11] Marvin C. Alkin, Richard Daillak and Peter White. *Using Evaluations: Does Evaluation Make a Difference?* (Beverly Hills: Sage, 1979). p. 13.

[12] For example, see W. James Popham, *Educational Evaluation* (Englewood Cliffs, NJ: Prentice-Hall, 1975) and Joint Committee on Standards for Educational Evaluation, *Standards for Evaluations of Educational Programs, Projects and Materials* (New York: McGraw-Hill, 1981).

[13] Bruce Thompson, Jean King and Ellen Pechman, "Evaluation Utilization: A Bibliography." Unpublished paper (New Orleans: Tulane University, 1971).

Implementation Evaluation. Implementation evaluation is defined as an assessment of the actual status of program activities as compared with the original program plan. Such a procedure permits an accurate description of exactly what the program being evaluated consists of in reality, rather than on paper. It provides a common frame of reference to all parties concerned with the evaluation and allows for the explanation of the differences between what was planned and what was actually done. An implementation evaluation can provide:

- a common definition within the professsion as to the characteristics of different types of foreign language programs;
- evidence that a program claiming to be of a specific type (i.e., "immersion," "individualized," "audiolingual") is in fact a program of that specific type;
- evidence that what is being measured by tests and other observations represents the learning environment which is in fact being fostered through a program's instructional and related activities.

Implementation evaluation data can allow educators to establish program categories and descriptions which will remain consistent across the nation. The implementation evaluation process will enable the profession to use a common set of terminology and thus, avoid some of the problems currently being encountered by the related field of bilingual education, which has seen a multiplicity of program types, purposes and methodologies all subjected to mislabeling under the general heading of "bilingual education." Implementation evaluation data will also allow for the setting up of comparative program studies since educators can be certain that programs which are in fact similar in design, orientation and methodology are selected for comparison purposes.

Formative Evaluation. Formative evaluation looks at a program during its early stages while program elements can still be changed, if desired, in response to local conditions. This stage can last from the first day of program implementation until such time as no further modifications can be made and the program stands intact as an entity. As Jarvis and Adams explain, "The purpose of formative evaluation is to improve the instruction: It asks, in effect, about the current status of the program so that it can be made better. It is evaluation that is carried out during the development, implementation and operation of the program" (p. 6).

Formative evaluation data are solely needed by (and lacking in) most foreign language programs.

Such data can be used to:

- examine the effectiveness of current instructional materials towards meeting the program's instructional goals;
- examine the match between program goals and learner activities;
- look at the match between testing content/strategies and learner instructional activities;
- develop criteria for program self-assessment;
- investigate the effectiveness of various teaching methodologies in achieving program goals;
- examine the relationship between teaching methodology and program conditions (i.e., which teaching methodologies are best suited to the various types of programs?);
- identify the most effective components to the "foreign language teacher style";
- look at "innovative" programs in their early years of implementation in order to identify the most promising elements and to increase their effectiveness (John Rassias' program at Dartmouth would be an example of such a program).

Rigorous formative evaluation procedures will provide defensible evidence leading to increased credibility for the profession's efforts to improve instructional programs. Demonstrated program successes can be more quickly and efficiently communicated, both to the profession and to the general public. Ongoing formative evaluation as an integral part of a foreign language program is a sine qua non for providing valid data to decision makers. The importance of formative evaluation for the profession is also stressed by Jarvis and Adams: "There is a need for a more formative evaluation in language programs. Rarely do we initiate evaluation even for the purpose of improving our programs. Even in the light of the very heavy loads of most instructors, evaluation is still far too important an activity to be so neglected" (p.7).

Summative Evaluation. Summative evaluation focuses on determining the merits of a program at its completion. As Jarvis and Adams summarize, "Summative evaluation is terminal evaluation of a program that is already operational. Its purpose is to make judgments about a program's worth . . . . Ultimately, summative evaluation is tied to decisions about support and continuation of a program" (p.6).

Jarvis and Adams go on to draw some useful distinctions between "formative" and "summative" evaluation:

> In a formative evaluation it is important to obtain day to day feedback on specific aspects of the instruction. This information is, however, often fragmentary and of limited use in a summative evaluation. Conversely, an assessment of the impact of a new program on

the image of language study in a particular school may be very important in a summative evaluation, but is inappropriate for a formative purpose. Formative and summative evaluators often do not 'behave' in the same way. A formative evaluator's procedures may be more partisan than a summative evaluator's approach. The summative evaluator must be objective and circumspect. A formative evaluator can use shortcuts, small samples and intuition in an effort to improve the program (pp.6-7).

Summative evaluation data is the most authoritative and defensible information the profession can provide to any interested parties. Such data can be used to:

- provide replicable outcome data for all types of foreign language programs;
- provide comparative program outcome data for competing program types;
- determine effective instructional materials and teaching methodologies;
- identify exemplary programs worthy of dissemination and/or replication;
- provide program cost-effectiveness data.

Improving the Utilization of Evaluation Information

Although data gathering and communication to relevant audiences should be ongoing activities throughout the entire evaluation process, there is a priority need for the foreign language profession to be able to communicate that information to the audiences that need to use it the most. In today's climate, these audiences are generally perceived to be the decision makers, people both within and outside the profession, at the local, state and national levels empowered to make decisions affecting FL programs.

Recent discussions of evaluation utilization have led to the point of view that evaluation data have many different kinds of use.[14] For example, such data can be considered by a local, client or by related external interested parties as an influence in making decisions, substantiating previous decisions or actions, or establishing or altering attitudes (Alkin, p. 232). Within this context, evaluation use can be instrumental, conceptual or symbolic: "Instrumental use occurs (where) action is taken in direct response to the results of an evaluation . . . In contrast, conceptual use refers to cases where evaluation results influence decision makers' current thinking about (and potentially their future action regarding) an issue or program . . . A third type of use has little to do with the actual content of the evaluation results. Symbolic use refers to cases where evaluation results are used indirectly for a variety of purposes, for example, to garner political support, to substitute for a decision, or to discredit a disliked policy" (Thompson).

Regardless of which type of use is being targeted, three major elements in the process have been shown to influence the degree of potential success. These elements are:

1) the "personal" factor. The personal factor is defined as the active presence within an evaluation study of at least one person who cares about the process and its results and who takes personal responsibility for promoting use of the findings. As Patton discovered in his studies on evaluation utilization: "Where such a person or group was present, evaluations were used; where the personal factor was absent, there was a correspondingly marked absence of evaluation impact. The personal factor represents the leadership, interest, enthusiasm, determination, commitment, aggressiveness, and caring of specific individual people."[15]

2) The political considerations factor. Within Patton's studies, (researchers) "found that the political factors consistently emerged to affect the utilization process, whether or not the decision makers and evaluators were fully aware of the political implications of the study from the outset" (p.49). Patton also identified the political nature of the evaluation process itself: "the fact that people are involved in evaluation research; the fact that classification systems and categories are involved; the fact that actions and decisions are the desired result of evaluation; the fact that programs and organizations are involved; and the fact that information is involved" (p. 47-8).

3) the nature and format of the evaluation report. As Alkin's team explains,

> The evaluator must call on both his personal and professional skills in determining the appropriate *substance* of evaluation information. First, s/he must have a clear idea of the people to whom the information will be addressed; second, s/he must decide what information, from all that has been gathered, will be provided to the selected person(s) and in what way. If it is well enough tailored to the users' specific concerns and questions about

[14] Carol Weiss, "Utilization of Evaluation: Towards Comparative Study," *Evaluating Action Programs: Readings in Social Action and Education*, ed. C.H. Weiss (Boston: Allyn and Bacon, 1972).

[15] Michael Patton, *Utilization-Focused Evaluation* (Beverly Hills: Sage, 1978), p. 64.

> the program, we can expect that potential utility will be greatly enhanced . . . Closely related to substance, of course, is the appropriateness of the *format* in which the information is presented. What the evaluation has to offer in the way of data and insights may be exactly what the program (or other external) personnel want or need to know, but unless it is in an intelligible form it may well be ignored or considered useless. Personal preferences must be considered; some people are simply uncomfortable with quantitative data, for example (p. 253-4).

These research findings convey specific suggestions and recommended courses of action to professionals desirous of ensuring that foreign language program evaluation data will be used. Foreign language educators need themselves to become conversant with the purposes and methodologies of evaluation; to look at their own instructional settings and programs to determine those evaluation activities most likely to help them both improve their programs and demonstrate success to the public; to take the initiative to embody that "personal" factor so crucial to the effective use of carefully obtained evaluation data and to local public relations efforts to counter unsubstantiated criticisms; to identify local and national decision makers and the political/social climate in which these persons are operating so that they may become knowledgeable and influential enough to participate in the decision-making process; and to identify the potential audiences for evaluation data (e.g., parents, news media, administrators, etc.) so that the content and format of evaluation reports can be tailored to suit the needs and preferences of each audience. Finally, and perhaps most importantly, foreign language educators need to make evaluation and evaluators their allies in the process to improve programs and thus ultimately to regain control over the direction of foreign language programs within the country.

Bibliography

Alkin, M.C.; Daillak, R. & White, P. *Using Evaluations: Does Evaluation Make A Difference?* Beverly Hills: Sage, 1979.

Benseler, David P. "The American Language Association: Toward New Strength, Visibility and Effectiveness as a Profession." In *Our Profession: Present Status and Future Directions*, ed. Thomas H. Geno. Middlebury, VT: Northeast Conference on the Teaching of Foreign Languages, 1980.

Benseler, David P.(ed) *Teaching the Basics in the Foreign Language Classroom: Options, and Strategies*. Skokie, IL: National Textbook Co., 1979.

Central States Conference on the Teaching of Foreign Languages. "Alternatives, Basics and Competencies in Foreign Language Instruction." *Modern Language Journal*, 63 (1979).

Clark, John L.D. "Measurement Implications of Recent Trends in Foreign Language Teaching." In *Foreign Language Education: A Reappraisal*, ed. Dale Lange. Skokie, IL: National Textbook Co., 1972.

Jarvis, G. & Adams, S. *Evaluating a Second Language Program*. Arlington, VA: Center for Applied Linguistics, 1979.

Joint Committee on Standards for Educational Evaluation. *Standards for Evaluations of Educational Programs, Projects and Materials*. New York: McGraw-Hill, 1981.

Omaggio, Alice. "Looking at the Results." In *Building on Experience, Building for Success*, ed. June Phillips. Skokie, IL: National Textbook Co., 1979.

Patton, Michael. *Utilization-Focused Evaluation*. Beverly Hills: Sage, 1978.

Paul, Regina H. "Needed: Stepladders of Foreign Language Learning." *Foreign Language Annals*, 14 (1981).

Popham, W. James. *Educational Evaluation*. Englewood Cliffs, NJ: Prentice-Hall, 1975.

President's Commission on Foreign Languages and International Studies. *Strength Through Wisdom: A Critique of U.S. Capability*. Washington, D.C., 1980.

Schulz, Renate A. "Back to Basics in the Foreign Language Classroom?" *Foreign Language Annals*, 11 (1978).

Simon, Paul. *The Tongue-Tied American*. New York: Continuum, 1980.

Thompson, B.; King, J. and Pechman, E. "Evaluation Utilization: A Bibliography." Unpublished paper. New Orleans: Tulane University, 1971.

Valette, R. and Linder, c. "Measuring the Variables and Testing the Outcomes." In *Building on Experience, Building for Success*, ed. June Phillips. Skokie, IL: National Textbook Co., 1979.

Weiss, Carol. "Utilization of Evaluation: Towards Comparative Study." In *Evaluating Action Programs: Readings in Social Action and Education*, ed. C.H. Weiss. Boston: Allyn and Bacon, 1972.

USE OF NATIVE LANGUAGE TESTS FOR PROGRAM PLANNING PURPOSES

Flora V. Rodríguez-Brown
University of Illinois-Chicago
Illinois Resource Center

A Rationale for Testing Needs in a Bilingual Classroom

Since bilingual education programs supported by Federal and State funds are to be transitional in nature, English proficiency is the most relevant factor in determining participation in bilingual programs. Most of the time, programs are designed to serve children's needs according to their different levels of English proficiency as shown by their performance on different language proficiency tests.

Tests currently in use which determine bilingual program participation measure proficiency from many different perspectives. Some tests measure only one aspect of language (i.e. vocabulary or syntax). Others include several variables which the authors define as being important aspects of language development derived from their own conceptual framework of language. While some tests tend to measure proficiency from a more holistic pragmatic perspective, others measure it from a discrete point perspective. A recent report from NIE (1981) describes findings of a comparison of results of the four major tests currently used to determine the English language proficiency of LEP children. In the study, the results of the different tests showed little correlation in terms of proficiency levels. It was found that the lack of correlation was due to the fact that each test selected different aspects of language to measure in determining proficiency.

Researchers involved in studies in this area (Hayes, 1982; Slaughter and Bennett, 1982; and Rodriguez-Brown and Elias-Olivares, 1981) have started to discuss the need to redefine the concept of language proficiency. Central to the discussion is the idea of looking at language from a holistic perspective. In terms of bilingual education, these researchers urge a redefinition of the concept to clarify whether we are talking of that proficiency necessary to succeed in school or in life. The idea of looking at language proficiency from the perspective of what the child can do, rather than what adults feel the child should be able to do at different levels of proficiency, has been posed as a more realistic approach. Another area of discussion is the relationship of the concept of language proficiency to communicative competence and functional language ability (Slaughter and Bennett, 1982; Hayes, 1982; and Rodriguez-Brown and Elias-Olivares, 1981). From the view that language proficiency should be measured as part of communicative competence, then, tests and other measures should take into account not only what the child can do in L_2 (English) but, what s/he can do in his/her native language (L_1) as well. It seems relevant to know not only the levels of proficiency in L_1 and L_2, but also how the child uses language (L_1 or L_2) in relation to the setting, topic, and formality levels. While natural language samples seem to give a more holistic perspective of what the child can do in each language, their use is restricted due to lack of standardization. Because of the need to look at language from a holistic framework, the need to look at proficiency in terms of L_1 and L_2 and the need to examine proficiency from a sociolinguistic perspective, an assessment of a child's language proficiency in both the L_1 and L_2 will be required. In this way, decisions regarding program planning will be made according to what children already know. This will facilitate the learning of children enrolled in bilingual programs.

Research reports by Skutnabb-Kangas and Toukomaa (1976), Cziko (1976), and Tucker (1975), among others, have suggested a strong relationship between the development of L_1 and the learning of L_2. These research results add support to the rationale for a need to assess L_1 and L_2 development in children attending bilingual programs.

Skutnabb-Kangas and Toukomaa's study (1976) has shown that Finnish immigrant children who were instructed in Swedish (their L_2) before they reached the preoperational stage of development tended to have problems learning their L_2 (Swedish) and achieving at grade level with other students in the system. In contrast, Finnish children who began instruction in Swedish (L_2) at an older age and who had established their first language skills before immigration approach the achievement levels of majority (Swedish) children. These data suggests that a certain level of development in L_1 is a prerequisite for learning L_2 and achievement in school. Cummins (1979) stated that before bilingualism can become an asset in learning, one must reach a certain threshold level of language development in L_1 and L_2.

Along with these developments which show the relationship between L_1 and L_2 and achievement, several researchers have shown the need to look beyond language when planning programs for "bilingual" [1] children (Bowen, 1977; Troike, 1981; Tucker, 1977; Rodriguez-Brown, 1979; and Cummins, 1982. These researchers emphasize the need to examine the child's established sociocultural traits to determine issues such as participation in the bilingual program and media of instruction in the program.

[1] "Bilingual" children is used in this paper to include children whose home language background is not English and who show low proficiency in English. As such, the term may include those who only speak L_1.

In spite this research evidence which can provide some direction in bilingual program design and implementation, most of the decisions currently being made are based solely on the results of English-language proficiency testing. Commonly tests used to measure English-language proficiency are narrow in scope, do not assess common areas or accurately indicate a child's ability to communicate. If we want to plan programs that better serve the needs of children attending bilingual programs, it seems necessary to emphasize the need for the teacher to know what each child can do linguistically in both L_1 and L_2, his stage of cognitive development or readiness for learning and sociocultural information that might help enhance the program's impact. If teachers had this type of information, they would be able to make better decisions regarding instruction to enhance children's participation and learning opportunities in the program. While teacher observations can aid in collecting most of the data described before, testing is a most useful tool in gathering program planning information for specific children.

Some difficulties with formal measures have caused teachers to acquire negative attitudes toward testing. They feel they have to administer too many tests and, in some cases, do not receive the results. Furthermore, when they do receive the results, they often do not know how to interpret and/or use them. Teachers find that time spent in testing reduces the amount of time available to deliver the curriculum prescribed by the school district. In the case of bilingual teachers, it is important that the benefits which testing results may have for designing and improving instructional programs become more evident. It is important to emphasize that whether bilingual programs are transitional or not, native language testing results offer a rich source of information which will facilitate decisions regarding the language used for instruction, teaching methods, grouping and selection of curriculum materials. Wise decision making will enhance achievement and compensate for environmental situations the teacher may not be able to control directly.

The use of English language proficiency test results have been overemphasized in bilingual program design and planning. It is crucial, following current research findings (Skutnabb-Kangas and Toukomaa, 1976, Cummins, 1979) that we try to look beyond English proficiency. For children who attend bilingual programs, it is important to look at native language proficiency as well as readiness skills and previous school achievement. The following section of this paper will address the need to administer native language testing in different areas to enhance the quality of instruction provided in bilingual programs.

Application of Native Language Test Results in Bilingual Programs

Presently in the state of Illinois, bilingual programs have to report native language testing results for children attending the programs. As such, testing in native languages will be required to fulfill state bilingual education guidelines. Teachers, then, should take advantage of this requirement to collect native language information to aid in planning classroom activities specific to the needs of children attending bilingual programs. In choosing tests, administrators and teachers should choose those tests which best correlate with the educational objectives of the district. Tests should provide users with varied information regarding both what children do and do not know, as well as their areas of strength and weakness.

In regard to native language testing, one of the first areas of concern is to determine with accuracy the native language of the child. This may not be a problem if we are dealing with Hispanic children, but it has proven to be a problem with Indochinese students. Some of these children have lived in several countries before they immigrated to the United States, and as a result, districts tend to assume that the language spoken in the last country (or area) of residence is the child's native language. Indochinese students may have been in contact with several languages and their knowledge of each will vary. Parents should be involved in the process of defining the child's primary language before any native language tests are administered.

Since studies (Tucker, 1975, Cziko, 1976) have shown a strong relationship between L_1 and L_2 development in bilingual children, it seems important, then, to look at the proficiency of the students in the two languages. It is necessary not only to compare the proficiency levels in L_1 and L_2, but to identify areas of strength and weakness in two languages. When looking at language proficiency in L_1 and L_2, it is important to look at the proficiency levels by analyzing the results of more than one measure. Multiple-part tests which measure different aspects of language, observational surveys such as the Student Oral Language Observational Matrix (SOLOM California State Department of Education, 1981) and teacher informal observations used together can give a more valid and reliable view of the students true proficiency, including areas of strength and weakness. For Hispanic students, a test such as the Woodcock Proficiency Test (Woodcock, 1981, in English and Spanish) may provide the teacher with useful diagnostic information for classroom planning.

Due to the high priority given reading in school curriculum in the United States, this aspect of proficiency receives special attention from program planners. Districts are interested in determining the children's reading ability in the native language to place them at an equivalent level in the test measurement providing program direction. The best use of native language reading test results is that of finding out what children can do in their native language. For program planning, it is important to know not only the highest grade level attained in L_1, but also how relevant the objectives of the program he attended in L_1 were to the L_2 instructional program the school district will be providing for the child.

We know, more from common sense than from research findings, that some reading skills transfer from L_1 to L_2. This tells us a child who has some background in reading in L_1 does not have to start to learn to read in English from scratch. It is for this reason that it is important to find out whether the child knows how to read in L_1 and which skills he brings to school with him. Native language test results could be a useful tool in determining a child's readiness to learn. Diagnostic tests in L_1 may help determine the child's strengths and weaknesses in reading. Besides the usefulness of these data for decision making in reading instruction, they could also be a valuable aid for grouping and general program planning.

Cognitive development and readiness levels determined via the native language testing results can affect decisions in the bilingual program. As educators, many times we judge bilingual children's achievement problems as related only to their lack of proficiency in English. Very seldom do we think of whether the children have the readiness skills or the cognitive development necessary to perform well in content areas in school. Tests such as the Boehm Test of Basic Concepts administered both in English and the native language of the children could help teachers determine if children have acquired concepts which are basic for learning. If the children have those concepts, they are functional in an English-speaking classroom. Otherwise they may need to simultaneously acquire both the concepts and English skills.

Current practices in referral of bilingual children for psychological testing seem to overlook the fact that children attending bilingual programs usually have attended schools in other countries where the language of instruction as well as the children's learning experiences varied. The fact that children come to the bilingual program or the regular school with a certain degree of development in L_1 should not be overlooked before making any judgment on the children's abilities.

It is important for teachers and school districts to collect enough observational and/or diagnostic testing data related to performance in the native language so that possible areas of concern hindering the student's learning capacity and/or producing behavioral disorders might be discovered. Problem areas may be related directly to cultural and linguistic differences, thereby not suggesting additional psychological testing, or they may be related to fundamental learning problems.

School districts should establish policies in which, as a general rule, bilingual children are not referred for psychological testing at least during their first six months of schooling in the United States. Additionally, any referral submitted after this period should be accompanied by a thorough report showing the process used by the teacher to determine the need for a referral. This report should include information concerning how the native language of the child was determined (especially in the case of Indochinese students), observations made by teachers and other school personnel, and previous and/or current school records as well as testing information collected in the native language of the child. This testing information could be used to identify areas of concern which should be further observed and assessed.

In spite of the lack of valid and reliable psychological tests constructed in languages other than English, school districts have been trying to test children with translations of English language tests. The fact that the district is conscious that the child does not speak English is commendable, but this is not enough. Mistakes have been made recently even in determining the native language of the children. There are cases where children have been tested in a different foreign language than their native language with translations of tests which were not reliable or valid when translated. Under these conditions, it is important to think of the number of children who have already been misplaced in special programs following this dubious process. Psychological testing is an area in which erroneous native language testing can be a liability rather than help for bilingual children. This is why a prereferral assessment period that allows for the collection of observational as well as diagnostic information from the teacher is so crucial.

Availability of Resources for Native-Language Testing

When talking about native language testing, two questions are consistently raised: first, "what native-language tests are available?" and second, "who can administer native-language tests when school personnel in a district are not available for specific languages needed?"

An effort to provide resources which show the availability of native-language tests is a survey done by the State of Illinois (Rodriguez-Brown and Starker-Rowe, 1982). The survey tried to determine a) whether the state offices of bilingual education or individual districts recommend specific native-language tests as part of the program design and planning procedures, b) whether districts have developed tests in native languages which can be useful for other programs, and c) the availability from publishers of native-language tests which can be used with bilingual students.

The survey showed that few states recommend any specific native language tests and furthermore that most districts surveyed did not require or see a need for native language testing. In addition, a majority were unaware of available tests and/or procedures. A few states and districts, though, referred us to published or district-made tests in such areas as L_1 language proficiency, reading, general achievement, math, readiness skills, etc. The languages of the tests found were varied. Among them were Arabic, Chinese, Greek, Hebrew, Italian, Russian and Vietnamese. A survey of current test bibliographies in wide use coupled with answers received from the states and districts have helped us identify as many as 450 tests in different areas (language proficiency achievement, etc.) in 39

different languages

Many of these tests have not been reviewed yet. Many of them are district made. They are not perfect, but they seem to serve the need for which the test developers made the test. These tests are not meant to provide a final assessment and/or diagnostic of student's problems, but they could be useful tools for teachers and administrators to find out what the child knows in the native language when s/he gets to school. The results of the survey, future updates of concerns available, data as well as a review of the different tests will be a useful source of information on the availability of native-language tests in a variety of languages and skill areas.

In regard to the administration of tests in languages for which bilingual teaching personnel are unavailable, teacher aides, parents and other community people proficient in the language (and in certain cases even peers) can provide valuable assistance in the testing. Another question which has arisen and requires attention is what to do when tests in specific languages do not exist. In this situation, Illinois is committed to train teachers in the use of alternatives to testing which can provide useful information especially about language development. Error (miscue) analysis, native language sample analysis and the cloze procedure are alternatives to testing which school personnel can be trained to use with bilingual students to sample native language proficiency. In this way, the design of various school programs can better serve the needs of these children.

Error (miscue) analysis, for example, can be used to determine reading comprehension skills in the native language, as well as the stage of development of reading skills in terms of the student's use of graphic versus semantic constraints in the text. This information provides the teacher with information which may facilitate the design of a program that serves better the needs of individual children. The cloze procedure can be used not only to determine reading comprehension skills, but vocabulary and syntax development. Natural language samples can provide the teacher with a wider perspective of what children can do linguistically without limitations to a particular context and setting. These samples, when analyzed, can provide invaluable diagnostic information in regard to the receptive and expressive skills of children. These analyses can be focused not only on the formal aspects of language but at the functional level as well. Natural language samples provide the teacher with a more global view of children's linguistic skills which, in turn, will facilitate the teacher's role in planning activities which are relevant to a particular child's needs. The school's programs then will be more conducive to learning and better serve the linguistic and academic needs of children in bilingual programs.

Summary

The purpose of this paper is to emphasize the need for native language testing in order to plan programs of instruction that better serve the needs of linguistically and culturally different children. Until now, programs have focused principally, if not exclusively, on English proficiency for program design and planning purposes.

Research evidence exists which shows a strong relationship between proficiency in L_1 and L_2 and achievement. Native-language testing, then, can provide important information to plan programs which take into account what the child knows and what s/he is ready to learn when s/he gets to school. Native-language testing has been carelessly used in psychological testing of ethnolinguistic minority children. It is suggested that a moratory period be established before referral of non-English-speaking children to special programs is made whereby teachers could collect data that will identify the true native language of the child, as well as collect observational and diagnostic data in the native language which will help determine the appropriateness of further referral. These data may show specific areas of concern which may require further testing.

Results of a survey of native-language tests presently available have shown that there are more available native language tests than most people believe. These tests are not perfect, but they serve a need. The results from these tests are not final, but they can help the teacher in decision making. Data from these tests can be collected not only by teachers, but by supportive staff, peers and community people. When tests are not available, alternative methods of data collection can be used. Alternatives discussed are error (miscue) analysis, native language sample analysis and the cloze procedure. In general, the paper presents an argument for the need for native-language testing in order to provide information to design relevant programs for children enrolled in bilingual programs.

References

Bowen, J.D. "Linguistic Perspectives on Bilingual Education." In B. Spolsky and R. Cooper (eds.) *Frontiers of Bilingual Education*. Rowley, Mass: Newbury House, 1977.

Cummins, James. "Linguistic Interdependence and the Educational Development of Bilingual Children." *Bilingual Education Paper Series*, September, 1979, 43, (2).

Cummins, James. "The Effectiveness of Bilingual Education." Toronto, Canada: The Ontario Institute for Studies in Education, 1982.

Cziko, Gary A. "The Effects of Language Sequencing on the Development of Bilingual Reading Skills." Manuscript. McGill University, Montreal, Canada, 1976.

Hayes, Zoe Ann. "Linguistic and Communicative Assessment of Bilingual Children." In C. Rivera (Eds.) *Entry and Exit Procedures in Bilingual Education: Education and Policy Issues*. Washington, DC: Center for Applied Linguistics, 1982.

National Institute of Education (NIE). "Report of the National Institute of Education on the Testing and Assessment Implications of the Title VI Language Minority Proposed Rules." Washington, DC: Manuscript, 1981.

Rodriguez-Brown, Flora. "The Effect of Language Used for Early Reading Instruction: A Bilingual Perspective." Report to NIE, Grant Number NIE-G-78-0134 submitted to the National Institute of Education, September, 1979.

Rodriguez-Brown, Flora and Elias Olivares, E. "Linguistic Repertoires, Communicative Competence and the Hispanic Child." Paper presented at the Language Proficiency Assessment Symposium (LPAS), Washington, VA, March 14-18, 1981.

Rodriguez-Brown, F. and Starker-Bowe, G. "Native Language Survey." Manuscript. Arlington Heights, Illinois: Illinois Resource Center, 1982.

Skutnabb-Kangas, T. and Toukomaa, P. *Testing Migrant Children's Mother Tongue and Learning the Language of the Host Country in the Contexts of the Monocultural Situation of the Migrant Family*. Helsinki: The Finnish National Commission for UNESCO, 1976.

Slaughter, Helen H. and Bennett, Adrian T. "Sociolinguistic Discourse Alternative for Language Assessment." Paper presented at the Annual Meeting of the American Educational Research Association in New York City, March 19, 1982.

State of California. "Student Oral Language Observation Matrix (SOLOM)." In *Individual Learning Programs for LEP Students*. Sacramento, CA, 1981.

Troike, Rudolph C. "SCALP: Social and Cultural Aspects of Language Proficiency." Paper presented at the Language Proficiency Assessment Symposium, Arlie House, Warrenton, VA, March, 1981.

Tucker, G. Richard. "The Development of Reading Skills within a Bilingual Program." In S.S. Smiley and J.C. Towner (Eds.), *Language and Reading*. Bellingham, Washington: Sixth Western Washington Symposium on Learning, 1975.

Tucker, G. Richard. "The Linguistic Perspective." In *Bilingual Education: Current Perspectives* (Vol. 2). Arlington, VA: Center for Applied Linguistics, 1977.

Woodcock, Richard W. *Bateria Woodcock de proficiencia en el idioma*. Hingham, MA: Teaching Resources, 1981.

ASSESSMENT CONSIDERATIONS FOR LIMITED-ENGLISH-PROFICIENT VOCATIONAL STUDENTS

Joan E. Friedenberg
Associate Professor of Bilingual Education
Florida International University

Curtis H. Bradley
Professor of Vocational Education
Florida International University

Language assessment of limited English proficient (LEP) students enrolled in vocational education programs requires some unusually interesting considerations. Unlike in most ESL and bilingual programs where LEP students must strive to master the total linguistic system of English, in a vocational training situation the language component is limited to that required for survival in the vocational education classroom, laboratory, or shop and for success on the job.

Before the reader can fully appreciate the uniqueness of the vocational situation, it is important to clearly illustrate the differences among vocational English language instruction and other types of ESL programs. Some of these program types are summarized below.

1. General ESL. The purpose of this program is to provide students with all of the lexical, structural, and phonological components of English. The sequence of instruction is usually based on grammatical structures, from the simplest and most frequent to the most complex. Instructional techniques may range from highly controlled (e.g. repetition, substitution, conversion, and dialogue memorization) to less controlled (question-answer, discussions, and role plays).

2. ESP (English for Special Purposes). The purpose of this program is to provide students with the specific English components and skills needed to function in a given situation. Although job-related ESL technically fits under this category, ESP programs in the U.S. have, thus far, exclusively served students with strong academic skills who are preparing for professional-level careers.

3. Prevocational ESL. The purpose of this program is to provide learners with the English necessary to survive in a U.S. community. The sequence of instruction is usually situational and includes such topics as "calling a doctor," "opening a bank account," "public transportation," and "reading want-ads."

4. VESL (Vocational ESL). The purpose of this program is to provide learners with the English language components and skills necessary to survive in a vocational education classroom and a job. The sequence of instruction is situational and is related to topics in the given occupation. In an ideal setting, VESL instruction is provided simultaneously with bilingual vocational education (BVE) or bilingual on-the-job training.

Despite the fact that job-related ESL programs are rapidly increasing, there is a serious lack of appropriate assessment instruments. For example, according to the *Administrator's Manual for the Bilingual Vocational Oral Proficiency Test* (Melton Peninsula, 1981):

> Although instruments have been developed to determine the level of English proficiency of children, these tests are not appropriate for an adult population. Furthermore, most of the English proficiency tests developed for adults were designed to test the language required for college and university work. (p. 2)

Because of the similarities between VESL and ESP, and because of ESP's "head start" in the areas of program and materials development, one might suggest that VESL educators model the assessment process after that employed by ESP educators. However, there is an equally serious lack of assessment measures in ESP, as pointed out by Strevens (1977):

> It remains a major shortcoming of ESP that very little work has been done to devise fresh methods of testing, examining and assessment that match the new courses of training. (p. 129)

The purpose of this chapter is to present some important considerations in assessing non- and limited-English-proficient vocational students. Although discussions are included about VESL placement testing and vocational skills assessment, the major focus will be on the process of deriving occupation-specific content for VESL curricula and classroom achievement tests. A discussion of appropriate testing techniques is also provided.

Types of Assessment for BVE & VESL

The assessment process for LEP students enrolled in vocational education programs is a complex one that requires a good deal of collaboration among VESL instructors, vocational instructors, and vocational counselors. We see this process as containing five essential components:

1. Initial assessment of English language proficiency.
2. Initial assessment of vocational skills.
3. Vocational skills achievement (for progress).
4. VESL achievement (for mastery).
5. VESL achievement (for progress).

Initial Assessment of Vocational English-Language Proficiency

An assessment of English-language proficiency should take place as soon as the LEP students enroll in the vocational education program. This assessment serves many important functions.

1. It can provide a measure for VESL placement purposes.
2. It can provide a measure to determine whether the assessment of vocational skills must be in the native language or whether it can be in English.
3. It can provide a measure to help determine growth in English proficiency (achievement) during the VESL training period.
4. It can provide a measure to determine whether and how much vocational instruction in the native language is needed.

Bradley and Friedenberg (1982) presented three important characteristics of an English proficiency test for LEP adults:

1. does not require reading,
2. is appropriate in content for adults,
3. will elicit the student's ability to communicate a message as opposed to assessing formal grammatical correctness (p. 147).

In addition to these characteristics, it is also important that the test not focus on content with which only those with strong academic backgrounds could feel comfortable and that it reflect the structures used most frequently in vocational education classes. Figure 1 illustrates the linguistic structures which have a particularly high frequency across vocational education classes.

Figure 1.

Linguistic Structures Which Have a High Frequency across Vocational Education Classes*

SENTENCE TYPE	
N V tr N	Noun Transitive Verb Noun
N be N (N V be N)	Noun Verb "be" Noun
N be Adj (N V be Adj)	Noun Verb "be" Adjective
N be Adv (N V be Adv)	Noun Verb "be" Adverb
Noun V tr N N	Noun Transitive Verb Noun Noun
N V int N	Noun Intransitive Verb Noun
N V int Adj	Noun Intransitive Verb Adjective
VERB TENSE	
Pres	Present
Past	Past
Pres Prog	Present Progressive
Modal	Modal
TRANSFORMATION	
Wh	Who? What? When? Where? Why? Question
Yes-No Ques	Yes-No Question
Comp	Compound
Imp	Imperative
Cl-Adj	Clause-Adjective
Cl-Adv	Clause-Adverb
Cl-Noun	Clause-Noun
Pass	Passive
Part	Participle
Ger	Gerund
Inf	Infinitive
Do	"Do" Insertion
Neg	Negative
MODIFIER	
Prep Ph-Adv	Prepositional Phrase-Adverb
Prep Ph-Adj	Prepositional Phrase-Adjective
Adj	Adjective
Adv	Adverb
Intens	Intensifier

*Melton Peninsula, 1981, p. 4.

Ideally, the assessment of English language proficiency, in which an appropriate standardized test is used, should be a careful, collaborative effort between the ESL instructor and vocational counselor. One test, which was developed expressly for the purposes stated here, is the *Bilingual Vocational Oral Proficiency Test* (BVOPT). However, if a test such as the BVOPT is unavailable or if there is no one able or willing to conduct a formal measure of English language proficiency, the vocational instructor should informally assess the students' English abilities by interviewing each of them in class or privately. This will, at least, provide a rough indication of whether the students are fluent, limited, or non-English-speaking.

Initial Assessment of Vocational Skills

The initial assessment of vocational skills is conducted by the vocational instructor to determine the point at which vocational instruction must begin for each individual student. The assessment activities normally consist of a series of instructor-developed psychomotor and cognitive items designed to measure student ability prior to instruction. Thus, in a typical pretest, the student might be asked to explain and/or demonstrate a procedure or process. Language-related problems often cause the LEP student to score less well on this test than is warranted by his or her actual vocational skills level. Therefore, this assessment should be conducted in the LEP student's native language.

Vocational Skills Achievement (for Progress)

The vocational instructor develops course objectives based on the tasks derived through occupational analysis. Learning activities are then developed that enable the student to attain mastery of each objective. Finally, *appropriate* methods of assessment are designed.

Appropriate assessment requires the student to perform what the objective asks him or her to perform, under conditions specified in the objective. For example, if the objective calls for the student to "give a shampoo," appropriate evaluation would require the student to actually give a shampoo. Asking a student to *tell* how to give a shampoo or to *list* the steps in giving a shampoo would *not* be appropriate assessment because the performance during the assessment differs from the type of performance called for in the objective.

Objectives may call for cognitive, affective or psychomotor performances, or any combination of these behaviors. Therefore, a wide range of assessment techniques is utilized with special emphasis on demonstration of psychomotor performance. In the case of LEP students, directions and even test items are often provided in the students' native language. For a more complete discussion, see Bradley and Friedenberg (1982).

Assessment of VESL Achievement (for Mastery)

The purpose of an achievement test for mastery is mainly to assess the degree of success of the instructional program. Such tests are usually used for research purposes to evaluate the effectiveness of a certain type of "experimental" instruction and for specifically funded programs to determine success and whether funding ought to be continued. They are also useful to concerned administrators who want a general picture of how successful a given program is.

In a VESL program, the instrument used for assessing overall VESL achievement should be the same instrument used for the initial assessment of vocational English-language proficiency. If the measure is to be used for research or program evaluation purposes, the testing must be formal (e.g. like the BVOPT); however, if one only wants a general picture of program effectiveness, informal procedures, such as interviewing, may be employed.

Assessment of VESL Achievement (for Progress)

Assessing the achievement (progress) of vocational English is the most complex, yet the most important part of assessing LEP vocational students. It is also the part which has been least explored.

Like in general ESL courses, the content of the assessment instrument for VESL should be a reflection of the course. The content of a VESL course is based on language derived from the vocational content. Thus, a good deal of close collaboration between VESL and vocational instructors is essential.

Deriving Curriculum Content

The goal of vocational education is to prepare individuals for initial employment, upgrading, retraining and advancement in business and industry. It is essential, therefore, that vocational education courses be based upon the actual job requirements of business and industry. The most widely used method of determining vocational education course content is called *occupational analysis* or *task analysis*.

The effective use of task analysis to develop a vocational education course assures course content based upon the skills and knowledge currently required of workers in the occupation being taught. In addition, task analysis provides the ESL instructor with concrete data regarding the language that must be taught and, therefore, assessed. Following is a brief description of some of the important steps involved in planning vocational education programs. For a more complete description of the process, see Bradley and Friedenberg (1982).

The Job Description

The first step in analyzing an occupation is to obtain or develop a written job description. A job description is a general statement that defines the scope of the course by describing all of the kinds of duties a specific worker performs. A job description includes the various duties involved and lists any special or unusual conditions under which the duties are carried out (e.g., works out-of-doors, long periods of standing, heavy lifting, special clothing, and so forth).

The job description promotes a job-oriented point of view that guides the selection of only the most relevant subject matter and learning activities for the course. The job description also helps to communicate to students, community members, and other educators — including ESL teachers — the purpose of the course.

Task Listing

Many occupations require more skills and knowledge than can possibly be fitted into the time available in a particular course. The instructor in such a situation must make certain decisions. Which skills are absolutely essential for the entry-level job? Which skills are required a little later on the job? Which skills are expected only of the experienced worker? The answers to these questions help the instructor make informal instructional decisions.

Duties and Tasks. When analyzing complex occupations such as automobile mechanic, cosmetologist, and nurse, it becomes immediately apparent that workers in each of these occupations perform numerous tasks and that certain tasks are logically related to each other. It makes sense for the course developer, the instructor, to place these related tasks into logical groupings. One method of placing tasks in logical groups is to begin by first identifying the major *duties* of that occupation, then listing each task under the related duty.

A duty is one of the distinct major activities or units involved in the work performed in an occupation. For example, analyzing the automobile mechanic's occupation would reveal that certain activities can be categorized as diagnosing, others as repairing, replacing, servicing and so forth. Each of these broad categories is known as a duty. Each of these duties is composed of a set of related tasks. Figure 2 presents the duties of a cosmetologist.

After the duties have been listed, each of the tasks is identified and placed under the appropriate duty. A task is a unit of work comprised of a set of related behaviors (steps) that make up a particular portion of a job. For example, one of the tasks of an auto body worker is to replace fiberglass panels. All of the steps involved in replacing fiberglass panels serve to make up the complete task. Part of a waiter's job is to take the customer's order. Taking a customer's order, then, is one of the tasks that make up the waiter's job. Typing a business letter is one of the tasks of the clerk-typist's job. Figure 3 presents an example of the tasks that would be listed under one of the duties of a cosmetologist.

Figure 2

Duties of a Cosmetologist

Mixing Supplies and Sanitizing Equipment

Shampooing and Rinsing Hair

Cutting and Shaping Hair

Setting and Combing Hair into Style

Cleaning and Styling Wigs and Hairpieces

Treating Special Scalp Conditions

Curling and Relaxing Hair

Bleaching and Tinting Hair

Manicuring and Pedicuring Nails

Giving Facial Treatments

Keeping Records

Figure 3

Example of a Duty and Related Tasks

Occupation: Cosmetologist

Duty: Shampooing and Rinsing Hair

Tasks:

1. Give a Plain Shampoo.
2. Shampoo Lightened Hair.
3. Give a Mild Acid Rinse.

Some jobs such as key punch operator, retail store cashier, and window washer are ordinarily referred to as single-duty occupations and a task listing is sufficient. However, when analyzing a multi-duty occupation, the process is normally begun by first identifying the major duties of that occupation and then the tasks related to each duty.

The instructor develops a task listing by reviewing the formal written job description. Then, since actual work experience in the occupation to be taught is required of each vocational instructor, the instructor reviews his or her own occupational experience and the literature to

develop an initial task listing. However, for accuracy and up-to-date information, the instructor finds it essential to observe and talk with workers who are currently performing the job being analyzed. Thus, the task listing provides a list of all of the tasks workers are currently required to perform in the occupation being analyzed.

Task Detailing

Detailing is the process of listing each step required to perform a particular task. Drawing upon occupation literature, personal experience, and observation of and discussions with experienced workers actually performing the task, the instructor writes in sequential order every step required to perform the task. For example, detailing the task of baking a sheet cake would produce the following steps: read the recipe, gather the necessary ingredients and equipment, wash hands, cream the sugar, shortening and salt, and so forth. The purpose of detailing is to be certain that every aspect of each task is included in the instruction. Figure 4 presents the steps required for a cosmetologist to give a patron a plain shampoo.

Figure 4

Example of Task Detailing

Occupation: Cosmetologist

Duty: Shampooing and Rinsing Hair

Task: Give a Plain Shampoo.

Steps:

1. Select and arrange all materials.
2. Sanitize hands.
3. Greet patron.
4. Seat patron comfortably.
5. Ask patron to remove neck jewelry or ear jewelry and glasses.
6. Drape patron.
7. Remove any hair pins from hair.
8. Examine condition of patron's hair and scalp.
9. Brush hair thoroughly.
10. Adjust shampoo cape over back of shampoo chair.
11. Adjust volume and temperature of water spray.
12. Wet hair thoroughly.
13. Apply shampoo.
14. Give manipulations.
15. Rinse hair.
16. Apply shampoo again.
17. Give manipulations.
18. Rinse hair throughly.
19. Towel-dry hair.

When detailing a task, experienced instructors also often identify the major type of performance involved in each step. Many instructors simply classify the behavior as cognitive or psychomotor. However, a growing number use the behavior categories provided by Mager and Beach (1967): Recall, Discrimination, Problem Solving, Manipulation and Speech. Identifying the type of performance enables the instructor to ascertain that students will practice the kinds of performances required in the task. For example, if speaking to the customer is a step in the task, the instructor notes speech as the performance and makes certain that when learning activities are selected, students have the opportunity to practice the exact performance required on the job — speaking to the customer.

The task listing also often includes *who* performs the task (e.g. entry-level worker or experienced worker) and how often it is performed. Instructors also note the tools and materials being used while the tasks are being performed. Therefore, task detailing also provides a list of tools, materials, equipment and supplies that must be obtained for the instructional program. Figure 5 presents a list of the tools, materials, equipment and supplies obtained while detailing the task of giving a plain shampoo. However, the primary reason that the vocational instructor details each task is to assure that every step of every task is included in the instructional plan.

Figure 5

Tools, Materials, Equipment and Supplies List

Occupation:	Cosmetologist
Duty:	Shampooing and Rinsing Hair
Task:	Give a Plain Shampoo
	Sterilized Towel
	Sanitized Comb
	Sanitized Brush
	Neck Strip
	Shampoo Cape
	Shampoo (According to patron's needs)
	Hair Rinse or Conditioner

Task Analysis for LEP Students

A limited, but growing number of vocational instructors are professionally prepared in *bilingual* vocational education methodology. These trained instructors recognize the importance of the language associated with the safe and successful completion of every task. Therefore, during task detailing, these instructors identify the special language associated with each task.

In addition to listing the names of tools, equipment and supplies, these instructors note the technical terms and phrases related to each task. They also record the types of questions and directions workers respond to when performing each task. These instructors recognize that this "language of the trade" is an important part of the preparation of all vocational education students, but that for LEP students, acquiring the language of the trade is a survival skill.

The specially trained vocational instructor of LEP students uses the language of the trade in at least two ways. First, the instructor includes these terms, phrases, questions and directions in his or her instructional plan. Second, the list of essential terms, phrases, questions and directions is shared and discussed with the ESL teacher. A cooperative plan is developed whereby the ESL instructor teaches the needed language at appropriate times and the vocational instructor provides opportunities for practice and reinforcement of the use of the needed language as it is being used. Such collaboration between the ESL teacher and the vocational instructor is essential to essure that ESL instruction is, in fact, job-related and presented in appropriate sequence.

Unfortunately, only a limited number of both ESL teachers and vocational instructors are trained in the techniques of collaboration. However, it is not difficult for the ESL instructor to determine what and when language must be taught even when the vocational instructor is not trained in bilingual vocational education.

Deriving Content and Sequence for ESL

When the vocational instructor is trained in bilingual vocational education methodology, the essential language of the trade is determined through task detailing and shared during the collaborative process. When the vocational instructor is *not* trained in bilingual vocational education methodology, the ESL teacher determines ESL content and sequence through discussions with the vocational instructor.

The ESL instructor is aware that the vocational instructor has a list of the tools, equipment, materials and supplies required for each task. Thus, a lexicon is readily available. Task detailing sheets provide a list of the imperatives that must be learned and the order in which they must be learned. The ESL instructor begins the meeting with the vocational instructor by requesting a copy of the appropriate task detailing sheet and a list of related tools, equipment, materials and supplies (inventory). A few questions provide the balance of the needed information.

The ESL instructor asks the vocational instructor questions, such as, "What oral questions or directions will be directed to the student while performing the task?" "What will the student be expected to say (to customers, peers, supervisor) while performing this task?" "What must the student read while performing this task?" The answers to questions such as these added to the lexicon, task detailing sheet, course hand-outs and text provide the ESL teacher with a comprehensive knowledge of the language that must be taught. Sequence is then discussed and agreed upon.

The ESL instructor then organizes the language samples collected from the vocational teacher. These samples will contain safety expressions, technical terms, sub-technical terms, classroom-related expressions, non-technical words, and various grammatical structures.

It must be remembered that the vocational instructor, whether or not trained in bilingual methodology, is *not* a language teacher. The vocational instructor must receive guidance and suggestions from the ESL teacher regarding how and when to reinforce the language learning. Thus, frequent collaboration is essential.

Assessment Techniques

Just as the content of the VESL achievement test should reflect the content of the VESL course, the testing techniques should also reflect the learning activities employed during the course. These learning activities should reflect authentic on-the-job or vocational classroom activities as much as possible. Hence, VESL learning activities are communicative and require students to integrate language components for comprehension and production, as opposed to manipulative where students focus on discrete elements.

Up until recently, few ESL techniques had been developed which gave practice in integrating language components meaningfully. Now several language educators have exciting and innovative learning activities that promote communication. These activities include discussions, information getting, reporting, and description (Savignon, 1972); group dialogue (Farid, 1976); microcounseling (Friedenberg, and Bradley, 1981); and others.

If few communicative teaching techniques exist, even fewer integrative testing techniques exist. The challenge is even greater for VESL students, many of whom possess no literacy skills. The following sampling of activities may prove successful for VESL achievement testing which requires neither reading nor writing.

1. Oral Presentation — students explain processes or procedures.
2. Pictorial Multiple Choice — students choose correct drawing of equipment or process based on what they hear (taped or live).
3. Acting Out Situations — students role play real-life situations.
4. Paraphrasing — students reword an utterance they hear.
5. Physical Response — students carry out commands.

Conclusion

The assessment of LEP vocational students is comprised of five essential components: 1) Initial assessment of English language proficiency, 2) Initial assessment of vocational skills, 3) Vocational skills achievement, 4) English language achievement for mastery, and 5) English language achievement for progress. This assessment process is a uniquely challenging one which requires collaboration among the vocational instructor, the vocational counselor, and the ESL instructor.

This chapter focused mainly on the considerations necessary to develop the content for vocational ESL achievement tests (for progress). Like general ESL courses, the content and techniques of the assessment instrument should reflect the content and learning activities experienced in the course. Unlike general ESL courses, however, the content of a VESL course contains only the language necessary to survive in the vocational education classroom, laboratory, or shop and for success on the job. This content is derived from the task analysis.

Although slow is coming, bilingual vocational education and vocational ESL programs have become important developments in our field. They help language-minority youth and adults to enter the job market and gain self-satisfaction and adequate pay, and they contribute to the economic growth of our nation. In order to increase the effectiveness of these programs, we must strive to develop well-trained staff, suitable instructional materials, and assessment instruments that evaluate LEP students reliably and fairly.

References

Bradley, C. & Friedenberg, J. *Foundations & Strategies for Bilingual Vocational Education: A Handbook for Vocational-Technical Education Personnel.* Washington, D.C.: Center for Applied Linguistics, 19B2.

Farid, A. "Communication in the Classroom: Student-Improvised Dialogues." *TESOL Quarterly*, 10(3), 1976, 299-304.

Friedenberg, J. & Bradley, C. "Communication Skills for the Adult ESL Student: A Microcounseling Approach." *TESOL Quarterly*, 15(4), 1981, 403-411.

Mager, R. & Beach, K. *Developing Vocational Instruction.* Belmont, CA: Fearon, 1967.

Melton Peninsula. *Administrator's Manual for the Bilingual Vocational Oral Proficiency Test.* Dallas, TX: Melton Peninsula, 1981.

Savignon, S. *Communicative Competence.* Philadelphia, PA: The Center for Curriculum Development, 1972.

Strevens, P. "English for Special Purposes: An Analysis and Survey." In J. Ronayne Cowan (Ed.) *Studies in Language Learning: Special Issue on Language for Special Purposes*, 2(1), 1977, 111-135.

Part III
Research and Policy

TEACHING AND TESTING BOTH COMMUNICATIVE AND ACADEMIC SKILLS IN ENGLISH AS A SECOND LANGUAGE[1]

Muriel Saville-Troike
University of Illinois at Urbana-Champaign

The focus in second language research, testing, and teaching has shifted during the past decade from the attainment of grammatical proficiency to the acquisition or learning of "communicative competence." This is generally defined as including all of the knowledge about grammar and vocabulary considered in earlier theory and practice, plus knowledge about appropriate use of the language in naturally occurring social contexts, ability to use the language for desired purposes (for example, to request, command, or inform), and skill in such interactional tasks as initiating and maintaining a conversation.

In spite of this change of focus, however, there have been, to date, very few studies which have examined such things as (1) the development of communicative forms in a naturalistic setting, (2) the nature and relative effectiveness of different communicative tactics that limited English-speaking children use while they still have very limited linguistic means at their disposal, or (3) the relationship of growth in communicative competence to academic achievement when that must be acquired and evaluated primarily through the medium of English.

For years, we proceeded under the assumption that learning to use a language was a unitary phenomenon and that whatever proficiency was developed would serve both communicative and academic needs. James Cummins (e.g. 1980) and others have made us aware in recent years, however, that some students may become quite effective in interpersonal communication in a language, and yet apparently not possess the linguistic skills required to succeed academically through the medium of that same language. This is true in first-language acquisition as well as second-language learning, as we observe many children (and older students) who communicate quite effectively in the highly contextualized situations of face-to-face interaction, and yet fail in the context-reduced linguistic tasks required in reading, writing, and more advanced cognitive processing.

Now the general assumption in both our teaching and testing is that interpersonal communicative skills are lower on a developmental continuum than the skills for academic linguistic competence. We assume, for instance, that the communicative training we now include in ESL lessons will contribute significantly to our students' success in learning through the medium of English, but we remain generally unaware of what the linguistic demands on them actually are when they are in science classes, mathematics classes, and other content areas. And our evaluative scales including FSI-type ratings assume that ability to function effectively in academic contexts in a language presupposes ability to function effectively in interpersonal contexts as well. But this is not necessarily the case.

Because empirical evidence to support our assumptions about such issues is still quite limited, I wish to present some results from research I have been conducting with Erica McClure and Mary Fritz (cf. Saville-Troike et al., 1982 and McClure et al., 1982). Then I would like to draw some tentative conclusions from our findings, to date, and discuss their possible relevance to research methods, testing procedures, and classroom practices.

Our research thus far has addressed the following more specific questions:

1. What do children who are new to an English-speaking milieu need to accomplish in various communicative contexts during their first weeks and months of school?

2. What tactics do they use to fulfill these functions, and how do these develop over time? (By "communicative tactics" I mean the linguistic or nonlinguistic means at their disposal that children select to try to accomplish their communicative goals.)

3. What is the relative effectiveness of different tactics for accomplishing the children's intent?

4. And finally, how do communicative tactics and grammatical development relate to how well the children succeed in learning through the medium of English?

I would like to begin by briefly describing our population and procedures.

The population for our research consisted of twenty speakers of Japanese, Korean, Hebrew, Spanish, Icelandic, and Polish who were enrolled in a multilingual program during 1981-82, of whom 18 completed the school year. In the elementary school where we carried out this research, almost half of the students are native speakers of a language other than English, and all native English-speaking students are required to study a second language.

The children in our sample ranged in age from 7 to 12 years (grades 2 through 6). At the time of their selection at the beginning of the fall semester, all of the subjects

[1]The research cited in this report was supported in part by the Center for the Study of Reading, under a grant from the National Institute of Education, and in part by a grant from the University of Illinois Research Board. Their help is gratefully acknowledged.

(a) knew very little or no English, (b) were already literate in their native languages, and (c) had well-educated parents (professors or graduate students at the University of Illinois).

The twenty children were divided into two groups for daily 30-minute pull-out ESL instructional periods, according to their age and maturity level. I videotaped these sessions once a week for 30 weeks during the 1981-82 school year, and one of my colleagues observed and audiotaped the first ten minutes of each session another two times a week. As part of the research design (but in a situation also occurring naturally in previous years), the children were videotaped while without adult supervision or intervention before the ESL teacher came into the room, as well as during the ESL lesson. Additionally, we observed weekly in regular classrooms and in other school contexts, videotaped once a month in regular classrooms, and audiotaped and videotaped irregularly in other situations (e.g. on the playground, in the library and halls, and in native-language reading sessions).

We began our determination of the communicative tasks these students faced with a microanalysis of the ESL instructional periods. Videotapes were scripted, and then all communicative acts in sixteen of the sessions were coded for inferred speaker intent, using a system adapted from Dore (1978). This system includes such categories of communicative acts as requests, responses to requests, descriptions of facts, statements of rules and beliefs, acknowledgments which recognize and evaluate responses, organizational devices that regulate contact and conversation, and performatives. Our most significant modification was in extending the system to include nonverbal as well as verbal communicative acts; additionally, we have added request and response types which may be unique to language teaching situations (e.g. "Use a complete sentence."), and new categories for translation and interaction between children (e.g. invitation, acceptance and rejection of interaction, display behavior, and phatic communion with no propositional content).

Our first relevant finding is that children need or want to communicate for very different reasons in different types of social or instructional events, even within one time slot that is considered a single lesson.

The ESL lessons, for instance, divided themselves into a regular sequence of events that required a variety of communicative skills. The first was "unstructured" or "child-directed," as the children interacted with one another before the teacher arrived and called them to the instructional area. When she did, there followed an event we called "claiming a seat," during which the children challenged, claimed, and negotiated for their favorite position at a single large table.

During these two events, almost all communication that took place was between children and generally between children that did not share a common native language. Their primary reasons for communicating were either Interactional or Performative, and they had little need to request or give information of any kind. The most common communicative goals were engaging another child in play, teasing, protesting, calling attention to things, and claiming objects or territory. To achieve these goals, children do not really need language, as these children demonstrated. The form of most of these acts was entirely nonverbal, i.e. gestures and physical contact accompanied by non-speech sounds. It is interesting to note that there was almost no increase in the amount of verbalization used in these events throughout the year, even as the children developed proficiency in English. One significant change in the verbalizations that did occur was from use of the native language to protest and claim (even with children who didn't understand it) to such English routines as "Don't do" and "That's mine." Observations and recordings on the playground yielded the same finding: Children can and do participate in a great deal of social activity, even when fairly complicated rules are involved, with little or no language. Other children taught non-English speakers games by demonstrating what to do and by correcting their mistakes with a simple "No," or with physical intervention.

The earliest tactics that our subjects used to converse with one another can be illustrated with an episode that occurred in the third week of videotaping. This sequence occurred as three boys — speakers of Japanese (J), Korean (K), and Spanish (S) — worked together to assemble a puzzle map of the world. Conversational rounds consisted mainly of one child naming a referent and the other children taking turns repeating the term in round robin fashion.

S: "America."
(He found a puzzle piece of the United States and put it in place.)

J picked up S's hand and moved it out of the way.

J: "Japani, Japani."
(He pointed to the place in the puzzle frame where the Japan piece will go.)

K: "Japani."
S: "Japani."
J: "Japani."
S: "Japani."
J: "Ah. Japani."
(He found the Japan piece to the puzzle.)

"Korea."
(He also claimed to have found Korea.)

Adding only "No," and using different stress and intonational contours, the children disagreed and questioned one another.

K: "No Korea."
J: "Korea."
K: "No."
J: "Korea."
S: "Korea?"
K: "No."

S: (Spanish) "Si?"

The Korean boy offered an explanation, which the others acknowledged by repeating it after him.

K: "Yes, yes, yes.
No Korea. Pink Korea." (Stress on "pink.")

"Pink" referred to the color of the Korean map piece: i.e., "It can't be Korea, because Korea is pink."

J: "Pink."
S: "Pink."
K: "Korea."
(He found the piece and put it in place.)

The ESL teacher entered the scene, telling the class to put away their games. The Korean boy unsuccessfully tried to protest, saying to her in Korean (which the teacher doesn't understand):

"Ahl I have to finish."

This sequence concluded with more repetition and nonverbal tactics as J and K leaped up and ran toward the table, and the teacher gently stopped K with her arm around his neck and shoulder.

K said "He," as he pointed toward the table, protesting that J got away with running to claim a chair. He then used the most complex English structure of this whole sequence when he yelled to J, "Hey, walking."

J responded in Japanese, "Damare!", which means "Shut up."

It is apparent that from the beginning these children have conversational skills for attention getting and turn taking, and even with minimal language, relate their moves in a coherent manner to the form and content of the one that precedes it. In Grice's (1975) terms, they know the cooperative principle of conversation, and are "relevant."

The earliest communicative forms available to these children (as for children acquiring a first language) are nonvocal behaviors, sound play and nonspeech sounds, simple routines, repetition of part or all of previous utterances, and single referential terms they are beginning to learn in English (cf. Keenan, 1974; Peck, 1978). Additionally, these second language learners have access to first language forms and tactics and have had experience interacting with other children.

This pattern held true for all of the children in this study who did interact socially with other children, but some of our subjects did not. I will return to this point to discuss the relationship of peer interaction to language development.

A second type of event during the ESL sessions was highly structured and teacher-directed, including such opening routines as "What day is it today?" and the focal instructional activity (usually introduction of new vocabulary or new grammatical structures). During these events, almost all communication that took place was between child and teacher. The aim of most of the children's communicative acts was to respond to a question or request from the teacher. The only other acts that occurred with any frequency were children's bids for the teacher's attention. The teachers' requests changed in form through the year from requests to repeat after them and identify single lexical items to yes/no questions and WH-questions. The children's responses matched these changes, going from primarily repetition, to nodding yes or no, to single words, to (primarily memorized) phrase and sentence patterns. Observation and recording in the children's regular classrooms indicated that this same participant structure occurred in small group instruction where all students in the group were at a similar level in their language proficiency, especially for reading, or when the regular classroom teacher worked with students individually, but this seldom accounted for more than another 20 minutes out of their 5 1/2 hour school day.

Occasionally we observed instances where the children's facility in responding to such questions (combined with the skill in learning procedures nonverbally from other children that I have already mentioned) misled an adult in the regular classroom into thinking children understood much more English than they actually did. Early in the year, for instance, children learned the names for shapes in ESL. Back in his classroom, one very limited-English speaker was building an elaborate pyramid for an art lesson by gluing toothpicks together when he was asked, "What geometrical shape is that?" He certainly understood only the word "shape," but appropriately responded, "Triangle." When the adult working with him was asked to evaluate his English at the end of the art lesson, she reported that he was quite fluent.

The third type of event during the ESL session was semi-directed by the teacher, but included children addressing each other as well as her. These events were characterized primarily by the need for a *real* exchange of information and constituted the largest time block in the classes we observed. This included the teacher asking the whereabouts of absent children before the structured lesson began and less structured follow-up activities, usually including explanation and implementation of an art project or game. It was only in this type of event throughout the year that the functions of the children's communicative production regularly included the categories of Requests, Descriptions, and Statements, and it was only in this type of event that they went beyond memorized patterns as they

struggled to obtain needed information, or to express themselves to the teacher and other children. The results were often ungrammatical, but it is interesting to note that grammar was never corrected when *real* communication was at stake. Successful communication, in fact, often included a combination of English, nonverbal miming, and the speaker's native language. A preliminary analysis of the correlation of the language forms presented in the structured lessons and the language forms actually used by the children for less structured communicative purposes shows little carryover except for vocabulary.[2] The "is" of "This is a pencil," for instance, was generally not used for weeks or months in other contexts, and (along with articles and past tense or plural inflections) is still absent for some who have by now achieved considerable fluency in both communicative and academic uses of English.

There was of course intragroup variability. A few of the children sought interaction only with the teacher during these events, and a few almost entirely with other children. When another child in the group spoke the same native language, some fulfilled these goals primarily through that language rather than English, especially during the first half of the year. Again, I will return to this point when I discuss the relationship of peer interaction to English language development, and to academic achievement through the medium of English.

Observation and recording during the rest of the children's five and one-half-hour school day showed much more carryover of these communicative needs and skills to other learning situations, with some interesting similarities and differences in tactics and results. Children who preferred to request information and express themselves in their native language continued to do so, either seeking out a bilingual adult or making use of bilingual classmates to explain directions that had been given in English. Not infrequently, these limited-English speakers joined forces to figure out problems they could not understand. Translations of the non-English portions of our recordings indicate a very high percentage of native language communicative acts was directly related to the assigned academic tasks. Some children used their native language in this context even when there were no bilingual adults or children in the class who shared their language, verbalizing extensively to themselves as they worked. Standardized achievement tests administered to all students in the school in English at the end of the year (the CTBS) show us in retrospect that this language preference group includes three of the five children who scored highest in English reading and science.

In contrast to the similarities in early conversational tactics, children used basically different means for describing events and actions during the early stages of learning English. One tactic was to use a single lexical term (usually the name of the topic the child wanted to convey information about). Completion of the proposition required a cooperative addressee, who asked one or more questions that allowed the child to respond with a simple "yes" or "no." This tactic was used more frequently with adults than with other children.

The following sequence, which exemplifies this tactic, was a Japanese speaker's (J) response the teacher's (T) question about why another child had not come to class.

J: "Picture."
T: "She's making a picture?"
J: "Yes."
T: "Will she come when she's finished?"
J: "Yes."

The same cooperative process continued for the construction of longer texts even after children could express a whole proposition themselves by stringing together multiple terms in English, with adults first providing expanded models, and later prompting expansions by asking WH- questions. This rarely occurred in child-child communication, and never between limited-English speakers.

The second tactic was to create a complete topic-comment proposition by using both verbal and nonverbal means, usually naming the topic in English and then providing a nonverbal comment.

In the following example, a Korean boy (K) named both agent and object as he communicated to the teacher (T) that Taki had thrown a pencil at him while passing out supplies.

K: "Teacher. Teacher."
(The teacher went over to him.)
K: "Taki. Pencil."
(He made a throwing motion.)
"Taki. Un ... he."
(He made a throwing motion again.)
T: "Taki, please don't throw pencils."

Most children used this tactic more frequently with other children than with adults, but a few used it almost exclusively for any situation. The general pattern that developed for longer narratives by children who preferred this tactic was: (1) getting attention (e.g. "Look, look"), (2) indicating focus (e.g. "this," "right here," tapping the object that was in focus), (3) naming objects, and (4) acting out the event.

Students who developed considerable early competence in communicating with a combination of English and miming to other limited-English-speaking children and the teachers in ESL did not have the same success with teachers in their regular classroom. This tactic, however, proved to be very successful for communicating with English-speaking children, and was used by all of the children who from the beginning

[2]The ESL teachers observed in this study made a particular effort to incorporate terms in their lessons that the children needed to know for content areas taught in English in their regular classrooms.

of the year interacted most with English-speaking peers at school. But this group included *none* who subsequently scored highest on the CTBS.

This was a surprising finding since our original assumption was that children who interacted the most socially with other English-speaking children would learn English faster, and thus do better in English-medium instruction; this turned out not to be the case. In fact, along with the three strong native-language-preference students who scored among the top five, the other two top achievers rarely spoke at all to the other children during the ESL sessions that were audio and videotaped. Even in their regular classrooms and on the playground, they appeared to make minimal use of their developing English for social purposes.

On the other hand, some of the other children who did not engage in much verbal interaction with their peers proved to be among the poorest in their language development and academic achievement in English, so I am not suggesting any simple correlation and certainly no cause-effect relationship. A striking observation, nevertheless, is that some of the more successful communicators, at least in the contexts we recorded, fossilized at fairly early levels of development, suggesting that their very success may have reduced their motivation to learn more complex linguistic forms. This fossilization was particularly evident among those who continued to combine verbal and nonverbal actions to express propositions.

In addition to the types of communicative events I have just described, two other kinds of instructional events were frequently observed and recorded in regular classrooms, but never in ESL (in part because of the smaller class size), and it was in these that we saw most of the limited-English-speaking children entirely unable to cope: staring out windows, doodling, poking their neighbors — one even crawling under his desk in retreat.

These are the teacher-to-whole-class participant structure, where the group was being talked to (or read to) as a whole, and its opposite, the fully independent work activity period, where students were given written or oral instructions in English and expected to proceed without additional interaction with adults or peers. The amount of time spent in these types of events varied greatly from class to class, but occurred to some extent in all. In the classrooms we observed through the year, these kinds of events constituted from ten to over fifty percent of the instructional day. The percentage generally increased with grade level, and I feel sure we would find students engaged in — or disengaged from — even more of these kinds of instructional situations if our research had continued into junior high and high school.

In spite of the very similar nature of our subjects' socioeconomic status, family educational level, prior literacy in their native languages, and even type of residence and play facilities (they all lived in the same university-owned housing complex), by the end of the year they differed greatly in the degree to which they had mastered English and in their level of accomplishment in reading and other subjects taught and tested through the medium of English. As I summarize the final results of our study, important questions to keep in mind are: What did ESL contribute to the achievement of the successful students? And what more can it do?

First, most children do not have to be taught to communicate with one another; while social interaction between them is certainly to be encouraged, we cannot depend on that alone for developing English language skills.

Second, there is a qualitative difference between the communicative tactics and skills children find effective for meeting their social needs and goals and those that are necessary for successful academic achievement in the classroom. As teachers of students who must learn to learn through the medium of English, we should be concerned that academic achievement — not just the learning of English — has a clear priority in our curricula.

Third, the portions of our lessons which focus on structural patterns, especially on English morphology, appear to have least applicability to students' immediate academic needs. Not only do most beginning students not use the grammatical inflections when attention is on receiving or imparting information, they really don't need to. One example is provided by a third grade student in our sample, a Japanese girl, who scored at the 65th percentile of national norms in reading on the CTBS after only one year of English, and above the 90th percentile in all content areas in the test battery. Yet, she still systematically omitted plurals, articles, and tense markers in her spoken English, as did four out of five others among the top six academic achievers. On the other hand, the Polish, Icelandic, and Spanish speakers ranking 13th, 14th, and 17th out of 18 in achievement scores had a far better mastery of English morphology because of positive transfer of inflections from their native languages.

This is also quite relevant for testing and placement, since we found no significant correlation between students' scores on the Bilingual Syntax Measure (which emphasizes morphology) and their ability to achieve in English-medium instructional contexts. The only language factor that does appear to be critical is vocabulary, which usually receives woefully inadequate attention in ESL teaching and testing. The determination of which vocabulary is critical for specific students requires closer coordination with regular classroom teachers than most specialized ESL teachers maintain and a better knowledge on their part of the content areas of the curriculum.

Fourth, as I reported earlier, the lowest academic achievers in our sample were among the most successful at interpersonal communication, especially with other children. Academic success requires competence in using and understanding language in

context-reduced situations in which students cannot rely on nonverbal elements of communication. The language skill which is most likely to develop this competence is writing, yet it is sadly last not only in the listening-speaking-reading-writing sequence we recite, but also in the time and attention allotted it in most ESL classes.

The few students in our sample who could cope with independent instructional activities in regular classrooms possessed skills that were not taught in ESL, but could be. One important skill they shared, for instance, was their ability to make good use of a bilingual dictionary. One mother had taught her child this skill by encouraging her to write sentences in her native language and then look the words up and translate the sentences into English. This yielded ungrammatical sentences, of course, especially since the mother did not speak enough English herself to provide correction (Koda, 1982). Still, the skill proved to be very useful in school and might well be added to elementary-level ESL objectives for the benefit of children whose parents do not provide this instruction.

Finally, I would like to emphasize again that most of the children who achieved best in content areas, as measured by tests in English, were those who had the opportunity to discuss the concepts they were learning in their native language with other children or adults. Even in linguistically heterogeneous classrooms such as those in our research, at least some degree of bilingual education is proving to be feasible and clearly provides the best context for conceptual development — and for learning English.

I began by saying that we in TESL have shifted our focus from grammatical competence to communicative competence, and I consider this a positive development. But I must close by saying that this still falls short of our responsibility to students who must learn to learn through the medium of English — which includes most of our students. We need to develop their *academic competence* as well, and this calls for even more changes in our priorities and in our procedures. Developing communicative competence alone is a desirable but insufficient goal for English teaching, and — ironically — may even interfere with academic achievement. We must begin to place more emphasis on vocabulary learning and less on grammar. (You will note I have not even mentioned pronunciation.) We must recognize the value and importance of the native language as a basis on which to build the learning of English. And above all, we must realize that for our students, learning English is never an end in itself, but always a means to an end. We must strive to keep our students' needs uppermost in mind in determining our own priorities.

References

Cummins, James. "The Construct of Language Proficiency in Bilingual Education." In J. Alatis, ed., *Current Issues in Bilingual Education*. GURT. Washington, DC: Georgetown University Press, 1980.

Dore, John. "Requestive Systems in Nursery School Conversations: Analysis of Talk in Its Social Context." In R. Campbell and P. Smith, eds., *Recent Advances in the Psychology of Language: Language Development and Mother-Child Interaction*. New York: Plenum Press, 1978.

Grice, H. Paul. "Logic and Conversation." In P. Cole and J. Morgan, eds., *Syntax and Semantics: Speech Acts*. New York: Academic Press, 1975.

Keenan, E. "Conversational Competence in Children." *Journal of Child Language* 1 (1974), 164-183.

Koda, Keiko. English Language Acquisition by Four Japanese Girls: Case Studies and Influencing Factors. Unpublished manuscript, University of Illinois, 1982.

McClure, Erica, Muriel Saville-Troike, and Mary Fritz. "Children's Communicative Tactics across Language Boundaries." In *Proceedings of the Chicago Linguistics Society*. 1982.

Peck, Sabrina. "Child-Child Discourse in Second Language Acquisition." In E. Hatch, ed., *Second Language Acquisition: A Book of Readings*. Rowley, Mass.: Newbury House, 1978.

Saville-Troike, Muriel, Erica McClure, and Mary Fritz. "Communicative Tactics in Children's Second Language Acquisition." Paper presented at the University of Wisconsin Symposium on Universals of Second Language Acquisition, 1982.

PROMOTING CONCEPT AND LANGUAGE DEVELOPMENT IN THE CLASSROOM

Rodolfo Jacobson, Ph D
University of Texas at San Antonio

Introduction

The route to choose in order for a child whose first language is not English to become truly competitive in a setting with children who are native speakers of English is neither the hasty pullout from a bilingual education program nor the premature submersion, labeled "immersion," into an English only mainstream setting. The answer in turn, seems to be in the successful association of two different but by no means contradictory, goals: the achievement of the threshold level in L_1 and a stable as well as balanced distribution of L_1 and L_2 while conceptual learning takes place. It is the objective of this paper to elaborate on these two goals and to report on some early results from the implementation of a project — the Title VII Demonstration Project in Bilingual Instructional Methodology — which precisely seeks to test the validity of this hypothesis.

Prior Research. The author's research leading to the instructional design of the Project funded by the U.S. Department of Education goes back a number of years including consultant work and research activities in Laredo and San Antonio, Texas (1975-82). The findings that resulted from these activities have appeared in journals and anthologies, and it may be in order to briefly summarize these to lend greater credit to the objectives of the ongoing project.

The author's studies during the last decade have focused on three topical areas: (a) community verbal behavior with emphasis on Spanish-English codeswitching, (b) Chicano English as an emerging speech variety of American English, and (c) bilingual instructional methodology with emphasis on language distributional aspects. The first and the last topic turned out to be most relevant to the formation of the program design, since the former determined the extent to which the two languages could and should be jointly utilized in the teaching of bilingual children and the latter led to the development of a bilingual teaching strategy knows as the *New Concurrent Approach* which upgrades the somewhat questionable Concurrent Translation Approach that underwent some criticism by scholars like Theodore Andersson, Bruce Gaarder, Christiana Bratt Paulston[1] and others.

The author's studies on codeswitching comprised "The Social Implications of Intrasentential Codeswitching" (Jacobson, 1977), "Anticipatory Embedding and Imaginary Content" (Jacobson, 1978b) and "Consensus and Divergence in Bilingual Co-occurrence Patterns: Some Mexican American Versions of 'Tis a Night before Christmas" (paper delivered at the UWM University of Wisconsin at Milwaukee, Linguistics Symposium, 1978, unpublished). The need to adapt the community switching to and/or limit it for classroom purposes became evident from the very beginning and the various papers on bilingual methodology emphasize the importance of avoiding any switching inside the sentence and developing a set of strategies designed to make these intersentential switching events purposeful, that is, to render them linguistically, psychologically and sociologically justifiable.

Following the publication of his paper presented at the 1975 TESOL (Teaching English to Speakers of Other Languages) conference, the author has attempted a constant refinement of the language alternation strategy until a satisfactory model would evolve that lent itself to demonstrate the effectiveness of the cited approach. The publications that reflect this gradual emergence of an approach controlling the dual language use of teachers and students show that what is sought here is neither random nor confusing, but rather sociolinguistically relevant as it reflects the norms found in multiple settings where two languages are in continuous contact (Jacobson, 1975a, 1975b, 1978, 1979a, 1979b, 1981, 1982a, forthcoming).

Threshold Level and Second Language Acquisition. The educational research in Canada (Cummins, 1980) has offered abundant evidence for the need of developing the child's home language and helping him/her to reach in it a threshold level that then permits the learner to gain proficiency in the school language. The analogy of the iceberg in attempting to conceptualize the problems of L_2 acquisition is by now widely known (Cummins, 1980: 29-36). Equally well known are the acronyms CALP and BICS in picturing the acquisition process controlling the success or the failure of the child expected to learn in school through a language other than his native language (Cummins, 1980: 28-40). Both arguments show the urgency of developing L_1 in the child and making him/her literate in it so that the child may transfer to L_2 the deeper cognitive skills that he/she has developed in his/her first language.

The question however arises whether L_2 should be acquired only after the threshold level in L_1 has been achieved or whether both languages can be acquired or developed in an almost simultaneous fashion. An L_1/L_2 consecutive acquisition is hardly viable in a setting like the U.S. where early exposure to English is unavoidable and the learning of the mainstream language is usually

[1] The difference between the NCA method and its earlier version, the Concurrent Translation Approach is still poorly understood. The opposition to the latter is merely transferred to the former without further study. See Christina Bratt Paulston's introduction to a collection of papers delivered recently in Racine, Wisconsin (Valdman, Hartford, Foster, 1982: XIII).

equated with mainstreaming and upward mobility. It has been rewarding to see that simultaneous or almost simultaneous acquisition is now also considered a possible option even for the Canadian setting (Cummins, this volume). This author has been certain — for some time now — that such a near concurrence of L_1 development and L_2 acquisition is most feasible and even necessary for settings where the pressure of learning L_2 is as formidable as in the U.S. In other words, the home language can be developed effectively through dual language learning tasks if one keeps in mind that the threshold level in the first language must eventually be reached.

As this latter objective is being pursued, the language of the majority is also learned producing in the child an early bilinguality that then permits the parallel development of his/her two languages. This early bilinguality, however, is not only the result of the child's exposure to two languages and his/her dual language development but also — and to a very considerable extent — to the notion of stability in language distribution that characterizes the truly bilingual person. As a matter of fact, the balanced distribution of his two languages, that is, the use of as much L_2 as L_1, denotes equal language status or prestige. Hence, any L_2 increase in the bilingual classroom that is accompanied by a decrease in the use of L_1 — however gradual this may be realized — will always be interpreted as a "put down" of one's vernacular, a situation that will ultimately lead to the apparent or even the actual loss of the home language.

Examine the following table that illustrates that instructional design observed in the K level of the cited Federal Project. Regardless of its identification as comparison or treatment classroom, the amount of Spanish (L_1) used to teach the children is such that the threshold in that language can eventually be reached. On the other hand, only the treatment kindergarten demonstrates the stability of language distribution discussed before (see above) as its ratio is a steady 90% to 10% in favor of the home language. The 10% of second language instruction, furthermore, is restricted there to language arts (ESL). The comparison kindergarten, in turn, shows the decrease-increase pattern as the ratio of 90%:10% in favor of Spanish at the beginning of the year is gradually changed to a 75%:25% ratio. Within this 25% of English, not only language arts but also math, science or health can be taught. This introduces the child to the trend in conventional bilingual education where the teaching in the vernacular is decreased as the teaching in English is increased. It also introduces the child to the separation of the two languages on the basis of school subjects, as such is found in every traditional program.

Table 1

Comparison		Treatment	
English	Spanish	English	Spanish
	90-75% of time		90% of time
10-25% of time (lang. arts & content)		ESL (10% of time as lang. arts)	

Answers to questions that are addressed in a project implementing such a design are of great importance to any bilingual instructor and likely to provide us with some interesting data as to what does and does not work in bilingual education. Some of these questions might be as follows:

1. Do all the children at the K level eventually reach the threshold level in their home language?
2. Does the decrease-increase pattern in the comparison classes, in effect, produce a feeling of English language superiority and Spanish language inferiority?
3. Does the stability of distribution in the treatment group, in turn, give the child greater security in regard to his/her attitude to English and Spanish?
4. Does the use of English in science-related subjects give the child the impression that only English can handle these areas of content?
5. Is the limited use of English in the treatment classes unrealistic for a child who is expected, in the following year, to receive half of his/her content teaching in that language?

Without any doubt, the kindergarten level presents itself as the ideal stage to set the groundwork for later achievements. It is here where we can develop further the language that the child brings from his/her home and, at the same time, attend to his/her language attitude formation. Since no specific academic achievement is required at the pre-primary level, there is sufficient time not only to use the home language in the classroom, but also to correct any biased perception that the child may hold of the language of the home or any overrating of the majority language that the child may feel to exist. The prejudices that the child does bring to school are rarely deep-rooted and can thus be removed quite easily so that he/she begins his/her

primary education with few, if any, psychological problems.

Concept Development through the New Concurrent Approach

The *New Concurrent Approach* has been defined and explained by the author in a number of recent as well as in forthcoming publications. "The New Concurrent Approach," he argued in a paper delivered at the Ethnoperspectives III Conference in Ypsilanti, Michigan (Jacobson, 1981: 14-29):

> resulted from the author's desire to bring together the child's two languages in a way that would further the latter's language development and, at the same time, lead to satisfactory school performance. To accomplish this objective, he had to address several issues:
>
> (1) the extent to which the native language must be developed in order to succeed in learning a second language; (see above)
>
> (2) the extent to which the home language should be used in school to develop a positive attitude toward it;
>
> (3) the extent to which first language maintenance in the primary grades would not interfere with the transition to English in post-primary education;
>
> (4) the extent to which the use of both languages would lead to an understanding of the bilingual functioning of some sectors of our society;
>
> (5) the extent to which school subjects would be learned through two language media.

In elaborating on this last issue, the author has emphasized that, in the primary grades:

> School subjects other than language arts utilize concurrent teaching techniques in which the teacher shifts smoothly from one language to the other as the lesson is carried on. This alternation enables the teacher jointly to develop the bilingual child's two languages without jeopardizing his school progress. Thus, concepts are forced or reenforced in both languages, the lexicon is expanded and many other meaningful activities go on in both languages simultaneously.
>
> (Ibid., 18)

In the actual proposal submitted to the Office of Bilingual Education and Minority Languages (OBEMLA), U.S. Department of Education, the author has attempted to define the approach more concisely as:

> A strategy through which the bilingual teacher teaches the school curriculum (except language arts) in the child's two languages by using both languages concurrently, that is, switching from one language to the other as the teaching/learning situation may require.

"Each switching instance," he continues saying, "shall be pedagogically justifiable in light of four criteria:

> (1) the two languages are distributed at an approximate ratio of 50%-50%;
>
> (2) the teaching of content is not interrupted;
>
> (3) the teacher is conscious of his/her alternation between the two languages; and
>
> (4) the alternation accomplishes a specific learning goal."
>
> (Jacobson, 1981: 14)

The intersentential switching for which the author is arguing is based on the utilization of a framework known as the *System of Cues*. "Cues, in NCA (New Concurrent Approach)," explains the author in the "The Role of the Vernacular in Transitional Bilingual Education" (in Hartford, Valdman, Foster, eds., 1982: 183):

> are signals that the teacher identifies in her class and that she wishes to respond to. As an enthnographer seeks to gather all the elements of a social situation and then arrange them in a meaningful order, so the teacher identifies the various aspects, pedagogical and social, that she feels she must address. These elements, or cues, are categorized depending on their relevance to classroom management, language development, curriculum, or teaching-student interactional norms.

Table 2 lists these pedagogical areas and the corresponding cues.

Table 2

System of Cues

(1) Classroom Strategies	(2) Curriculum
a. Conceptual Reinforcement	a. Language Appropriateness
b. Review	b. Content
c. Capturing of Attention	c. Text
d. Approval/Disapproval	

(3) Language Development	(4) Interpersonal Relationships
a. Variable Language Dominance	a. Intimacy-Formality
b. Lexical Enrichment	b. Courtesy
c. Translatability	c. Free Choice
	d. Fatigue
	e. Self-Awareness
	f. Rapport

The planned language alternation, as discussed in the quotes above, is serving as the teaching method in the treatment group of the cited Title VII Demonstration Project. The dual language use in the comparison group, on the other hand, is conventional as there, the two languages are strictly separated on the basis of content, that is, social studies, music, art, and physical education are taught in Spanish, whereas math, science and health are taught in English. It is, in particular, in the treatment group, that the close correlation between language development and conceptual reenforcement is observed. As a concept is taught, the teacher assesses the need for further language development, whether in the area of phonology, grammar or lexicon. By the same token, as the language is being developed during content classes, the specific concept to be taught is never disregarded.

As for the switching from one to the other language during content classes, three reasons can be cited that justify the dual language use in class. It is justifiable on pedagogical grounds as it promotes desirable language behavior in each one of the child's languages. It is equally justifiable on psychological grounds as it promotes the status of the home language and places both codes on equal grounds. The distributional balance found in the 50-50 ratio and the utilization of both languages in the teaching and learning tasks cannot but convey the message that nothing is learned in one language that cannot also be learned in the other. Finally, the switching is justifiable also on sociolinguistic grounds in view of the similarity that exists in respect to certain community behavioral patterns that are being observed. Obviously, the difference between intrasentential and intersentential switching is merely one of degree and as long as alternations do occur, they will impress the child as familiar strategies present in home and neighborhood.

In addition to the theoretical justification on whatever grounds, it must be ascertained that the NCA Method is being implemented successfully. The prerequisites of successful implementation are, however, such that they require special training either in the form of a university course[2] or as part of inservice workshops. Of four such prerequisites, two have been discussed above when a working definition for the approach was proposed and when the system of cues was suggested as a viable framework for triggering language alternation with a preconceived objective in mind. To be more specific, the author has contended that the success of this method depends on the *balanced distribution* of the two codes and the ever-present *goal conciousness* as the teacher responds to the cues that he/she has been trained to identify and to react to.

A further prerequisite could be labeled *self-monitoring of dual language use*. It is the need for the teacher to be able to detach him/herself from the teaching task and monitor the relative time that he/she stays in either language, as well as the extent to which he/she observes the educational objectives agreed upon in advance. In other words, was as much Spanish as English used in teaching, say, a social studies lesson, and were the language switches actually geared to tasks, such as, conceptual development, lexical enrichment, the notion of language appropriateness and so forth?

More from the methodological perspective, an effort shall be made to convey *continuity* in whatever content a lesson is taught. The language alternation must not be disruptive. As a matter of fact, planning an NCA lesson differs from preparing a lesson in a single language only in the language medium, not in the content nor the teaching strategy.

The strict observance of these prerequisites, i.e. (1) balanced distribution, (2) continuity of class performance, (3) self-monitoring of dual language use, and (4) goal consciousness, all contribute to the effective use of intersentential codeswitching in the classroom allowing for concept development and dual language growth.

Community Behavior in Contact Situations: A Rationale for NCA

It may be in order to elaborate somewhat further on the relationship between the NCA method and language distributional patterns in bilingual settings. Wherever two languages are in continuous contact, three situations may prevail. Only one language is spoken at a time for reasons of language appropriateness or out of courtesy to the L_1 or L_2 dominant member of a given bilingual community. Accordingly, a balanced Spanish-English bilingual will address interlocutor A in Spanish only and interlocutor B in English only.

More frequent than a unilingual pattern of this nature is the switching of codes, whenever pertinent, such that segments of both languages occur either in the same sentence (intrasentential) or from sentence to sentence or between two discourse episodes (intersentential). To summarize, the bilingual's distributional patterns to use in his/her bilingual community are of three kinds:

(a) unilingual — one language at a time, but not both;
(b) intersentential — one language in one sentence and the other language in the next;
(c) intrasentential — both languages in the same sentence.

Recent research on codeswitching has focused mainly on *(c)* although some data have also been gathered — especially in the author's research on *b* (Jacobson, 1977). The following examples from the study "The Social Implications of Intrasentential Codeswitching" *(ibid.)* illustrate these two switching strategies.

[2]The course Special Problems: "Sociolinguistic Approaches to Bilingual Instruction" has been offered at the University of Texas at San Antonio since 1979 and provides exposure and practice in NCA methodology.

Table 3

INTERSENTENTIAL	INTRASENTENTIAL
1. —It takes — *Es más despácio la manera esa.* (EC-24.14-15) (False Start)	1. —And I tell you another thing que I'd shoot body . . . (FM-1.14) (Substratum)
2. —. . . I wished they'd come more often! —You ought to get on the phone—y *dijo mamá que vinieran a visitar, ¿ves?* (EC-3.20-23) (CODE: Initiation of response)	2. —I lose my temper *porque a mí me da mucho coraje* . . . (FM 1.15-16) (Emotion)
3. —Sabe lo que me gusta a mí ves! *Man! That's all kinds of beers ahí!* (EC-25.10) (CODE: Continued speech after switching)	3. —. . . she would tell me things like—*este*—you know (FM-5.11) (Hesitation)
4. —No he podido grabar la conversación. Qué es que dijo? —*I was unable to record the conversation.* (RJ—memory) (CODE: Clarification)	4. —. . . ¿Le borro y le pongo *speak?* (FM-3.1.-2) (CODE: Prior code use)
5. ¿No? —Si. *He is going to be training for a manager right now.* (EC-6.7-9) (Employment Domain)	5. —Ah, no—it's not a sound problem, it's more of a like *como donde acentúa uno la palabra.* (VC-12.9-10) (CODE: As topic)
6. —Well, you do a lot of P.R. *Cuando vienen las mamases muy bien, tienes que calmarlas.* (VC-1.3-5) (Culture: Persons)	6. —. . . se no 'biera ella dicho eso, *We would take for granted that it was the last Saturday?* (CODE: Anticipatory embedding) (Specialized terms)
7. —(Wife to mother) Con la misma chaqueta por cuatro años. —(Wife to husband) *You are the one wearing the same jacket.* (EC-7.19-21) (between spouses).	7. —"How long have you been here?" *Pos le decía* "twenty-nine, thirty years." (EC-10.1-2) (CODE: Quote)
8. —¿Me aprobó mi sopa? *Ah, that's good.* Este no es macaroni de la bolsa. (METHAPHOR: For contrast)	8. —. . . un hombre precabido vale por dos.—Ahora digo yo si toca lo de malas, *OK but I did what I could to prevent it. (FM-2.5-7)* (CODE: Precoining)
9. —. . ., íbamos allí siempre cada año. *We went there every year.* (VC-6.506) (Metaphor: For emphasis)	9. —*Mis sobrinas* are the typical—you know—they can understand it. . . (FM-4.12-13) Home/Family (DOMAIN)
	10. —Oh, si. Si porque *I notice* que *if I write something down I can remember it better.* (ED-28.7-8) (School Domain)
	11. —It was the day you went *al parque.* (R-1.6) (Culture: Environment)
	12. —Cuando comenzó esto todo lo que andaban mechudos, they believe that they were no good . . . (EC-9.16-18) (CULTURE: Attitude/Bias)
	13. —Los doctores que vienen de México igualmente. hacen lo mismo *after being here for a while.* (FM-6.10-11) (CULTURE: Language-Locales)
	14. —I went only for one sole reason to Mexico—*porque no me contaran como era.* I went when I was twenty *a la capital.*, (FM-9.22-23) (CULTURE: Heritage)
	15. —Ah...bueno, como me estás grabando—*I'll take the fifth amendment*—on that one. (VC-1.3-5) (CULTURE: Social/Political Institution)
	16. —And maybe it's part of the culture, too, because, you know, like *con nosotros "¿qué?"* you know, *se oye muy mal.* I mean, *en mi casa todo el tiempo* es "mande." (VC-13.19-21) (CULTURE: Language)

140

Given the presence of both strategies in our community and one of them being acceptable on educational grounds, this language alternation strategy, if transferred to the classroom, could accomplish one of the most cherished but also least attained goals, not only in bilingual education, but in all education settings: that of bringing closer together the school and the community. Hence, the use of intersentential codeswitching in the NCA class makes an important contribution toward balancing two communicative competencies — that of Spanish and English, both of which must be brought to a common denominator if one wishes to improve the self-image of the bilingual minority child.

The Long-Term Objectives of the Demonstration Project

One can identify in the cited project at least six distinct strands or objectives which, when interrelated with one another, represent the long-term goal of the Title VII Demonstration Project in Bilingual Instructional Methodology. More specifically, the author has attempted to address in it issues that can be identified under the following headings:

(1) Methodology,
(2) Ethnography,
(3) Pedagogy,
(4) Education in the broadest sense,
(5) Network of Information,
(6) Evaluation.

Methodology is addressed in the sense that the potentials of the NCA Method are investigated. On the other hand, before conducting the teaching of content through alternation between the child's two languages, the teacher must come to grips with a series of related aspects. How is the child in this project assisted in reaching the threshold level in his/her home language? What is the child's attitude to the home and the majority language as he/she strives for the L_1 threshold level? How does the child envision the future role of his/her two languages? The latter question obviously relates to the equal or subordinated status of L_1 as an overall means of communication.

The rapprochement between school and community discussed earlier is obviously a concern of *ethnography* if the investigation of parental language attitudes and patterns is included here. The acknowledgement of the importance of ethnography of communication (Hymes, 1974) as a scholarly field of study and its incorporation into this project points to the fact that the children's attitudes to language in general, and their two languages in particular, can only be developed in school if support in this respect is also found in the homes and the immediate community.

The training of the teaching staff is a necessity in a project as pedagogically demanding as this one. Pedagogy comprises two different phases of the personnel's intellectual growth, inservice training and continued education. In the former, teachers and aides are acquiring the specific expertise to successfully implement the teaching method to which they may have been assigned; in the latter, they are upgrading their professional education at the undergraduate or graduate level, whichever may apply. The advantages of providing training and college education to the participating teaching staff are many, some benefiting the teachers and aides, others the project itself. As the teaching staff become more knowledgeable, they also contribute more effectively to the project.

The impact of the cited project on the schools, the school district and the immediate community is a noteworthy aspect and addresses *education in its broadest sense*. More than merely accepting the project *per se*, one deals here with a much broader issue: the feasibility of bilingual education for minority children, as long as the ingredients of such a program are comparable to those observed in this project. Also, is the implementation of a project of this nature actually leading to the institutionalization of the approach on a wider scale? As for the district under consideration here, this impact is already recognizable as additional bilingual classrooms are being established with local funds, and enrollment figures for the project grades are showing far greater stability than what is usually noted in comparable grades of the two schools.

Finally, for the information on the project to become available to the interested educator, a *network* has been conceived that will allow participants to share the experiences being gained in the Demonstration Project. Within this general framework, the awareness and the demonstrability phases are designed to release information on its progress locally and also nationally. In addition, the project in operation shall be made available for site visitations in the hope that these may ultimately lead to the replication of the NCA model on school campuses where instructional innovation in the field of bilingual education is currently sought.

Conclusion

It is expected that incoming data during the following years of the grant period will provide hard evidence on the greater success of second-language acquisition when the first language is also being developed and a comparable success in conceptual learning when bilingual children are made comfortable with the presence in their lives of two languages and two cultures, as they grow up and learn to adjust to their roles in a mainstream setting.

Andersson, Theodore and Boyer, Mildred. *Bilingual Schooling in the United States*, 2 vols. Austin, TX: Southwest Educational Development Laboratory, 1970.

Cummins, James. "The Entry and Exit Fallacy in Bilingual Education." *NABE Journal*, 4.3 (Spring, 1980).

_____. "Conceptual and Linguistic Foundations of Language Assessment." Paper delivered at the Second Annual Language Assessment Institute (6128). Chicago, IL., 1982.

Gaarder, A. Bruce. "Bilingual Education: Central Questions and Concerns." In Franceso Cordasco, ed., *Bilingual Schooling in the United States*. New York, N.Y.: McGraw-Hill, 1975. 150-158.

Hymes, Dell. *Foundations in Sociolinguistics*. Philadelphia, PA: University of Pennsylvania Press, 1974.

Jacobson, Rodolfo. "The Dilemma of Bilingual Education Modes: Duplication or Compartmentalization," in Maria Burt and Haydee Dulay, ed., *On TESOL'75: New Directions in Second Language Learning, Teaching and Bilingual Education*. Washington, D.C.: TESOL, 1975a.

_____. "Teaching Strategies for the Education of Bilinguals." ERIC Microfiche ED 132836. 1975b.

_____. "The Social Implications of Intrasentential Codeswitching," *New Scholar*, 6(1977):227-256.

_____. "Codeswitching in South Texas: Sociolinguistic Considerations and Pedagogical Applications," *LASSO-Journal*, 3.1 (1978a).

_____. "Consensus and Divergencies in Bilingual Co-ocurrence Patterns: Some Mexican American Versions of 'Tis a Night before Christmas'." Paper delivered at the Seventh Annual UWM Linguistics Symposium (3/18/78) (unpublished). 1978b.

_____. "Anticipatory Embedding and Imaginary Content: Two Newly Identified Codeswitching Variables," in Anthony G. Lozano, ed., *Swallow VII Proceedings of the Seventh Southwest Areal Language and Linguistics Workshop: Bilingual and Biliterate Perspectives*. Boulder, CO: University of Colorado, 1978c.

_____. "Beyond ESL: The Teaching of Content Other Than Language Arts in Bilingual Education," in Richard Bauman and Joel Sherzer, eds., *Working Papers in Sociolinguistics*. Austin, TX: Southwest Educational Development Laboratory, 1979a.

_____. "Can Bilingual Teaching Techniques Reflect Bilingual Community Behavior? — A Study in Ethnoculture and Its Relationship to Some Amendments Contained in the New Bilingual Education Act." In Raymond V. Padilla, ed. *Bilingual Education and Public Policy in the United States*. Ypsilanti, MI: Eastern Michigan University, 1979b.

_____. "Can and Should the Laredo Experiment Be Duplicated Elsewhere? — The Applicability of the Concurrent Approach in Other Communities." In Phillip G. Gonzales, ed., *Proceedings, Eighth Annual International Bilingual-Bicultural Education Conference, Seattle, WA*. Rosslyn, VA: National Clearinghouse for Bilingual Education, 1981a.

_____. "A Title VII Demonstration Project in Bilingual Instructional Methodology." Proposal submitted for funding to the Office of Bilingual Education and Minority Languages, U.S. Department of Education, Washington, D.C. (unpublished). 1981b.

_____. "The Implementation of a Bilingual Instruction Model, the New Concurrent Approach." In Raymond V. Padilla, ed., *Bilingual Education Technology*. Ypsilanti, MI: Eastern Michigan University, 1982a.

_____. "The Role of the Vernacular in Transitional Bilingual Education." In Beverly Hartford, Albert Valdman and Charles R. Foster, eds. *Issues in International Bilingual Education: The Role of the Vernacular*. New York, N.Y.: Plenum Press, 1982b.

_____. "The Title VII Demonstration Project in Bilingual Methodology." In Hernan Lafontaine, Barry Persly, and Leonard Golubchick, eds., *Bilingual Education*, Second edition. Wayne, N.J.: Avery Publishing, Forthcoming.

TEST AND SPONTANEOUS LANGUAGE BEHAVIOR OF LATE PREADOLESCENT SPANISH-ENGLISH BILINGUALS: MORPHOLOGY VS. SYNTAX*

Benji Wald
National Center for Bilingual Research

1. Introduction

The most immediate aim of this paper is to explore and discuss the relationship between the spontaneous and test speech of a sample of 10-12-year-old Hispanic bilinguals. This aim, in turn, serves a larger purpose of interpreting the implications of the findings for the testing of oral speech and the projection of school achievement in literacy.

Certain questions of relevance to educational concerns are involved. In somewhat direct form, the two most prominent question are:

1. What does an oral language test tell us about what a speaker knows and can be used for further language learning?

2. What role does knowledge of oral language play in learning to read and write?

As the title suggests, the answers to both of these questions depend on distinguishing various components of language (components shared by both spoken and written forms of language). The most fundamental components of language which carry meaning and yield to precise analysis are morphology and syntax. To anticipate later discussion, the distinction between morphological and syntactic behavior will reveal the theories of language acquisition, which have emphasized language development up to the age of six, have underestimated the complexity of the purposes which syntax serves and how knowledge of syntax is integrated into increasingly complex demands made on communication in the course of the social development of children into adults, whether or not literacy is part of this development. Syntactic behavior, in particular, will be shown to have relevance to certain widely known hypotheses concerning the relation of bilingualism to literacy achievement among lower SES minority groups, hypotheses generated from information grounded exclusively in test language behavior.

2. Basic concepts

The project reported on here was designed to investigate the influence of situation and topic on the language behavior of 10-12-year-old Spanish-English bilinguals who are from lower-than-middle socio-economic backgrounds in a highly concentrated Mexican-American community in Los Angeles. One of the overall objectives of the study was to gain insight into which aspects of each language were most influenced in quality and/or quantity by changes in situation.

It is well-known to sociolinguistic research that language behavior changes according to situation, but that not all aspects of language change in the same way (cf. Wald:1980, 1981 for reviews of sociolinguistic studies of the relation of language behavior to situation). Accordingly, we reasoned that, to the extent that language proficiency tests are intended to measure behavior representative of the actual language abilities (knowledge, competence, or resources) of speakers in communicative situations, some tests may be more appropriate than others for this purpose, depending on the particular aspects of language which are emphasized in their test designs. For example, among the language proficiency tests widely used for classifying students as limited or fluent, the BSM (Bilingual Syntax Measure) and the BINL (Basic Inventory of Natural Language) are virtually diametrically opposed in their emphases (see Herbert, 1979; Burt & Dulay, 1976).

1. BSM: emphasizes morphology, i.e. word-composition, particularly the use of inflections, e.g. *fall* + Past = *fell*.

2. BINL: morphology is intentionally ignored in favor of sentential syntax, i.e. the organization of words into sentences. The BINL measures mean length of utterance and syntactic complexity. The use of subordination (the adjoining of additional clauses to some element of the main clause of a sentence) is highly valued, e.g. the use of the *relative clause*, as in: there was a silly old monster *who liked to drink pink lemonade*.

In discussing the language behavior of the speakers, it is useful to define and distinguish test and spontaneous speech as follows:

Test Speech: speech elicited by an oral language proficiency assessment instrument following the

*An earlier version of this paper was presented at the Hispanic Research SIG of the AERA in New York City on March 20, 1982. I would like to thank Jim Cummins for comments on the discussion of his ideas. The research reported was funded by NIE through the National Center for Bilingual Research. However, this acknowledgement does not imply any agreement or endorsement of anything contained in this paper by the funding sources.

instructions for administration of the instrument. In the present study, the Spanish and English versions of parts of the BSM-1 and the LAS-1 (Language Assessment Scales) production task were used (see De Avila & Duncan, 1977).

Spontaneous Speech: speech produced in peer situations where speakers were free to initiate and/or reject topics, and to speak in whichever language they preferred (during the first hour).

Each type of speech is situationally distinct, but both types reflect aspects of each speaker's total language resources.

At issue is the correspondence between the two types of speech, and subsequently, the relevance of the correspondence for educational purposes, in particular the achievement of literacy.

Below we consider, in turn, certain aspects of morphology and syntax.

3. Morphological behavior

For purpose of display, the speakers will be divided by age of arrival into three groups. Due to the narrow age-range of the sample, 10-12-year-olds, age of arrival is introconvertible with length of residence and grade of entry into the U.S. school system. The three age-of-arrival groups will be referred to as the *early* (age of arrival 0-5 years, mostly second generation), *middle* (age of arrival 6-8 years), and *late* (age of arrival 9+ years). This grouping is a heuristic found to correspond closely to the language preference of the speakers in spontaneous speech, so that English was preferred by earlier ages of arrivals, and Spanish was progressively preferred by later arrivals (cf. Wald, 1981a).

Figure 1 below displays the percentage of speakers in each group using the indicated morphological form in response to the BSM sections designed to elicit the standard versions of the features. The figure only displays the speakers who responded to the item in either a standard or non-standard form, not those who did not respond at all.

Figure 1

Percentage of Speakers in Each Age of Arrival Group Using the Indicated Morphological Form in Response to the BSM-1 English Proficiency Test

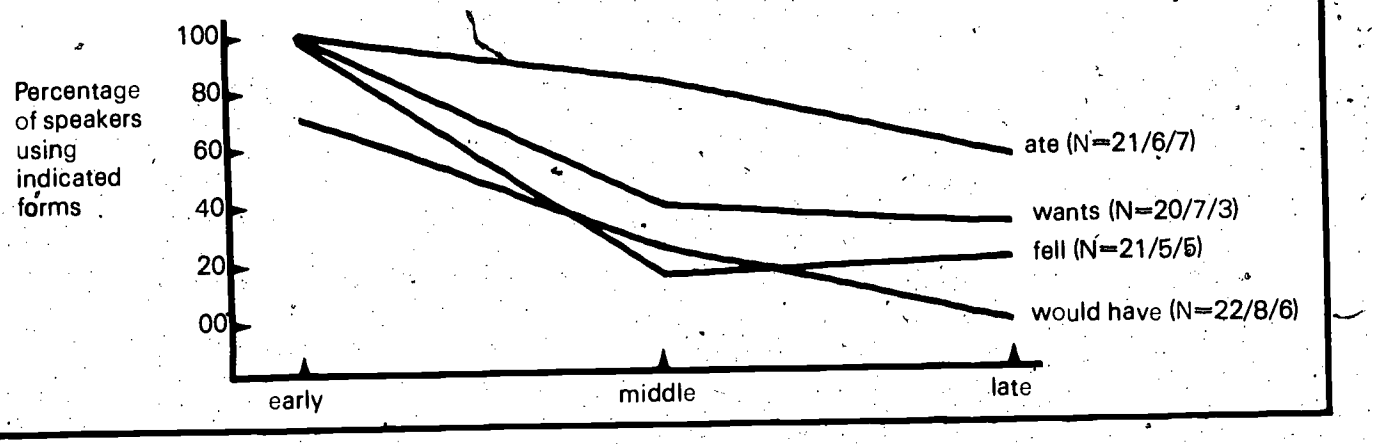


Figure 1 shows a *gradient* pattern, indicating that speakers with earlier ages of arrival (or longer lengths of residence) more frequently use the morphological features tested. Such a pattern is familiar to many studies correlating various aspects of language acquisition with length of residence (although there are studies which show that length of residence is an important indicator of linguistic development only for a while, about 4-5 years, among adults and adolescents, e.g. Heidelberger, 1978; Snow & Hofnagel-Hohle, 1978). The implication is that the *use* of these features reflect the *acquisition* of the features by the speakers.

Similar gradience is found for the speakers in spontaneous speech. For example, Table 1 below gives the unweighted average (average of individual means) of each group for English subject-verb agreement (the suffixation of -s to a present tense verb with a third person singular subject, e.g. *she look-s* vs. *they look).*

On the whole, correspondence between test and spontaneous speech preserves the gradience by age of arrival and length of residence. This indicates steady development of the morphological process with length of residence.

A similar correspondence is found with the use of irregular (strong) past tense forms of verbs in spontaneous speech (represented in BSM test speech by the pasts of the verb *eat* and *fall*). The LAS story retelling presented a greater number of possible contexts for strong pasts than the BSM. Table 2 below shows the closeness of fit between test and spontaneous speech for speakers whose use of the strong past was not fully developed in spontaneous speech.

Table 2 indicates that as the number of possible contexts in test speech increases, the percentage of use converges with the percentage of use in spontaneous speech. At 4-15 tokens (elicited by the LAS story retelling), test behavior in irregular past use corresponds to spontaneous behavior with 85% accuracy.

So far the correspondence between morphological behavior in test and spontaneous speech looks good. The gradient patterns in both situations indicate that the tests are tapping actual developmental patterns. For these speakers and the communities they represent, English morphology and its measurement in test situations is relatively unproblematic.

At this point, we turn to a syntactic feature which shows a different pattern and requires a different explanation.

TABLE 1

Unweighted Average of Subject-Verb Agreement for the Three Age of Arrival Groups in Spontaneous Speech

Age of Arrival	Average	N Speakers	N Possible Contexts
0-5	99	22	648
6-8	70	7	135
9+	55	3	37

(Table includes only speakers with at least 5 possible contexts each.)

TABLE 2

Correspondence between Test and Spontaneous Speech for Speakers Showing Variation in the Marking of Irregular Past Tenses

	Test		
	BSM	LAS	Spontaneous
Range (N of contexts)	2-4	4-15	5-128
Average N contexts/speaker	2.0	8.2	48.6
Average % difference from spontaneous speech	31%	15%	—

4. Syntactic behavior

The most crucial difference between syntax and morphology is in the difficulty of specifying possible contexts for complex syntactic features. Possible contexts in morphology tend to be straight-forward and are specifiable in terms of linguistic contexts alone. An example given above is English subject-verb agreement. On the other hand, complex syntax, which involves more than one clause in a single sentence, is not specifiable in this way. Various possibilities often exist for integrating a number of clauses into spoken discourse. To use the example to be discussed at length below, the information contained in a *relative clause* can also be expressed in a main clause without violating any known linguistic norms of any form of spoken or written English or Spanish. Both of these languages show a great deal of surface structural similarity for this and many other features of syntax. Example (1) below shows relative clauses used in both languages in test speech.

(1) a. It was about a silly old monster *who useta like m pink lemonade*. (JS 12m)
b. Se trata de una giganta *que le qustaban los helados de-de fresas*. (EG 12f)
(Parentheses following the citations identify the speaker by code name, age, and sex, in that order.)

The same content can also be expressed in two independent clauses, as exemplified by the test speech in (2) below.

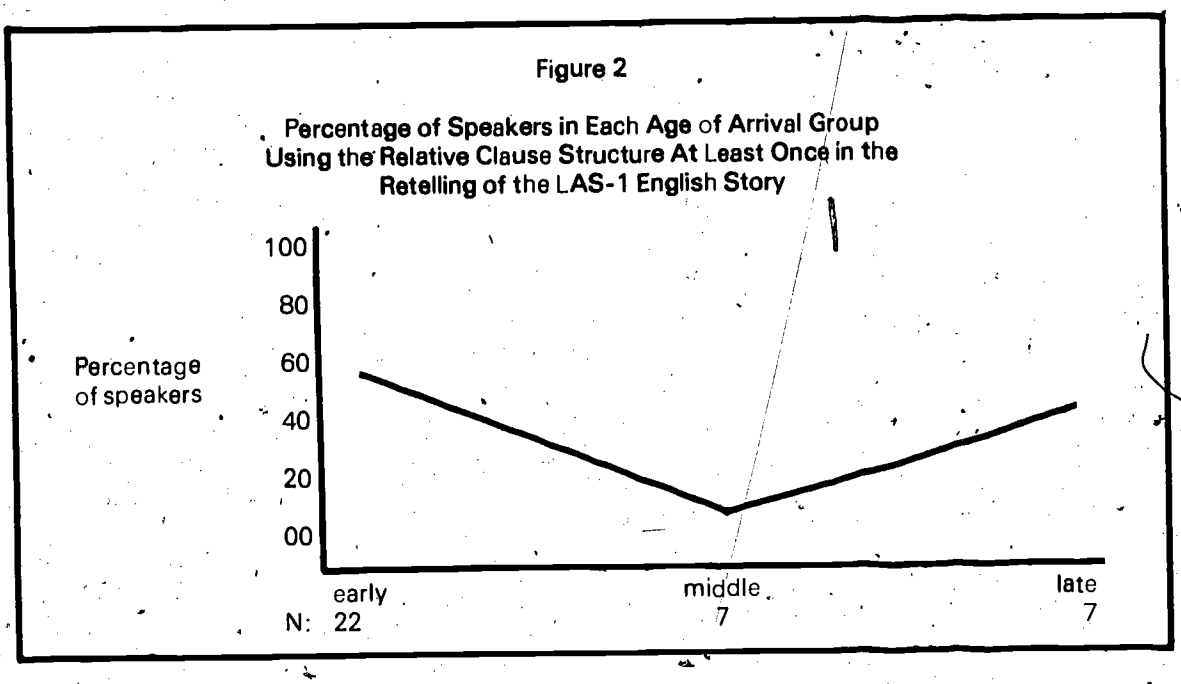
(2) a. Once there was . . . a big monster n he liked . . . lime ice cream. (YL 11f)
b. . . . habia una giganta . . . y -este- le gustaba las fresas. (JS 12m)

In measuring syntactic complexity, testers tend to disregard the problem of possible context in favor of scoring for frequency of use of all syntactic constructions. Complex syntax is highly valued by tests concerned with syntax, so that the examples in (1) contribute more to a syntactic score than the examples in (2) above.

Following this tradition, Figure 2 below displays the percentage of speakers in each group who used a relative clause structure at least once in the English retelling of the LAS-1 English story. The LAS-1 stimulus itself contains three instances of the relative clause structure.

Figure 2

Percentage of Speakers in Each Age of Arrival Group Using the Relative Clause Structure At Least Once in the Retelling of the LAS-1 English Story



The striking feature of the pattern is the dip in the middle group, those speakers first exposed to a predominantly English environment at ages 6-8 years and entering the school system between first and third grade. Since we have divided the sample of continuous ages of arrival into three groups, it is convenient to refer to this pattern as the *pattern of the depressed middle group*.

The pattern is striking because we might have expected to find a gradient pattern suggesting that age of arrival (or length of residence) monotonically influences development and use of relative clauses in English. The assumption that leads to this expectation implicitly equates use in test speech with acquisition and development of the feature. At first glance, this seems reasonable if we transfer our expectations from morphological features of test speech discussed in the preceding section. Although our data on spontaneous speech allows us to investigate this assumption directly, it is useful to try to explain the pattern in terms of a number of recent and current hypotheses which are limited to test language behavior. It will turn out that these hypotheses are inadequate. Nevertheless, they are worth considering because they have been influential in the recent educational research on bilingualism, and because the observations on which they are based must be explained in some way. These hypotheses are discussed immediately below.

5. Cross-Language Hypotheses: Development and Transfer

One of the crudest hypotheses that can be brought to bear on the depressed middle group pattern is Skutnabb-Kangas and Toukomaa's *semilingual hypothesis*. Based on a study of tested Swedish language skills of Finns in several Swedish primary schools, Skutnabb-Kangas and Toukomaa conclude that middle age of arrival Finns, especially those arriving between the ages of 7-8 years, are less likely to develop language skills equivalent to monolingual Swedes than either those Finns who arrived before the age of 6 or those who arrived at age 10+ years, if educated only in Swedish after arrival (Skutnabb-Kangas and Toukomaa, 1976,75). Skutnabb-Kangas and

Toukomaa's hypothesis implies that the arrivals before age 6 have an advantage over the age 7-8 arrivals, presumably because of their headstart in the acquisition of Swedish, but that those Finns who arrive at later ages, 10+, are in an even more advantageous position to learn Swedish, presumably due to the lack of interruption of the development of their Finnish language skills and the ease with which those skills can be transferred to Swedish at older ages.

As quickly pointed out by critics of the hypothesis familiar with the findings of sociolinguistic research, the semilingual hypothesis is extremely crude since it is creates the impression in consumers of the theory that the so-called "semilingual" Finns are less able to express themselves in either language than monolingual Finns or Swedes (cf. Brent-Palmer, 1979; Leap, 1979). Since the Finnish findings are restricted to tests in academic contexts, extension of the findings to spontaneous speech is unwarranted.

In a number of influential papers adding original analyses from Canadian L2 studies, Cummins modified the semilingual hypothesis in a direction responsive to the sociolinguistic criticisms, while attempting to preserve its apparent insights. Particularly in a reanalysis of a study by Wright & Ramsey of a mixed batch of L2 English speakers in fifth, seventh and ninth grade in the Toronto school system, Cummins noted that when length of residence was held constant, older students showed a greater tendency to perform better on many language skills than younger students (see especially Cummins, 1981). For example, Cummins observes that, as a group, students arriving at the adolescent ages of 14-15 learned more vocabulary in one year, as measured by a picture vocabulary test, than students arriving at ages 4-5 learned in 7 years (op cit, 146). This leads to the conclusion that older speakers are learning the same skills at a much more rapid rate.

Cummins further noted that not all language skills favor older over younger speakers in rate of development. For example, language skills such as native-like pronunciation, fluency, and oral comprehension appear to favor younger speakers, according to the studies Cummins reviews, while vocabulary, sound discrimination, morphology, and the aspects of syntax implicit in the test items appear to favor older speakers (cf. Cummins, 1980). Using the criterion of whether rate of development favored younger or older speakers, Cummins suggested that language skills favoring older over younger speakers are different in kind from the other language skills and that they are more relevant to school achievement (Cummins, 1980, 177ff).

By positing differences in relevance of different aspects of language to academic achievement, measured by test performance, Cummins constrains the semilingual issue more clearly than Skutnabb-Kangas & Toukomaa. However, unclarities still remain about precisely what skills are most relevant to school achievement, and, even more importantly, why this is so. This point will recur as discussion proceeds.

Cummins reformulates Skutnabb-Kangas & Toukomaa's hypothesis with much greater precision in the *interdependence hypothesis*. According to this hypothesis, those skills which are most relevant to academic achievement can be transferred from L1 to L2 at a rapid rate as the speaker approaches adolescence. On the other hand, Cummins suggests that, for speakers from minority populations historically discriminated against and concentrated in subordinate statuses in the society represented by the educational system (conditions fulfilled by both the majorities of the Finnish community in Sweden and the Mexican-American community among others), if students do not acquire the relevant skills in L1 first, they will take much longer to develop in L2, and consequently the student will fall further behind grade level (cf. especially Cummins, forthcoming, 12ff).

The interdependence hypothesis has two basic features: *development* and *transfer*. If this hypothesis is applied to the pattern of Figure 2 above, the simplest interpretation would be as follows:

Both extreme groups (i.e. the early and late groups) show similar relative clause *use* because they have similar relative clause *development*. The *early* group shows relatively advanced English relative clause development because early acquisition of English has allowed sufficient time for the structure to develop. The *late* group shows English relative clause development comparable to the early group because its speakers have had sufficient time to develop the structure in *Spanish* and have been able to *transfer* that skill relatively quickly as their English develops.

Following this logic, the depressed position of the middle group reflects a disruption of Spanish syntactic development without compensation in English development, due to a relatively late age of arrival and short length of residence compared to the early group.

6. Group Cross-Language Comparison

If the interdependence hypothesis is taken literally, we would expect to find a patterning of relative clause use across groups in Spanish similar to the English one.

Figure 3 below superimposes the pattern of relative clause use in Spanish on the English pattern already seen in Figure 1 above, for the same speakers. The Spanish data comes from the Spanish version of the LAS story retelling task. The Spanish LAS stimulus story is identical to the English LAS story in all relevant respects, up to the use of the same three stimulus relative clauses in their Spanish guises.

Unlike the English pattern of the depressed middle group, the Spanish pattern shows the gradience usually associated with monotonic development.

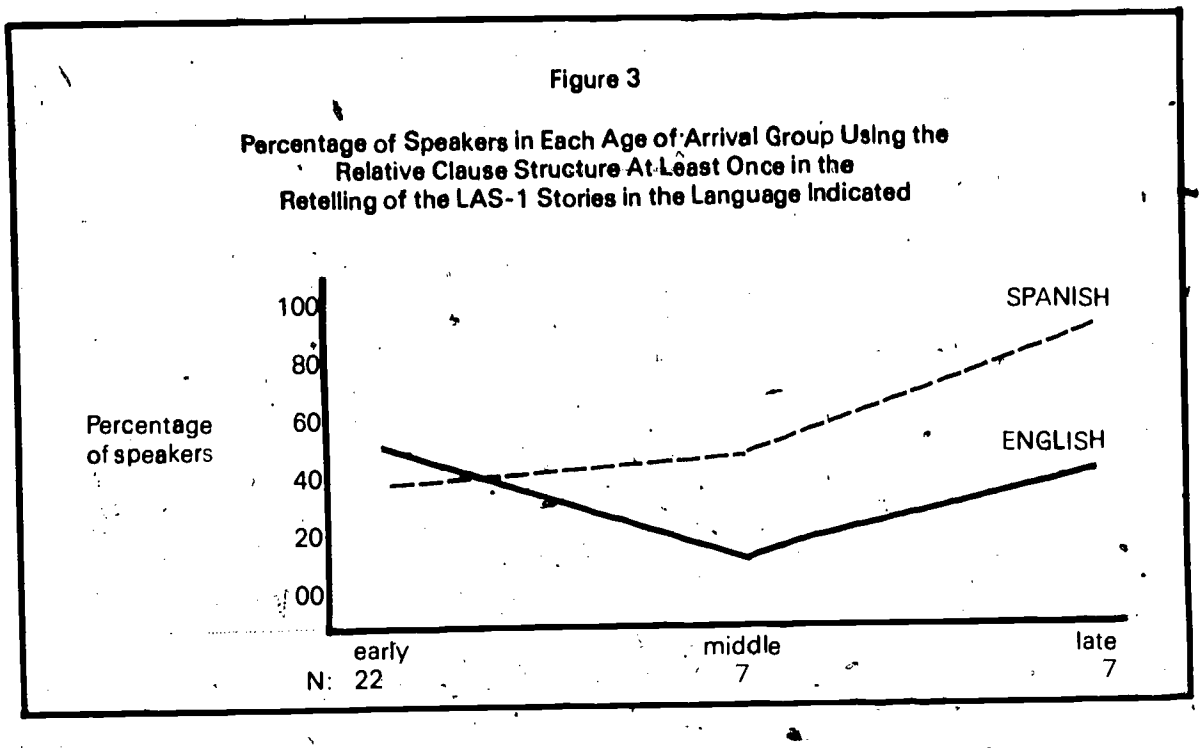
While the comparison of the Spanish and English patterns does not confirm the interdependence hypothesis, it does not disconfirm it either. For example, one might argue as follows:

The early group shows equivalent development in Spanish and English, but the later groups show more development in Spanish. In this case, the problem of transfer is restricted to those who have developed the relative clause in Spanish, but not in English. One might propose that those who have developed the relative clause in Spanish, but not in English, have not developed their English well enough to effect transfer.

This still makes the middle group look the most underdeveloped, since they show relatively little difference from the early group in apparent Spanish development, along with a relatively depressed development in English. Considering the longer length of residence of the middle group, along with the apparent depressed development of their English skill compared to the late group, the middle group appears not only to be relatively lacking in development, but also in ability to transfer across languages. Finally, the apparent lack of development of many members of the early group in either language makes them look inferior to the late group, and only superior to the middle group in English.

Figure 3

Percentage of Speakers in Each Age of Arrival Group Using the Relative Clause Structure At Least Once in the Retelling of the LAS-1 Stories in the Language Indicated



7. Individual Cross-Language Comparison

The finer resolution of Table 3 below alters the perspective of Figure 3 to some extent.

Here we see the actual match of use of the relative clause in English and Spanish for the individuals in each group.

The interdependence of relative clause development in Spanish and English is now seen to be very unclear for the early group. Very few speakers show use of the structure in both languages. Most use it in only one of the two languages without any appreciable difference of likelihood in either language. On the other hand, the middle and late groups preserve the impression that the relative clause must develop in Spanish before it develops in English. This allows us to continue to maintain that for non-early English speakers, the structure develops first in Spanish and is then transferred to English, rather than independently developing in English regardless of Spanish development.

The depressed middle group pattern is reflected in the proportion of members of each group who used the structure in *neither* language. The middle group shows the greatest proportion of individuals failing to exhibit development in either language. This is followed by the early group, which shows less use than the late group in at least one language.

8. Syntax: Test Speech Compared to Spontaneous Speech

Table 4 below now adds the situational variable to relative clause use of the students. It compares use of the structure in test speech, discussed above, with its use in spontaneous speech.

The results reveal serious discrepancies between observed performance in spontaneous speech and in test speech. Only two speakers show no evidence of the relative clause in any observed context. Both of these speakers were members of the early group. They talked relatively little in either language. For them, as for all the others, we cannot assert that failure to *use* the relative

clause in either language implies failure to have *developed* the structure. All other speakers had developed the structure in at least one language. Thus, the patterns of use reflected in the test speech are not reliable indicators of spontaneous speech for this syntactic structure. Most importantly, the patterns of test speech are not accurate indicators of the linguistic *knowledge* and *resources* underlying the use of this syntactic structure in either spontaneous or test speech.

These results immediately belie any attempted semilingual hypothesis in explanation of the test speech patterns. For the interdependence hypothesis, the question becomes not only the problem of *transfer across languages*, but of *transfer across situations within a single language*. Thus, although *all* the middle group speakers used the relative clause in English spontaneous speech, only one of them transferred this use to the test situation.

TABLE 3

Comparison of RC Use in Spanish and English Story-Retelling by ADA

ADA	Spanish Only	Both	English Only	Neither	N
0-5	.27	.22	.22	.27	(22)
6-8	.43	.14	.00	.43	(7)
9+	.57	.43	.00	.00	(7)

TABLE 4

Comparison of Use of RC in LAS-E and -S Story-Retelling and Use of RC in English Spontaneous Speech

	RC use in LAS			
	Sp. Only	Both	Eng. Only	Neither
Percent of Speakers using English RC in D1 (N)	.62 (13)	.55 (9)	1.00 (5)	.78 (9)

9. Discussion

This final section discusses the implications of the behavior discussed above for relevance to educational concerns.

A. The use of test speech without control or possible contexts for complex syntax leads to an unreliable measure of the actual language resources of the speaker tested.

First, the possibility of using complex syntax may depend on how much the speaker chooses to say. Setting aside actual language resources, one cannot safely assume that all students approach the same task of story retelling or picture description with the same interest or enthusiasm.

We may expect syntax to be richest when the speaker has greater access than the tester to the information to be expressed in speech (cf. Guthrie & Steffersen, 1980). To the extent that syntactic organization reflects the speaker's evaluation of how information should be presented (e.g. by a subordinate or independent clause), the speaker's additional task of guessing which information is most important in the test situation (communicating to a tester who may be assumed to already know the information) complicates the process of formulating a test response. The speaker must not only evaluate *which* information to present, but *how* to frame that information in some syntactic form.

The lack of control of possible context for relative clause vs. independent (or conjoined) sentence, as in (1) and (2) above, can be contrasted with the control of possible context within narrower, more rigorously defined syntactic contexts. For example, in the literacy standards of both English and Spanish, relativization on the object of a preposition allows the preposition to precede the relative marker (literary Spanish requires it while literary English is more flexible in allowing alternation between the two syntactic types), as illustrated in (3) below:

(3) a. [redacted] car *about which* I told you
b. el hombre *de quien/que* (yo) estoy hablando

However, in spontaneous speech from all speakers for whom prepositional relatives were observed, the constructions represented in (3) were never observed. The equivalents are presented in (4) below.

(4) A the car *(that)* I just told you *about* (AL 12f)
b el *que* yo estoy hablando (OM 10m)

In the speakers' spontaneous English, the preposition was always left stranded. In the equivalent Spanish, it was always unexpressed.

Because relativization on a prepositional objective is strictly defined in terms of linguistic elements, the possible context for the literary construction is clear. Speakers' use of the types in (4), but not in (3), indicates that the spoken versions of English and Spanish used by the speakers differ from the standard literary versions of both languages for this feature.

In contrast, the possible contexts which include the relative clauses of (1) and the adjacent main clause pairs of (2) cannot be strictly defined in linguistic terms. No matter how many adjacent main clause pairs we find, we cannot conclude that the speaker has not acquired the relative clause structure.

It is possible to create favorable contexts for eliciting the relative clause structure, e.g. by asking "what's a teacher?" (expected answer: somebody *who teaches*, cf. Chiang, 1980). However, the holistic measurement of syntactic development would require a large number of specially designed discrete-point items. Until proficiency testing has decided precisely what points of syntax it wants to measure for what purposes, it is unlikely that such a test will be widely used in evaluating students' language proficiency for academic purposes. But until such decisions are made, language proficiency testing for syntax continues to run the risk of seriously underestimating the student's linguistic resources and supporting programs which retard the student's academic progress in either or both languages.

B. The transfer aspect of the bilingual interdependence hypothesis is complex. It applies not only to transfer across languages, but also to transfer within a single language across situations.

From the perspective presented here, the primary value of the interdependence hypothesis is that it recognizes that different aspects of a language develop independently and are differentially related to literacy achievement. In assigning special attention to features of language which show a more rapid rate of development in L2 for older speakers, it has potential for contributing to the development of rational, well-motivated programs for teaching oral- and literacy-related skills to limited-English speakers.

One case where it may be justified in its assignment of differential importance to aspects of language for the purpose of literacy is in its discounting of pronunciation as an important literacy related skill. The common observation that native-like pronunciation of L2 is developed more rapidly by younger speakers implies that it is *not* among the most important literacy-related skills, according to the hypothesis. This has direct implications for the procedures used to teach reading. Convergent evidence comes from a microethnographic study reported by Moll (1981). Moll discusses a group of sophisticated third grade readers able to draw inferences from their Spanish-reading material. Drawing inferences is a high-level reading skill. At the same time, in a totally independent English reading class, the same students are only required to practice pronunciation and to perform low-level decoding tasks. The implication is that these students, who are capable of much more demanding tasks in English reading, are held back by a methodology which places unduly heavy emphasis on phonics and narrow decoding skills, as if reading skills in the two languages were independent of each other. The issue of unduly emphasizing pronunciation and even of misinterpreting pronunciation differences between standard and nonstandard varieties of a language as decoding errors has been identified by Labov (1972) for teachers of students speaking black English vernacular. Thus, pronunication represents a case of convergence between the predictions of interdependence hypothesis and the observations of scholars concerned with misinterpretations of language behavior in instructional situations.

However, the findings reported in this paper indicate that conclusions about the role of complex syntax in literacy achievement may be of a quite different order. The present case of the relative clause suggests that many tests do not simply test the language resources of students, but also their test-taking abilities. The relative clause, a feature of high value in the test-taking situation, is commonly used in spontaneous speech by virtually all speakers, but reflects a pattern associated with literacy-related language skills in test speech. It is, therefore, not a straightforward task to separate aspects of L2 and claim that they have a faster rate of development for older speakers without adding *in a test situation*.

The evidence suggests that syntactic behavior is very much influenced by features related to the school situation, both in testing and achieving literacy. It follows that difficulties in the acquisition of literacy and in language proficiency testing proceed not only from lack of language development, but also, and perhaps more

commonly, from difficulties in transferring already developed language resources to academic situations. Thus, the route to improved literacy achievement must include identification and academic development of resources which the speaker already possesses. How to do this will be labelled the transfer problem.

C. Production and comprehension of complex syntax is a literacy-related skill. The relation between syntactic complexity in vernacular and literate standard languages needs to be recognized in order to solve the transfer problem.

Complex syntax contributes to the organization of information in spoken discourse and written texts. The written language has a highly developed hierarchical information structure which, in its full form in the school textbook, includes chapters, sections, subsections, paragraphs and sentences. The sentences include those with more than one clause, either conjoined or subordinate. Recognition of this organization and ability to use it for information retrieval is a high-level goal of reading instruction. It is used, for example, for finding passages in a text which support inferences made beyond the actual text, as well as for locating information actually expressed in the text. Complementarily, a high-level goal of writing instruction is to provide students with the knowledge necessary to reproduce this information structure (e.g. in the extended book report).

The relation of the organization of spoken to written language is the subject of some current research. Evidence is beginning to accumulate that there are differences between some of the devices used to organize information in the spoken language of some lower SES communities and the standard written language, and that primary school students transfer these spoken devices to the written language in violation of the standard literary norms (cf. Michaels, 1981).

At the same time, the unqualified assertion that written language is less contextually dependent than spoken language has been challenged for some genres of speech and writing by some scholars (e.g. Prince, 1981 contra Olson, 1977; Fillmore and Kay's work on inferencing strategies used in real-time reading and in choosing among multiple choice answers is also relevant, cf. Fillmore & Kay, 1980). Certainly the stereotype that characterizes spoken languages as largely dependent on what is present in the speech situation has been exaggerated where school-age children are concerned. In our own research, we found that virtually all students in the 10-12 age range are highly capable of formulating complex narratives of personal experience about people and events not present or evident in any way in the speech situation.

Within the domain of the sentence, there has long been evidence that cognitively and structurally complex forms are used by speakers of all SES levels, whether or not they conform to standard written norms (cf. Labov, 1972).

In the case of the relative clause, as an example of complex syntax, there is some evidence accruing that it is sensitive to medium so that it tends to be used more often in written than in spoken presentation of the same informational content (A. Kroch, p.c. on a University of Pennsylvania project comparing spoken and written versions of the same text produced by undergraduate, largely working and lower middle class, college students). This connection between structures like the relative clause and literacy indicates why it is highly valued by some test designs.

Our findings show that there is a non-literate spoken base for the relative clause (among other types of complex syntax). For the most part, the educational problem is not one of teaching the form of the relative clause or even of transferring its form from one language to another (for L1 Spanish speakers), but rather of teaching the difference between the *uses* of the relative clause (and other syntactic features) in spoken and written language. Toward this goal, it is essential to know the difference between the vernacular uses in the community served and the uses in the standard language.

As a final note, it is also important to recognize whatever formal differences occur between written and spoken language. As discussed above, the vernacular forms of Spanish and English of the community of speakers studied differ from the standard forms of both languages in relativizing on a prepositional object (as in (3) and (4) above). The standard forms of Spanish and English are congruent across these languages as a result of a common origin in the same Latinate literary tradition. The vernacular forms have evolved through an independent spoken tradition. In the case of English, the vernacular form is much older than the standard form, and reflects a durable tradition acquired by L1 Spanish speakers in L2 *without* transference from Spanish (cf. Jespersen, 1965 on the history of the different forms of the relative clause in English). This indicates acquisition of these features of L2 is effectively taking place outside of the academic context. The impact of vernacular syntactic forms of English for the bilinguals discussed here stands in striking contrast to the lack of impact of syntactic forms restricted to the standard literary language. The standard English form of relativization on a preposition is already found in second grade textbooks, although we have seen that it is not used in speech even at the sixth grade level (cf. Language Skills Framework, 1980:382).

Possible difficulty in interpreting the standard construction in reading on the part of such students

is not necessarily the result of bilingualism (although the first impulse of the educator insulated from the realities of vernacular speech is often to draw that conclusion), but rather of the acquisition of a nonstandard form of English shared by most lower SES groups.

The consequences of our findings are:

1. Hypotheses about syntactic development based on test speech exclusively, while possibly symptomatic of literacy-related problems, do not rest on a reliable data base as indicators of language resources of bilingual students of lower SES backgrounds.

2. Transference of language resources across both languages and situations are literacy-related problems which must be distinguished from developmental problems in oral second language development.

3. Differences between standard and non-standard syntax known to exist in many monolingual English (and Spanish) also exist in (at least) some lower SES bilingual communities and must be carefully distinguished from differences caused by direct bilingual transference.

In separating the various sources of problems involving language resources toward the goal of literacy, it is necessary to develop a firm informational base distinguishing both forms and uses of syntactic structures in vernacular and standard languages — no less for bilinguals than for anyone else.

References

Brent-Palmer, C. "A Sociolinguistic Assessment of the Notion 'Immigrant Semilingualism' from a Social Conflict Perspective." *Working Papers on Bilingualism*. 17:135-180. April, 1979.

Burt, M. & Dulay, H. *Technical Manual for the Bilingual Syntax Measure*. New York: Harcourt, Brace and Jovanovich, Inc., 1976.

Chiang, D.L. "Predictors of Relative Clause Production." In S. Krashen & R. Scarcell eds., *Research in Second Language Acquisition*. Rowley, Mass.: Newburn House. 142-145. 1980.

Cummins, J. "The Cross-Lingual Dimensions of Language Proficiency: Implications for Bilingual Education and the Optimal Age Issue." *TESOL Quarterly*, 14:2, 175-187. 1980.

Cummins, J. "Age on Arrival and Immigrant Second Language in Canada: A Reassessment," *Applied Linguistics*, 11:2, 132-149. 1981.

Cummins, J. "The Effectiveness of Bilingual Education." To appear in *Focus*, forthcoming.

De Avila, E.A. & Duncan, S.E. "Language Assessment Scales." *LAS 1: Examiner's Manual*. Corte Madera, Ca: Linguametrics Group, Inc. 1977.

Fillmore, C. & Kay, P. "Progress Report: Text Semantic Analysis of Reading Comprehension Tests." UC Berkeley Depts. of Linguistics and Anthropology. Unpublished ms. 1980.

Guthrie, L. & Steffersen, M.S. "Effect of Situation on the Verbalization of Black Inner-City Children." Paper presented at the LSA Summer Meeting. Albuquerque, N.M. Aug. 3, 1980.

Heidelberger Sprachsforschung Projekt. "The Acquisition of German Syntax by Foreign Workers." In D. Sankoff, ed. *Linguistic Variation: Models and Methods*. New York: Academic Press, 1978. 1-22.

Herbert, C.H. *Basic Inventory of Natural Language (BINL). Instructions Manual, Revised*. San Bernardino, CA: CHECpoint Systems, Inc., 1979.

Jerspersen, O. *A Modern English Grammar*. London: George Allen & Unwin, 1965.

Labov, W. "The Logic of Nonstandard English." In *Language in the Inner City: Studies in the Black English Vernacular*. Philadelphia: University of Pennsylvania Press, 1972.

Language Skills Framework. *Vol 1: Oral Language: Resources for Developing a Student Placement System for Bilingual Programs*. SWRL project for the U.S. Office of Education. Los Alamitos. March, 1980.

Leap, W. "On the Implicit Politics of Educational Research: Indian English and the 'Semilingualism Conspiracy'." Unpublished ms., 1979.

Michaels, S. "'Sharing Time': Children's Narrative Styles and Differential Access to Literacy." *Languages in Society*, 10:3,423-442. 1981.

Michaels, S. & Kadar, L. "Children's Discourse Style and Differential Assess to Literacy." Paper presented at the annual meeting of the American Association of Anthropolgists. December, 1981.

Moll, L.C. "The Microethnographic Study of Bilingual Schooling." In R.V. Padilla, ed., *Ethnoperspectives in Bilingual Education Research, Vol. III: Bilingual Education Technology.* Ypsilanti: Eastern Michigan University, 1981. 430-446.

Olson, D. "From Utterance to Text: The Bias of Language in Speech and Writing." *Harvard Educational Review*, 47, 257-281. 1977.

Padilla, R.V., ed. *Ethnoperspectives in Bilingual Education Research, Vol. III: Bilingual Education Technology.* Ypsilanti: Eastern Michigan University, 1981.

Prince, E. "Toward a Taxonomy of Given-New Information." In P. Cole, ed., *Radical Pragmatics.* New York: Academic Press. 1981, Pp. 223-255.

Skutnabb-Kangas, T. & Toukomaa, P. *Teaching Migrant Children's Mother Tongue and Learning the Language of the Host Country in the Context of the Sociocultural Situation of the Migrant Family.* Dept. of Sociology and Social Psychology. University of Tempere, Finland, 1976.

Snow, C.E. & Hoefnagel-Hohle, M. "The Critical Period for Language Acquisition: Evidence from Second Language Learning." *Child Development, 49,* 1114-1128. 1978.

Wald, B. "Report on the Study of Limited Language Proficiency." *Working Paper Series.* National Center for Bilingual Research. 1980.

Wald, B. "Reaction to Troike's 'SCALP'." Presented at the Language Proficiency Assessment Symposium. Warrenton, Va. March, 1981.

Wald, B. "The Relation of Topic/Situation Sensitivity to the Study of Language Proficiency." In R.V. Padilla, ed., *Ethnoperspectives in Bilingual Education Research, Vol. III: Bilingual Education Technology. Ypsilanti: Eastern Michigan University, 281-306. 1981a.*

GRADE LEVEL AND ENGLISH ACQUISITION: AGE, YEAR OF BIRTH, OR SOMETHING ELSE?

Gerald E. DeMauro
New Jersey State Department of Education

Abstract

Recent analyses of achievement in New Jersey's categorical programs have shown that the amount that students gain, relative to test norms, is related to grade level. Younger students make the largest gains. The characteristic grade function seems to describe gains made by New Jersey bilingual education students on the *Language Assessment Battery* (Cumbo, O'Neill, Tilis and Weichun, 1976). However, on closer examination, differences in scoring gains reflect changes in test levels, rather than changes in grades. The phenomenon is related to changes in aptitude noted on recent intelligence tests. Younger children have been scoring higher on these tests in recent years (Herman, 1979). Perhaps cultural influences are causing temporary verbal aptitude changes. Decreases in these changes over time (Lazar and Darlington, 1982) may coincide with declines in the rate of language acquisition over grades. Other theories including the influence of acculturation, etc., are presented as possible alternate explanations of the grade effect.

Gains made in reading (Figure 1) and math (Figure 2) scores by New Jersey's state compensatory education and Title I students in recent years have been negatively related to grade level (DeMauro, 1980, 1981a). Since analyses are conducted in Normal Curve Equivalent (NCE) scores, which gauge student performance relative to test norms (Tallmadge, 1976), these results show that younger program students make larger improvements relative to grade peers.

Echternacht (1980) reports that early attempts were made to standardize Title I program effects by transforming mean student scale scores from 23 states on fall, 1977 pretests and spring, 1978 posttests to NCEs. The results, termed "normal growth," agree with New Jersey's findings of greatest gains in the lowest grades.[1]

Similar results are common in bilingual education. Douglas and Johnson (1981) claim that reading and math gains of first and second grade Title VII students surpass what could be expected from the test norms. The effect has some generality over academic years and treatment populations.

In a 1979-80 study of English acquisition by New Jersey bilingual students (DeMauro, 1981b), the relationship of gains in NCEs to grade level was similar except in grades 1 and 7. The scores in Figures 1 and 2 represent means of average grade level gains reported to the New Jersey Department of Education by local districts. The scores in Figure 3 are student mean gains.

The data in Figure 3 are derived from the fall, 1979 and spring, 1980 administrations of the *Language Assessment Battery*[2] (Cumbo, O'Neill, Tilis and Weichun, 1976) to 729 bilingual education students. The sample, from grades 1-12, was stratified to represent New Jersey's bilingual education population in educational environment (pullout, center approach, etc.), socioeconomic status, region of the state, and years of program experience.

The English skills acquisition data is the focus of the discussion because:

a) The single instrument requires no assumptions concerning the equivalence of NCE scores from various tests;

b) The sample represents New Jersey's bilingual education population;

c) The sample stratification permits analyses of meaningful effects, such as program experience.

The current paper examines grade level (and age) as a determinant of English skills acquisition by bilingual education students. It discusses the general grade effect evidence when standardized scores (such as NCEs) are used as well as other grade effects specific to the acquisition of English as a second language. Finally, several cognitive and noncognitive explanations for the observed effects are discussed. First, the effect is defined as accurately as possible through covariance.

Regression Scores

To reduce the large error component of difference scores (Thorndike and Hagen, 1969), the pretest was covaried from the posttest of the New Jersey English acquisition data and the adjusted means were plotted by grade and years of program experience (Figure 4).

[1] The author recommends caution as program selection was based on pretest scores, and some gain may be attributed to regression to the mean.

[2] This instrument has four subtests, listening, speaking, reading and writing. The scores used in analyses were NCEs derived from New York City's total test norms.

Figure 1

Mean Math Gains (in NCEs) of
New Jersey Title I and State Compensatory
Education Students, 1978-1980, By Grade

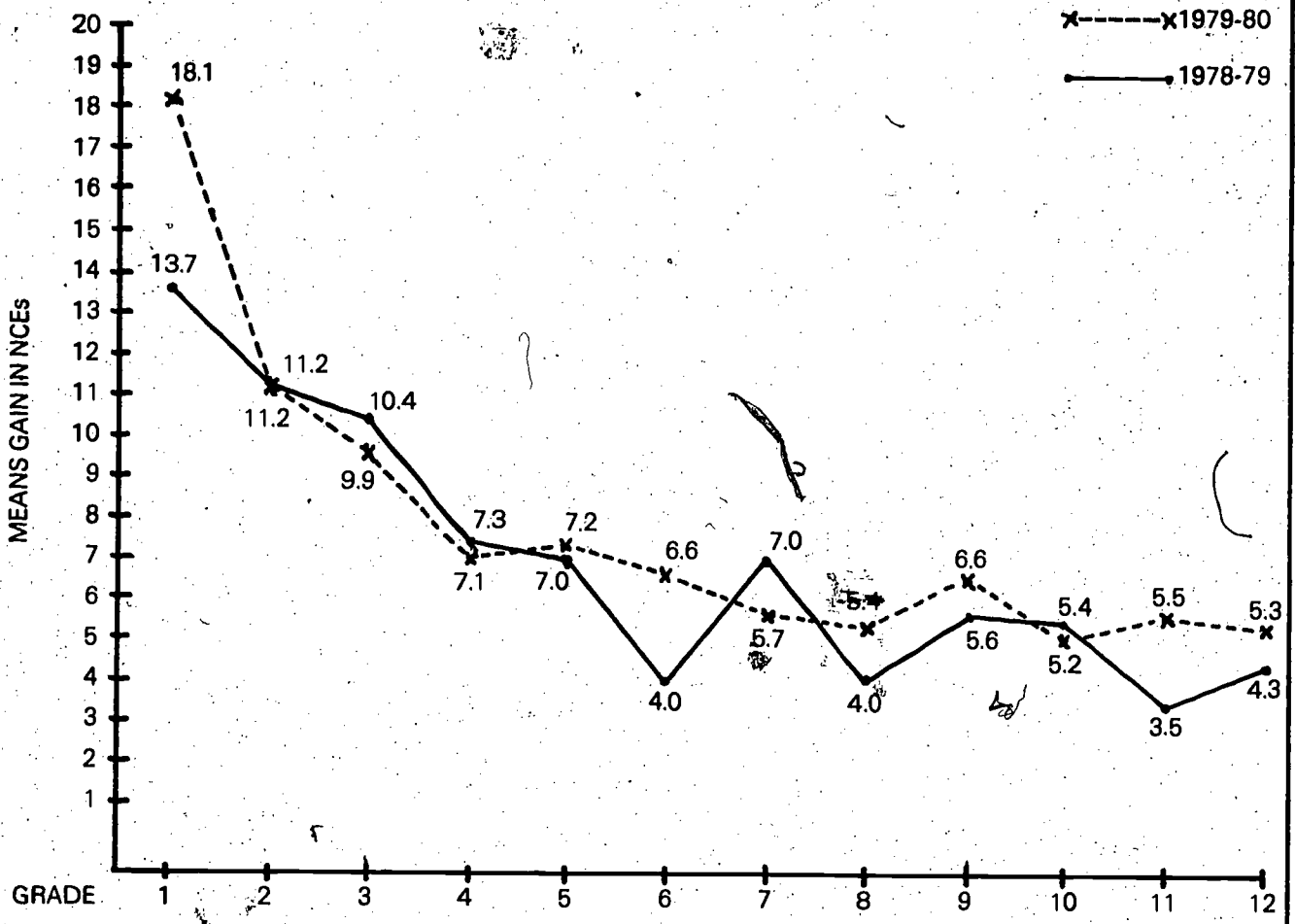


Figure 2

Mean Reading Gains (in NCEs) of
New Jersey Title I and State Compensatory
Education Students, 1978-1980, By Grade

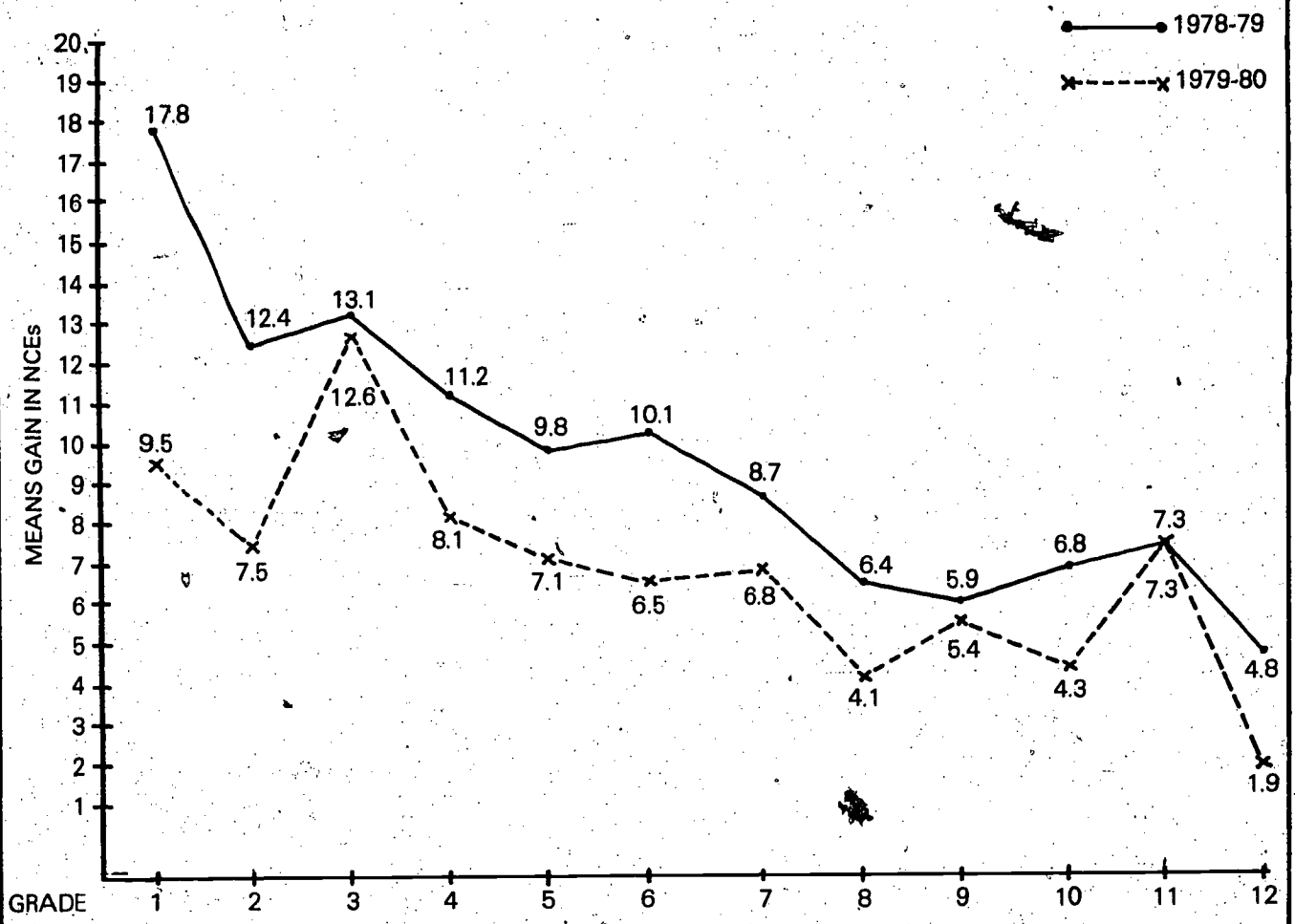


Figure 3

Mean Gains (in NCEs) of
a Representative Sample of
New Jersey Bilingual Education Students
on the LAB, 1979-1980

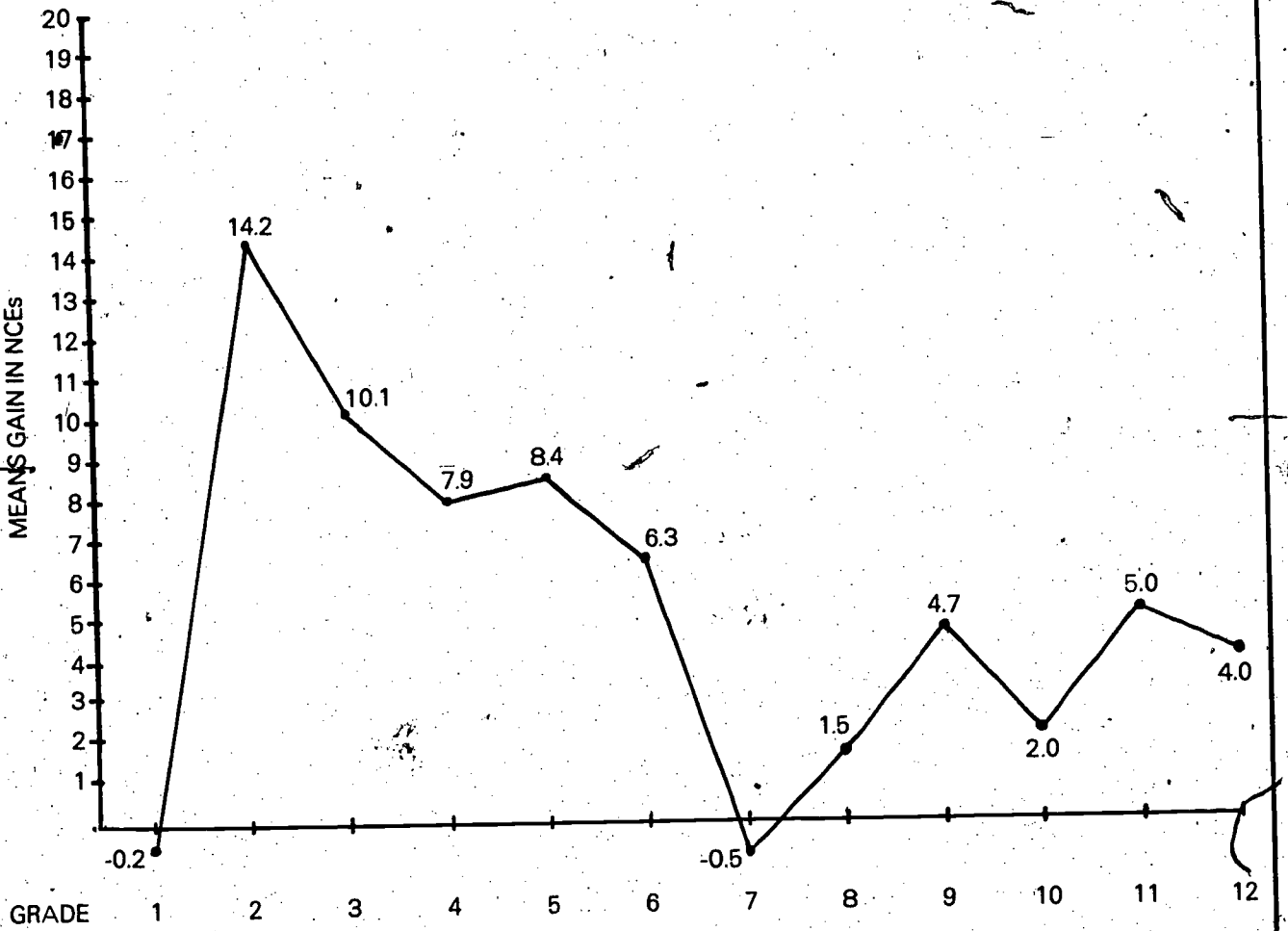
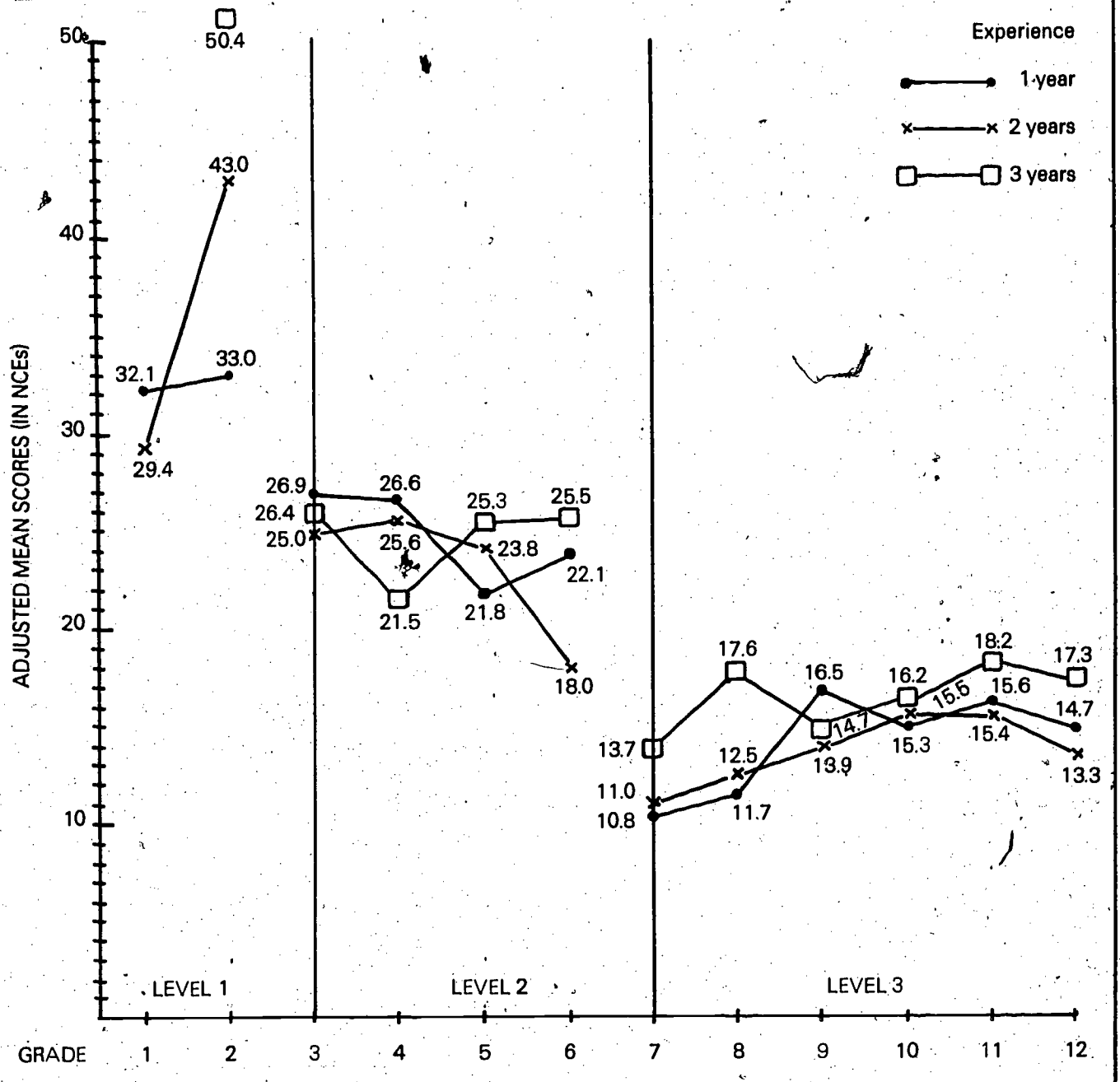


Figure 4

Adjusted Posttest Means (in NCEs) of a Representative Sample of New Jersey Bilingual Education Students on the LAB, 1979-1980



The *LAB* is divided into three levels, spanning grades K-2, 3-6, and 7-12. The covariance analysis is conducted within each level. Main effects are program experience and grade, the dependent variable is the posttest and the covariate is the pretest.

Significant effects are only found in level 1. The grade effect (F (1,257)=5.09, $p < .025$) shows that second graders outscored first graders. Students with more program experience outscored those with less (no first graders had three years of experience) (F (3,257)=3.56, $p < .015$). The grade and program experience interaction (F (3,257)=3.93, $p < .048$) shows that more experienced students scored higher in second grade, but lower in first.

As shown in Figure 4, the characteristic grade effect is better described in these data as differences among test levels and not grades within levels. One major assumption underlying the use of English norms is that tests that are more difficult for monolingual English students are proportionally more difficult for limited-English-proficient (LEP) students. Therefore, a harder level of a test should yield the same distribution of NCE scores because the test is harder for both the English norming sample and the LEP target population. In fact, a more difficult test may depress LEP student performance much more than it depresses the performance of the monolingual English norms group. This would explain the relatively poorer performance of LEP students on different levels of a test that is normed on monolingual English samples of equivalent abilities. Although the *LAB* standardization procedures attempted to control for student abilities across levels (see the technical manual), these levels may have had varying difficulty for LEP students.

Alternately, the differences in scoring across levels may reflect real differences in language acquisition that coincide with test levels. Younger children may learn English more quickly. This argument is consistent with the fact that grade has been a significant effect in all of the achievement analyses in New Jersey and is not an artifact of a particular test. Moreover, analyses of English language achievement on various tests administered by New Jersey local bilingual education programs show that grade is significant both as a main effect (F (3,70)=10.241, $p < .001$) and in interaction with pretest to posttest gains (F (3,71)=7.028, $p < .001$) (Table 1).

Critical Period Hypothesis

Perhaps Figure 3 may be explained by a critical period phenomenon. Younger children appear to approach English skill levels of monolingual grade peers more rapidly.

When English gains are examined without reference to grade norms, the superiority of younger children vanishes. In their examination of the critical period hypothesis, Snow and Hoefnagel-Hohle (1978) report that children acquire Dutch morphology and syntax and listening comprehension more rapidly. Contrary to the hypothesis, 12-15 year olds learned Dutch most rapidly and 3-5 year olds, least rapidly.

After reviewing language acquisition literature, Krashen, Long and Scarcella (1979) conclude that adults and other children acquire syntax and morphology faster, at least in the first year, than younger children. However, people who are naturally exposed to second languages as children become more proficient than those who are first exposed as adults.

Table I

Achievement Test Gains in NCEs of Bilingual Programs in New Jersey During the 1980-81 Academic Year in Reports to the State

Grade Cluster*	Number	Pretest Mean	Posttest Mean	Mean Gain
1-3	21	27.9	52.2	24.3
4-6	16	28.0	42.0	14.0
7-9	16	24.4	34.2	10.2
10-12	18	10.4	15.8	5.4

*Because the unit of analysis is program means, the number of available means is small, and grades are clustered to increase stability.

These studies do not conflict with the achievement data findings. The achievement data do not show that younger children outperform older children. Rather, they show that today's younger children acquire English more rapidly than the younger children in the test norms. Similar results have been noted when comparing performances in IQ tests between today's children and age peers from years ago.

Herman (1979) notes that the same raw score yields a lower IQ on the *WISC* (Wechsler, 1949) than on the WISC-R (Wechsler, 1974). He theorizes that these differences may either reflect differences in standardization samples or true increases in abilities that result from improvements in underlying capacities, test wiseness or cultural exposure to toys, books or television.

Herman's data on the *Binet* (Terman and Merrill, 1972) show that interedition IQ differences are more pronounced for younger children. This implies a curvilinear relationship between measured abilities and year of birth.

As an illustration, suppose that the improvements in measured abilities were predicted by the year of the child's birth. Six-year-olds (first graders) who took the test in 1979 would average three score units higher than six-year-olds who took the test in 1976 under a simple linear model (growth=birth year + constant). The same three units would separate children of any age.*

*Actually, the same three units times the regression slope.

Under the simplest quadratic model, e.g. growth=(b)(birth years)^2+ c, the differences for six-year-olds would follow: $b(y+3)^2+c-(b(y)^2+c)=6by+9b$, where b is the slope of year on growth, y is a linear function of the birth year and c is a constant. For 11-year-olds, this difference would be smaller, because they would have been born five years earlier, and y would be five units less.

IQ differences between the *WISC* and *WISC-R* are greater on the performance scale than on the verbal scale and interact with age (Doppelt and Kaufman, 1977). Below age 11, IQs average about four points lower on the verbal scale and about seven points lower on the performance scale. Above age 11, the performance scale still averaged about six points difference, while there were virtually no differences on the verbal scale.

Herman (1979) argues that today's children are "brighter." In rates of language acquisition, the differences between contemporary children and age peers of a few years ago are larger for younger children. This is consistent with the finding that differences in verbal IQ fall off after age 11 (interestingly, at about sixth grade — the last grade in *LAB* level 2).

Lazar and Darlington (1982) similarly report that IQ gains made by children who participate in early school intervention projects are not permanent. They are largest after one or two years. The authors claim that IQ and achievement tests are related indices of developed abilities. The early intervention promotes cognitive and noncognitive skills, such as attentiveness, perserverence, etc., which influence IQ and school achievement.

The aptitude gains which Herman cites may result from cultural influences (books, toys, television, etc.) which act like the early school projects. If so, they may well drop off after a certain period. Tests measuring these aptitudes would show large (and perhaps nonlinear) grade effects. Interestingly, Lazar and Darlington (1982) report such achievement test gains by program participants, but attribute these to methodological artifacts.

Troike (1981) believes that IQ and achievement scores may be largely determined by acculturation to middle class American. If so, the patterns described by Title I and bilingual education students may be accounted for by a cultural interference hypothesis. Younger children make larger cultural gains in the same time period simply because they have less experience with (and less interference from) other cultural values.

This is a simplistic elaboration of Troike's hypothesis. In fact, other factors may be operant in acculturation of various grade levels. For example, the hypothesis does not account for the significant grade effect in Level 1. In the original study analysis, (repeated measured) pretest-posttest growth was significantly larger for second graders (F (1,247)=11.72, $p < .001$), when by hypothesis, it should have been smaller.

The First Grade Anomaly

The linguistic interdependence hypothesis (Cummins, 1979) states that a child's first-language competence at the time of intensive exposure to the second language predicts second-language acquisition. Cummins (1981) also believes that skills in both languages are predicted by a common underlying proficiency. The *WISC-R* would certainly be sensitive to such a construct. Moreover, this underlying abillty may well contribute to achievement in math and reading.

Cummins' two hypotheses are consonant with the data. First graders with more program experience actually score lower than those with less program experience. More experienced students may have been prematurely exposed to intensive English. By second grade, these differences favor students with more experience. Perhaps any growth-inhibiting effects of premature English exposure have been overcome (third-year second graders scored highest) because the bilingual program and maturation sufficiently strengthen the native-language skills by the third year of the program.

The interdependence of language skills explains the nonlinear age (year of birth) function. Verbal skills, which are more developed among contemporary children, stimulate cognitive growth which aids further verbal skills development (Cummins, 1979).

Summary

The current paper easily raises as many questions as it answers. Much of the scoring variance on language acquisition tests is determined by program effects. However, there are obviously non-program effects which contribute to this variance as well.

Primarily, there is a general grade effect, in which the amount of gain relative to test norms, is typically very large in the early grades. Younger children are not outscoring older children, but are scoring higher compared to grade peers from years ago than are older children. This effect is, at least in some part, related to IQ increases among today's children, particularly in verbal scales.

The relationship between grade and score improvement appears to be nonlinear. However, this could not be shown with the current data, in which there was no effect for grade within any level.

The large differences in test scoring among test levels probably has several components. Since this instrument is not vertically equated, one cannot assess if the difficulty of the levels are different for bilingual populations. However, one must question whether variations in test difficulty have a larger effect on depressing the performance of LEP students than on monolingual English students.

Recent test literature and achievement score analyses suggest that the grade effect is not so easily explained. In fact, the scores are so similar to findings in other

areas that improved verbal abilities of today's children may well explain various findings.

Finally, longitudinal research shows that interventions which raise aptitudes may not be permanent. Perhaps factors such as improvements in cultural stimulation available to contemporary children account for improved performance on IQ tests and achievement tests up to the sixth grade. Beyond sixth grade, the flatter effect of improvements over grades is attributed to the constant influence of the bilingual program.

References

Cumbo, R.F.; O'Neill, F.J.; Tilis, H.S.; and Weichun, W. *The Language Assessment Battery.* Iowa City: Riverside Publishing Company, 1976.

Cummins, J. "Linguistic Interdependence and the Educational Development of Bilingual Children." *Review of Educational Research, 49*(2), Spring 1979, 222-251.

Cummins, J. "Four Misconceptions about Language Proficiency in Bilingual Education." *National Association of Bilingual Education Journal, 5*(3), 1981, 31-45.

DeMauro, G.E. *Basic Skills Preventive and Remedial Program, 1978-1980.* Trenton, New Jersey: New Jersey State Department of Education, 1980.

DeMauro, G.E. *Achievement in Basic Skills Improvement, 1981.* Trenton, New Jersey: New Jersey State Department of Education, 1981a.

DeMauro, G.E. *The Impact of Bilingual Education on English Acquisition in New Jersey.* Trenton, New Jersey: New Jersey State Department of Education, 1981b.

Doppelt, J.E. and Kaufman, A.S. "Estimation of the Differences between WISC-R and WISC IQs." *Educational and Psychological Measurement, 37,* 1977, 417-424.

Douglas, D. and Johnson, D.M. "An Evaluation of Title VII Evaluation: Results from a National Study." Paper presented at the meeting of the National Association of Bilingual Education, Boston, March, 1981.

Echternacht, G. "Title I Evaluation and Reporting System: Development of Evaluation Models." *New Directions for Testing and Measurement, 8,* 1980.

Herman, D.O. *The WISC-R, Its Development and Usage: Some Findings from Sequential Standardizations.* Paper presented at the meeting of CASSP/NASP, San Diego, March 1979.

Krashen, S.D.; Long, M.A.; and Scarcella, R.C. "Age, Rate and Eventual Attainment in Second Language Acquisition." *TESOL Quarterly, 13*(4), 1979, 573-581.

Lazar, I. and Darlington, R. "Lasting Effects of Early Education: A Report from the Consortium for Longitudinal Studies." *Monographs of the Society for Research in Child Development, 47* 1982, (2-3, Serial No. 195).

Snow, C.E. and Hoefnagel-Hohle, M. "The Critical Period for Language Acquisition: Evidence from Second Language Learning." *Child Development, 49,* 1978, 1114-1128.

Tallmadge, G.K. *Interpreting NCEs; ESEA Title I Evaluation and Reporting Systems Technical Paper #2.* Mountain View, California: RMC Research Corporation, 1976.

Terman, L.M. and Merrill, M.A. *Stanford-Binet Intelligence Scale: Manual for the Third Revision, Form L-M.* Boston: Houghton Mifflin, 1972.

Thorndike, R.L. and Hagen, E. *Measurement and Evaluation in Psychology and Education.* New York: Wiley, 1969.

Troike, R.C. "SCALP: Social and Cultural Aspects of Language Proficiency." Paper presented at the Language Proficiency Assessment Symposium, Warrenton, Virginia, March 1981.

Wechsler, D. *Manual for the Wechsler Intelligence Scale for Children.* New York: The Psychological Corporation, 1949.

Wechsler, D. *Manual for the Wechsler Intelligence Scale for Children—Revised.* New York: The Psychological Corporation, 1974.

SYMBOLIC FORMS BETWEEN YOUTHFUL SAME-SEXED, CLOSEST FRIENDS.

S. Kirson Weinberg, Ph.D.
Sociological and Psychological Associates

Rita M. Weinberg, Ph.D.
Department of Psychology
National College of Education

Abstract

In this inquiry, we analyze the relationship between the rudimentary linguistic forms and same-sexed closest friendship among college youths. Our concern is to determine the types of verbal and gestural symbols that arise and characterize this mode of socially intimate relations. The forms of address between intimate equals are differentiated from those between status unequals and between socially distant equals; the specific types of personal address between closest friends and their meaning for intimacy are presented. The situations and circumstances which influence the spontaneous formation of idiomatic verbal forms and gestures and the categories of these improvised and expressive symbols are described and their effects upon the intimate relationship elaborated. The generic hypothesis between socially intimate relations and cognition and modes of speech is characterized and distinguished from the theory of established culture as heritage and its cumulative property as it affects language and thought.

SYMBOLIC FORMS BETWEEN SAME-SEXED, CLOSEST FRIENDS

Symbolic expressions of special words, phrases and gestures are aspects of cognitive communication between same-sexed, closest friends. Since social relations are the matrices from which symbolic forms as meaningful consensus emerge, the modes of verbal and gestural expressions within personal relations would differ from the verbal and gestural forms in impersonal relations (Mead, 1934, 1938). Consistent with this perspective, the social styles of speech, the turns of phrases, the "slangy" terminologies arise from the ways social relations arrange themselves (Sapir, 1949). This approach is applied to the study of the modes of linguistic expressions of same-sexed, closest friendships among single undergraduate college students in the urban community. Our specific concerns include 1) the variations between the elaborate language and argot systems in specialized subcultures and the sparse and simple verbal forms in closest personal relations; 2) the modes of address between persons of unequal status, and those of equal status as well as between social intimates; 3) the situations and circumstances as contexts within which words, phrases and gestures are incorporated into their peculiar stream of communication; 4) the types of words, phrases and gestures that characterize the social intimacy between closest friends; and 5) the effects of these modes of expressive language upon these relationships.

The subjects consisted of single, undergraduate students from four urban colleges and universities.[1] The subjects were urban-reared, 98 percent native American with some foreign-born who had immigrated during childhood, single or never married, predominantly white — 97 percent — and in the broad middle class of white collar and skilled laboring or artisan fathers. The age interval of 18 through 25 was used for these integral reasons: First, they were accessible. Second, they were in a time span when they were disengaged or disengaging from the family and were not completely committed to a mate. As individuals, they are reaching out for peers, so that the same-sexed friendships are perhaps the closest of any that they have formed or will form in the future. By disengaging from the family, their friendships are individually rather than family motivated and activated. Since peer adolescent relationships generally share a pervasive subculture, their modes of interaction also share a distinctive language at variance from that of the adult subculture. This adolescent subculture with its emphasis on informality provides the basis for the modes of speech between closest friends who tend to incorporate selected words, phrases and modes of speech to express their social intimacy.

Since this particular inquiry is a phase of a larger study, the information analyzed was drawn from 210 semi-guided personal documents. The subjects who composed these documents were asked to describe the development and crucial features of their friendship. The aim of the analysis of these documents was to determine the degree of social intimacy between the closest friends. From these personal documents we constructed a series of questionnaires and scales for measuring specific traits of social intimacy. One questionnaire concerning the linguistic expressions of social intimacy between same-sexed, closest friends was administered to 51 subjects.

Closest friendships, as we found, vary considerably in degree of social intimacy. Hence, their modes of speech may also vary. The same-sexed, closest friends comprise the peers selected by the subjects because only each subject can designate his or her closest friend. The degree of social closeness between these closest friends then became the objective of inquiry, and the modes of speech between closest friends would reflect the variations in their social intimacy. This paper, however, is to present the linguistic expressions of the closest friends who are indeed intimate because those closest friends, who are somewhat distant despite pretensions of a close relationship, have

[1] The four colleges and universities included in this study are Roosevelt University, Portland State University, Loyola University of Chicago, and National College of Education.

symbolic expressions similar to others in their age and sex categories

II

An interest in the idiomatic verbal symbols between friendly intimates as a phase of linguistic inquiry may be regarded an extension of the study of specialized languages and argots of subcultures including ethnic, deviant and occupational types. The languages of foreign groups as ethnic units settling in and becoming transplanted into the American pluralistic urban community were noted as integral and even central parts of their specialized subcultures. The basic feature of these linguistic systems is their cumulative character of an established tradition and heritage (Thomas and Znaniecke, 1921; Wirth, 1927.) The techniques used in studying ethnic subcultures and their languages were extended to the study of deviant and occupational argots. Thus elaborated argots were synthesized for varied criminal groups such as the pickpocket, the professional thief, the career criminal and the drug addict (Maurer, 1936, 1938, 1955; Sutherland, 1937; Lindesmith, 1947; Weinberg, 1942). The argot as a specialized form of linguistic system was noted too among doctors, lawyers, and other groups (Becker, 1964).

The distinctive trait of these argots is their cumulative property, in that the language grows and changes through a process of accretion of "new generations" of recruits as these become incorporated into specialized units, and they in turn add words to the specialized vocabulary, as these terms are transmitted further to recruits and members. Verbal accumulation is lacking in the sparse symbolic forms generated between same-sexed, closest friends whose particular modes of expression arise *de novo* from the closeness of their relationship and dissipate with its dissolution or disruption.

III

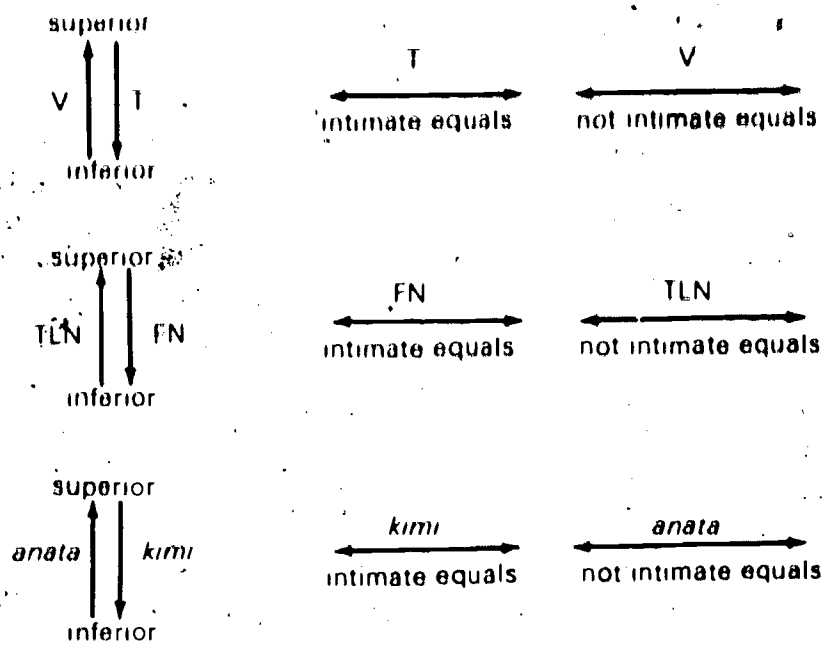
At the outset, it is evident that in several cultures, e.g., French, German, Japanese, Lower English, the modes of address for social intimates vary from those in unequal status positions and in some degree from equals generally. Status equals, whether intimate or distant, address each other in French by "Tu," while superior and inferior address each other by "Tu" and "Vous," respectively. A superior is addressed by a title, e.g., Mr., Mrs. or Miss, and a last name (TLN), while the subordinate is addressed by a first name (FN) (Brown, 1965; Silverberg, 1949; Fay, 1920). Equals address each other at least by their first names, but intimate equals may use abbreviated first names as well as nicknames. In Japanese the pronoun for you, "Kimi," is the mode of address for inferiors but also between equals, while "anata" terms for you, and in German for "sie" is the address to superiors in contrast to "du" and to strangers. Brown has regarded these modes of address as invariant norms in linguistic behavior. (See Chart I).

The change of address between dyads from Vous to Tu represents a change from unequals to equals. In Germany this change in mode of address is done by a rite of passage, a brief informal ceremony called the "Bruderschaft." One person who waits for a congenial mood, probably over a glass of wine, suggests: "Why don't we use Tu to address each other?" The response is favorable and means a shift in image that is reciprocal. However, one necessarily non-reciprocal phase of the relationship exists, specifically that one of the persons most be explicit in verbalizing the suggestion. But in the American society, these German modes of address do not exist. Indeed, it is characterized by informality which is especially pervasive in the adolescent subsociety.

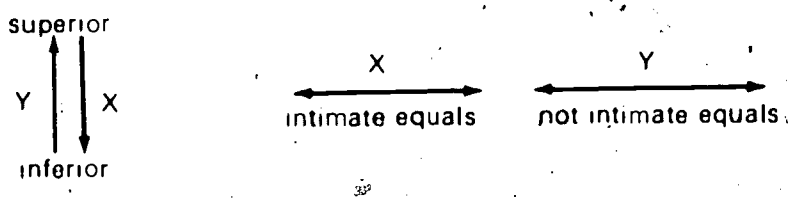
Among our subjects, 55 percent addressed their closest friends by full or abbreviated first names, while 37.3 percent addressed their closest friends by a variation of a first name or a last name — e.g., Mac for Mackenzie — or by a nickname. Only 3 percent addressed their friends by their last names and one percent used the formal title of Mr. or Miss. Most of those who addressed or were addressed by their closest friends with nicknames had known them from childhood. Since the preponderant 85 percent of the subjects had lost their childhood friends, a minority of the subjects as we have indicated used a variation of the first name or a nickname. Seemingly, those in the upper socioeconomic stratum are more likely to sustain their friendships from childhood; they call each other in private by their nicknames or twisted first names.

Chart I

Forms of Address among Superiors and Inferiors, Intimate Equals and Non-Intimate Equals.*



a. Examples of the invariant norm



b. General form of the invariant norm

* From Roger Brown, *Social Psychology*, (New York: The Free Press, 1965.) pp.93.

Another mode of address among equals and especially between intimates is the diminutive, characterized in the suffix by an added "y" or "ey." For example, Joseph is abbreviated to Joe and in the diminutive becomes "Joey." Gertrude is abbreviated to Gert and in the diminutive becomes "Gertie." The subjects intermittently added the diminutive to the abbreviated first name, but few used the diminutive continually, except for certain names, such as Jerry for Gerald or Jerome. The subjects felt that the diminutive is more appropriate for other socially intimate relations, particularly between parent and child. The subjects as adolescents and youths seemingly wanted to grow out of their childhood role, "the kid stuff." Hence they regarded a diminutive reference as not desirable continually, although occasional references were acceptable, even desirable. Another influence in the use of the diminutive is the recourse of some politicians to accept the informal diminutive of their first names. President Carter insisted upon the use of the name "Jimmy," rather than Jim or James. Governor Brown has acquired the name "Jerry," while former Vice-President

Nelson Rockefeller was called "Rocky." But it should be emphasized that in daily interaction with subordinates and formal associates the politician is not called by his diminutive first name anymore than President Eisenhower in his daily relations was called "Ike."

Does the mode of address affect the self-conception of the friend in the relationship? The reciprocity of an informal role-identity of friendship by the manner of address indicates the closeness of the relationship and, at times, the manner of affection from and for the friend. An abbreviated first name may create the opening of an informal relationship and also a feeling of potential approval and acceptance. Although the diminutive of a first name is used infrequently and for some friends not at all, its use pertinent to a given mode of affection indicates a reaction of closeness and emotional warmth. Parenthetically, between nonintimates, it can be used and interpreted as a sign of familiarity and depreciation. None of the subjects used the diminutive first name in this latter manner.

IV

We have determined that about 70 percent of the subjects are socially intimate with their closest friends. About 23 percent were informal, but not socially intimate with their closest friends, and about 7 percent seemingly lacked same-sexed, closest friends (Weinberg, 1970). Among the present group of 51 subjects, all of whom claimed to have closest friends, 45.1 percent and 35.3 percent claimed that their modes of speech differed "markedly" and "slightly," respectively, from those with same-sexed peers generally; 19.6 percent of the subjects averred that their mode of speech with their closest friends did not differ from that with other same-sexed peers.

Table I

Mode of Difference in Speech between Same-Sexed, Closest Friends and between Others of Same Age and Sex

Mode of Difference	Percent (N = 51)
Markedly Different	45.1
Slightly Different	35.3
Not Discernibly Different	19.6
Total	100.0

Since continuous, socially intimate relations between closest friends intensify mutual confidence and trust so that more personal secrets and private reactions are disclosed, they tend to develop an "In" frame of reference. As a consequence, their mode of communication becomes condensed so that a word or a phrase or at times, even a gesture may connote an elaborate episode of experiences. This kind of talk also excludes others as "outsiders" from understanding the communication. This type of close friendship and a consequent mode of communication is described in the following statement: "Our relationship during the major part of our friendship was somehow characterized by a mutual exclusion of other friends. I spent most of my spare time with Diane and participated in few activities or groups without her. Besides Diane, I had no other close friends, though I knew some Girl Scouts. But these friends did not know Diane, who was not a Scout. So, most of the time I spent with Diane was unaccompanied by others. I remember sharing secrets with her which we pledged never to repeat to others. Since neither of us had other close friends, it was very likely that we confided to a great extent. We devised a code by which we could relay messages through the classroom. The code was part of our secrets which was known to no one but us. Because of the way we acted, we recognized each other as best friends and were identified as best friends by all our school mates."

Since closest friends are informal, they are also spontaneous with each other and careless of their mode of speech and sentence structure or syntax.

"I think there is complete lack of formality between Edith and I. When I talk about Edith I usually call her by her first name, but when I talk to her, I usually call her Edie. She calls me Maureen or Maur.

"I certainly talk freely in just about everything, and I don't think that either of us hesitates to express ourselves better, or more formally. We say what we want to say when we want to say it. One reason I feel is the cause of our lack of sentence structure and very frequent use of symbols is the fact that Edith's parents are foreign-born and Edith is not very good in English. She gets her sentences mixed up sometimes, or sometimes she uses the wrong word to express herself. I do not hesitate to laugh when she expresses something in a funny way. I do not know how Edie takes my joking around like this, but I feel that others who do not understand her language handicap will even be worse than I, and I feel that she should trust in a friend. What I am trying to say is that when I laugh at her, I am not laughing at her but at what she says. Sometimes when I make errors in my speech, I kind of compare her errors with mine. In this way, I am hoping that she will not feel bad when she makes these errors in front of strangers or those she wishes to make a good impression upon."

V

Since spontaneous and expressive speech emerges from close friendship, the mode of conversation may become idiomatic on this level. These verbal and gestural forms arise during circumstances of relaxation; such as during and after parties or dances, during a stay at the friend's home, 31.4 percent; on weekends or when up late, 35.3 percent. The friends slough their speech, relax and regress. They resort to silly behavior, as the following instance illustrates.

Table II

Circumstances Stimulating Spontaneous and Informal Speech between Same-Sexed, Closest Friends

Circumstances	Percent (N = 51)
Before, during or after Party, Dance or Celebration	31.4
Staying Up Late, At Friend's Home, Weekend Meeting	35.3
Being "Together a Lot"	19.6
After Argument or Trouble	7.8
No Response	5.9
Total	100.0

"Whenever possible one of us would spend the weekend at the other's home. We followed a pattern of activities each time. We ate, danced to the latest records, set each other's hair, then ate some more and read Mad comic books all hours of the night and started using gushy words. Then we would start laughing ourselves sick, turn off the light and laugh some more. Some of the time we were in absolute hysterics over absolutely nothing. Our parents or one of them would then tell us to stop acting like kids. We would then try to settle down and go to sleep."

In another situation, the movies, the following two friends reacted in hilarity and generated a word of singular meaning to the two friends.

"Mary and I have always had words and names for each other. I recall one typical experience that led to new words. Once we went to the movies and saw a jungle picture; the story was about a man's struggle against red ants, or marabunta. For some reason we thought the situation was hilarious. We came away from the movie roaring with laughter and never forgot the word. For years afterward, one of us would interject the word "marabunta" into serious conversation and it would provoke laughter. People around us were baffled by our behavior and wondered what we were laughing about."

Thus the more hilarious and relaxed the circumstances, the more apt are the friends to select or to distort words about which to laugh which would lack the inside idiomatic meaning between the friends.

Another characteristic of communication is the manner of expression between closest friends that occurs during relaxed, hilarious, and regressive experiences. Words may be slurred and syntax may change to incoherence. Thus one subject admittedly omitted the "r's" from words so that "better" became "betta," harder" became "harda," "bother" became "botha," and this change was known by the other friend as an acceptable vocal mannerism.

This spontaneous speech leads to spontaneous distortion of words or to syntax incoherence, as the following characterization attests.

"We speak incoherently, in a sort of shorthand that others have a hard time following. Neither of us are very organized in speech; and through our knowing each other we have developed patterns of speech which we use mainly when we are alone with each other."

Closest friends who are socially intimate differ in their speech with each other from their mode of speech with others. As can be seem from Table III, the subjects claim that they more frequently use slang, swear and employ pet phrases. In addition, those persons in bilingual families incorporate these bilingual words into their dialogue. Some have devised new words and even phrases.

Table III

Modes of Speech Used by Subjects More Frequently with Closest Same-Sexed Friends Than With Peers of the Same Age and Sex

Mode of Speech	Percent (N=73)*
Slang Expressions	24.7
Swear Words	20.5
Pet Phrases	19.2
Foreign or Bilingual Words or Phrases	12.3
Newly Devised Words or Phrases	9.6
Other:	
"Open-up" More	4.1
Sarcastic and Teasing	4.1
Seemingly No Difference	5.5
Total	100.0

*The number exceeds 51 because the subjects used more than one "mode of speech."

In addition to the words and pet phrases that some closest friends incorporate into their dialogue, they also may adopt certain physical gestures which may have a specific or even a unique or idiomatic meaning for them. The specific meaning of a given gesture may be similar to the interpretation which others in the adolescent culture express. The types and frequencies of the gestures indicate that innovative and original gestures tend to be sparse between closest friends. By

far, most gestures are through eyes, 27.1 percent and fingers, 28.8 percent, with the body generally, 18.6 percent; the mouth, 13.6 percent; and the hand, 11.9 percent (Table IV).

Table IV

Types of Gestures Used in the Interaction between Same-Sexed, Closest Friends

Type of Gesture	Percent (N=59)
Fingers Insert finger in mouth (disapproval), Number of fingers for time and place of meeting	28.8
Mouth Speak inaudibly with slow mouth movements; Quick, short smile (Bad); Prolonged smile (Good)	13.6
Eyes Roll eyes (Confused) Blink eyes (Approval) Shift eyes from side to side (Doubt)	27.1
Hand Insert fist under chin (disapproval) Place hand on chest	11.9
Body Generally Bow to each other Japanese style (No specific meaning) Hug friend (general approval) Pat on the back (no "lie")	18.6
Total	100.0

*Since subjects used more than one gesture, the total is more than the number of subjects, 51. Six of the subjects indicated that they used no particular gestures that have a special meaning for them.

VI

We have pointed out that social intimacy between same-sexed, closest friends tends to influence a mode of speech, consisting of words, word enunciation and gestures, that serves to sustain and reinforce the intimate relationship. On the other hand, the recourse to critical, hostile words or phrases serves to cool and recede or even disrupt the relationship.

On a positive level, informal, spontaneous symbolic interaction pertinent to an intimate friendship relationship may contribute to such personal gratification that the expressive process makes the mode of speech and the substantive symbols end-values. By contrast, the impersonal, instrumental or utilitarian relationships are means to ulterior ends so that the modes of speech and words used are designed for effect to achieve a different objective.

When the mode of speech, modes of address and the words used become formal, hostile or depreciating, then the relationship may cool and recede to one of distance. But when the modes of address and words used emotionally support and enhance the self-esteem, the relationship may become sustained in its closeness.

The verbal symbols as "idioms of social intimacy" imply an in-group reaction as well as attitudes of solidarity and unity. The friends as interactants realize that they alone have the meaning of these terms so that others are "outsiders." The more frequent the use of idiomatic words and gestural signals, the more pervasive the close friends represent an "in" social unit (Slater, 1963; Newcomb, 1961.)

The spontaneous recourse to slang, slurred and swear words characterize verbal forms of informality and ease of presence as well as expressive release and catharsis in some situations. Hence these substantive aspects of communication indicate the interpersonal moods of trust and relaxation between the friends. These types of expression in interaction make the relationship satisfying and pleasurable and tend to intensify the bond between the friends.

Conclusions

The demonstrated consistency between socially intimate friendships of the same sex and the modes of emerging symbolic forms as an object of inquiry is a discontinuous extension of the study between culture, thought and language. This approach to linguistic forms is unlike the conception of language as an aspect of a growing, accruing cultural heritage which is transmitted to incoming novices and reinforced among those of the same generation. Instead this relational approach to the formation and retention of verbal and gestural forms emphasizes that these arise from the modes of intimate interaction, lack any cumulative characteristics and dissipate with the receding closeness of the friendship. The cultural approach to cognition and language is consistent with the Durkheimian approach and developed largely by Sapir (1949) and Whorf (1956). The relational approach is consistent with the Meadian view (Durkheim, 1951; Schneider and Homans, 1955; Barnow, 1973; Mead, 1934; Weinberg, 1952; Gibson, 1969; Horowitz, 1970; Neisser, 1976.)

The modes of intimate social interaction are the bases for distinct if fragmentary types of idiomatic linguistic forms and gestures that arise from and reinforce this type of relationship. We have presented the circumstances and types of verbal forms that arise in this interactional context. Although Whorf has emphasized that the language as an integral aspect of the culture influences the modes of thought, we have found in socially intimate relations the improvised and

spontaneous terms that are spoken are influenced by the interpersonal feeling, tone, and uncritical expressive thinking. Thus the words and gestures as these are incorporated into their substantive interaction are consequents of moods, especially euphoric ones, and of permissive, even regressive thinking.

References

Barnow, Victor. *Culture and Personality.* Homewood, Illinois: The Dorsey Press, 1973. pp. 80-88.

Becker, Howard. *The Outsiders.* New York: Free Press, 1964.

Brown, Roger. *Social Psychology.* New York: Collier-Macmillan Ltd., 1965.

Clemmer, Donald. *The Prison Community.* Boston: The Christopher Press, 1940.

Cole, R. A. and Scott, B. "Toward a Theory of Speech Perception." *Psychological Review.* 81, pp. 348-374.

Durkheim, Emile. *Suicide.* New York: The Free Press, 1951.

Fay, P. B. "The Use of 'Tu' and 'Vous' in Moliere." *University of California Publications in Modern Philology.* 1920, 8, pp. 227-286.

Gibson, E. J. *Principles of Perceptual Learning and Development.* New York: Appleton-Century-Crofts, 1969.

Horowitz, M. J. *Image Formation and Cognition.* New York: Appleton-Century-Crofts, 1970.

Lindesmith, Alfred C. *Opiate Addiction.* Chicago: Principia Press, 1947.

Maurer, David W. "The Argot of the Underworld Narcotic Addict." *American Speech,* XI, 1936. pp. 116-127.

Maurer, David W. "The Argot of the Underworld Narcotic Addict." *American Speech,* XIII, 1938, pp. 179-192.

Maurer, David W. *The Big Con.* Indianapolis: Bobbs Merrill, 1955.

Maurer, David W. *Whiz Mob: A Correlation of the Technical Argot of Pickpockets with Their Behavior Patterns.* Gainesville, Florida: American Dialect Society, 1955.

Mead, George H. *Mind, Self and Society.* Chicago: University of Chicago Press, 1934.

Mead, George H. *The Philosophy of the Act.* Chicago: University of Chicago Press, 1938.

Neisser, Ulric. *Cognition and Reality.* San Francisco: W. H. Freeman and Company, 1976. Chapter VIII. "Perceiving and Using Speech."

Newcomb, Theodore M. *The Acquaintance Process.* New York: Holt, Rinehart and Winston, 1961.

Sapir, Edward. *Culture, Language and Personality.* Berkeley, California: University of California Press, 1949. pp. 15-17.

Schneider, D. M. and Homans, G. C. "Kinship Terminology and the American Kinship System." *American Anthropologist,* 57, 1955, pp. 1194-1208.

Silverberg, W. V. "On a Psychological Significance of 'du' and 'sie'." *Psychoanalytic Quarterly,* 1940, pp. 509-519.

Slater, Phillip E. "On Social Regression." *American Sociological Review,* June, 1963, pp. 349-350.

Sutherland, Edwin H. *The Professional Thief.* Chicago: University of Chicago Press, 1937.

Thomas, William I. and Znaniecke, Florian. *The Polish Peasant in Europe and America.* Chicago: University of Chicago Press, 1927.

Weinberg, S. Kirson. "Aspects of the Prison's Social Structure." *American Journal of Sociology,* 1942.

Weinberg, S. Kirson. "Primary Group Theory and Closest Friendship." *Human Nature and Collective Behavior.* Edited by T. Shibutani. Englewood Cliffs, N. J.: Prentice-Hall Inc., 1970.

Whorf, Benjamin Lee. *Language, Thought and Reality. Selected Writings of Benjamin Lee Whorf.* Boston: Technology Press of Massachusetts Institute of Technology, 1956.

Wirth, Louis. *The Ghetto.* Chicago: University of Chicago Press, 1927.

THE INFLUENCE OF PUBLIC POLICY ON LANGUAGE ASSESSMENT OF BILINGUAL STUDENTS

Rudolph C. Troike
University of Illinois at Urbana-Champaign

From a policy-oriented point of view, it is interesting to see how concern in recent years for language assessment in bilingual education has been primarily the consequence of legislative and judicial decisions, rather than of purely educational decisions. If we recognize that all educational policy is fundamentally political (indeed, the word *politica* in Spanish comprehends both *policy* and *politics),* this situation should perhaps not be surprising. Nevertheless, it is true that few other areas of the educational arena have been so directly affected by the force of legislative and judicial events as this one.

Two strands can be discerned in the policy decisions which have led to the present demand for language assessment and language assessment instruments. One of these is accountability to funding sources or courts, and the other is determination or limitation of access to bilingual services. Policy decisions requiring language assessment arise from several sources:

1. Congressional legislation,
2. Court decisions,
3. State mandates.

We shall examine each of these in turn below.

Congressional Legislation

The original bilingual Education Act (PL 90-247), which was signed into law by President Johnson in 1968, was the first major piece of federal legislation to recognize the special educational needs of children of limited English-speaking ability. By providing funds for demonstration programs, which were required to conduct evaluations, the Act created a limited demand for language assessment instruments to measure student progress.

The 1974 amendments to the Act (PL 93-380) restated its aim as being to establish equal educational opportunity by supporting programs designed to enable students, while using their native language, "to achieve competence in the English language" (Sec. 702a). Discussions about "maintenance" programs to the contrary, notwithstanding, the attainment of competence *in English* was specifically one of the declared goals of federal policy in providing program assistance for limited English-speaking students, with the use of the native language merely serving as a means to that end.

The 1978 legislation (PL 95-561) narrowed this focus even further by defining a bilingual program as one in which instruction was designed "to allow a child to achieve competence in the English language" (Section 703a). This should have dispelled — though it did not — any illusions or misconceptions, either by opponents or proponents of bilingual education, that the legislation supported maintenance of other languages or linguistic separatism. The title "Bilingual Education Act" was clearly a misnomer; it would have been more accurate to refer to it as the "English Assimilation Act" since it was designed, in fact, primarily as an instrument for linguistic assimilation.

The 1978 amendments also explicitly revised and expanded the designation of the target population in two ways: (1) by changing "limited English-speaking ability" (LESA) to "limited-English proficiency" (LEP), which included reading and writing, and (2) by allowing the inclusion of students from non-English language background (NELB) homes, even though the students might be, at least superficially, fluent English speakers. The latter provision enabled American Indian communities which had largely lost their native languages to apply for Title VII funds, since research had shown that their English was still limited in ways that caused academic difficulties.

The emphasis on English in Title VII has continued to the most recent steps toward reauthorization, with a proposed Administration bill which would specifically allow funding for nonbilingual instruction in English, and another bill by Sen. Huddleston, which would require program exit based on English-proficiency test scores.

Although at no time have evaluation guidelines for Title VII-funded programs *required* English-proficiency tests, these have widely been used as a means of assessing program effectiveness. The further use of English proficiency to screen students in program intake was given a strong impetus by the Congressionally adopted limitation on percentages of non-LEP students included in programs, a provision sparked by the finding of the American Institutes for Research national evaluation (Danoff, 1978) that large numbers of (apparently) competent English speakers were enrolled in Title VII programs. Since lack of English proficiency had been the motivation for offering bilingual education in the first place, it should not be surprising that the attainment of proficiency in English should be seen as the basis for qualifying or not qualifying for inclusion in a program. At the same time, the move to adopt English proficiency as a criterion for program exit, as part of a policy of limiting bilingual services, also appears plausible though research (Cummins, 1980a and b; Troike, 1978) suggests that it is only speciously so.

The Civil Rights Act of 1964, under Title VI of the act, forbade discrimination on the basis of race, color, or national origin in the operation of federally assisted programs. In an important memorandum issued on

May 25, 1970 by the then Director of the Office of Civil Rights J. Stanley Pottinger, the protection of the Act was extended to cover children who were "deficient in English language skills" by virtue of their national origin minority background. Schools were required to provide equal educational opportunity to such children and were forbidden to discriminate on the basis of language. This broadened interpretation of the Civil Rights Act brought about compliance efforts on the part of OCR, though they were not strongly pursued until the *Lau* decision by the Supreme Court gave them added force.

Court Decisions

The most far-reaching court decision was that by the Supreme Court in the case of *Lau* vs *Nichols*, which held that the Civil Rights Act (and the OCR interpretation of it) was being violated by failure to provide special assistance to students of limited-English ability. The opinion of the Court, written by Justice Douglas, observed that "Basic English skills are at the very core of what these public schools teach . . . students who do not understand English are effectively foreclosed from any meaningful education."

Although the Court specified no remedy, the Office of Civil Rights developed guidelines for enforcing compliance with the decision, known as the "Lau Remedies." These guidelines specified that school districts found in noncompliance must identify students' primary or home language and their functional ability in the home language and English. This requirement, more than any other, coming as it did with the weight of compliance with a Supreme court decision, stimulated the national search for valid, reliable, and easily administered instruments to classify students by their English ability.

Specific court decisions, most notably the *Aspira* consent decree in New York City, have had the effect of reinforcing the requirements for language assessment at the local level.

State Mandates

Beginning with the State of Massachusetts in 1971, states have passed legislation providing for or mandating bilingual education. Many of these acts or their enabling regulations specify language assessment criteria for program placement, and some states also specify criteria for program exit.

In Illinois, school districts are required by the State Board of Education to have a district assessment procedure plan which must specify the criteria and instruments to be used, training for administration, and a description of procedures to be followed. The annual bilingual census classifies students into those categories according to their level of English.

A number of states have mandated the use of particular language tests which are considered acceptable, such as the LAS (Language Assessment Scales), BSM (Bilingual Syntax Measure), or BINL (Bilingual Inventory of Natural Language). Most states require or encourage multiple criteria for program placement or exit, though in some cases English test scores are the sole criterion.

A recent departure, taken in California and Texas, is to require the use of a standardized achievement test given in English, rather than a language proficiency test per se. (This move accords with my own belief that such a test provides a far better integrative assessment of the language necessary for successful achievement at a given level than would an isolated language-focused instrument.)

Discussion

We have seen that the growth of language proficiency testing for English, with notably lesser attention to other languages, was initially stimulated by the Bilingual Education Act of 1968 and given particular impetus by the *Lau* decision and related court actions. State mandates have added to the demand for tests, though the withdrawal of the proposed *Lau* regulations by the Department of Education in early 1981 and the shift of some states to requiring standardized achievement tests may signal that the demand has now peaked and will begin to decline. However, possible moves in some states and even nationally to tighten exit requirements in terms of language proficiency, being pushed most strongly by those who see bilingual education primarily as an assimilative device, could in fact increase the demand for tests.

Thus the market for English language assessment instruments has been created almost entirely by federal and state legislation and regulations and court decisions. (A few ancillary demands have appeared, such as the need for instruments and procedures to collect and validate census data on the number of limited-English-proficient persons of school age in the country, but this was generated by the same Title VII legislation in 1978). It is significant from the larger social perspective, however, to recognize the lack of interest nationwide in such tests on the part of major testing organizations and marketers. Test development and marketing in the area of language proficiency has been largely a cottage industry, in spite of the size of the potential market created by these legislative and judicial actions. This situation results, I believe, from the deep-seated antipathy toward bilingual education which is endemic in the educational-industrial complex and which ultimately reflects the hostility toward cultural and linguistic differences that characterizes our profoundly xenophobic society. No other explanation on purely rational economic grounds would account for this apparent lack of interest.

To return to the point with which we began, the demand for language assessment has been created by the imposition of public policy decisions on the educational system. Had educational institutions and educators been appropriately concerned about the best interests of their linguistic minority students, such decisions would not have had to be imposed from outside. Policies requiring bilingual education have

been perceived (and denounced) as political because they threaten the status quo — obscuring the fact that the status quo itself is political. The existing power relationships are fundamentally political and serve to determine allocation of resources and policy development within the system. However, these matters are rarely viewed as political since it is not in the interests of those in control to have them recognized as such.

One unfortunate fact about educational policies developed outside the educational system is that, desirable as they may be, they sometimes lack the research base or educational realism that policies which arise from informed educational practice might have. In this instance, educational practice was little better informed than legislative and judicial decision making.

As I have suggested earlier, the whole policy of language assessment for program entry and exit is seriously flawed, for it gives unwarranted status to isolated language skills as indicators of ability to "participate effectively in the educational program" (*Lau v. Nichols*). As recent research has shown (Saville-Troike et al., 1982; Cummins, 1980 a and b), we, in fact, know little about this relationship, certainly not enough to justify some of the educational decisions which are being made on the basis of language-proficiency assessment alone. What is needed is closer cooperation among policy-makers, researchers, and educators to develop more rational, effective, and educationally sound policies and procedures for their implementations. In the meanwhile, educators should insist that language measures alone are not enough for making educational decisions, and that other educationally grounded criteria must be used in addition, or even instead.

References

Cummins, James. "The Cross-Lingual Dimensions of Language Proficiency: Implications for Bilingual Education and the Optimal Age Issue." *TESOL Quarterly*, 14(2): 175-188, 1980a.

_____. "Psychological Assessment of Immigrant Children: Logic or Intuition?" *Journal of Multilingual and Multicultural Development*, 1(2):97-111, 1980b.

Danoff, Malcolm N. "Evaluation of the Impact of ESEA Title VII Spanish/English Bilingual Education Program: Overview of Study and Findings." Palo Alto: American Institute for Research, 1978.

Saville-Troike, Muriel, Erica McClure, and Mary Fritz. "Communicative Tactics in Children's Second Language Acquisition." In Fred Eckman (ed.), *Proceedings of the University of Wisconsin Symposium on Universals of Second Language Acquisition*. Milwaukee, 1982.

Troike, Rudolph C. "Research Evidence for the Effectiveness of Bilingual Education." *NABE Journal*, 3(1):13-14, 1978.

THE ISSUE OF LANGUAGE PROFICIENCY TESTING AS A REQUIREMENT FOR BILINGUAL EDUCATION TEACHER CERTIFICATION

Guillermo Duron
Governors State University, Park Forest South, Illinois

Introduction and Statement of the Problem

The issue of language proficiency requirements for bilingual education teachers has been one of controversy since the first Title VII programs were initiated. While no one will disagee that bilingual teachers must possess some minimum level of proficiency in the language or languages in which they are expected to teach, disagreement arises when the details are addressed. Various answers have been given to the following questions:

1. What are the areas of language to be addressed in setting minimum requirements (oral, written, vocabulary, syntax, etc.)?
2. Should the minimum language proficiency testing be conducted in the language of instruction or in both the target language and English?
3. Should the minimum language proficiency requirements be the same both in English and the target language of certification?
4. Should the criteria vary according to target language (i.e., less stringent requirements for languages for which there is a teacher shortage)?
5. What criteria will be used to determine language proficiency?
6. Who will set the minimum language proficiency requirements?

Several individuals and organizations have examined the issue of language proficiency requirements of bilingual teachers. A review of the relevant literature reveals that publications in the area of teacher certification of bilingual/bicultural teachers tend to encompass general guidelines which state education agencies or institutions of higher education can use to develop their own policies and procedures. In addition, all of the states which offer certification or endorsement in bilingual education require some evidence of proficiency in a language other than English.

In 1974 the Center for Applied Linguistics published a set of guidelines for teacher preparation and certification in bilingual/bicultural education. Included in these guidelines are language proficiency criteria which the authors felt should be required of bilingual/bicultural teachers. The guidelines indicated that bilingual/bicultural teachers should demonstrate the ability to:

1. Communicate effectively, both in speaking and understanding, in the languages and within the cultures of both the home and the school. The ability will include adequate control of pronounciation, grammar, vocabulary, and regional stylistic and nonverbal variants appropriate to the communication concept.
2. Carry out instruction in all areas of the curriculum using a standard variety of both languages. (Center for Applied Linguistics, 1974. Pp. 2-3.)

The National Association of State Directors of Teacher Education and Certification has adopted several standards which they designated as being appropriate to all college programs preparing teachers in bilingual/bicultural education. Standard Two addresses the area of language proficiency. As stated, "The programs shall develop and/or evaluate the ability of the prospective teacher to function and instruct students both in English and in the language of the target student population with fluency and accuracy, good pronounciation and intonation." (Waggoner, 1976, p. 41.)

Acosta and Blanco (1978) recommended that students completing a bilingual teacher education program should have a command of English and the target language in the areas of listening, speaking, reading and writing. They recommended that graduating students be able to provide instruction using standard language both in English and the target language as well as possess comprehension and speaking ability in the language variety of the student. They strongly encourage the use of an exit language proficiency test evaluating both general language skills and educational terminology in both languages.

National Perspective

In a survey conducted by the National Center for Education Statistics, Waggoner (1976) found that eleven states had adopted special requirements for bilingual education teachers.* All eleven required proficiency in a language other than English, while five required proficiency in English (see Table 1). Regulations concerning language proficiency requirements varied from a single sentence indicating that oral and written proficiency in the target language was required to a full description of criteria used to determine competencies in language skills.

*Delaware has since prohibited bilingual education.

172

Table 1

TYPE OF BILINGUAL LEGISLATION, CERTIFICATION AND REQUIREMENTS FOR TEACHERS IN BILINGUAL/BICULTURAL PROGRAMS IN THE U.S.[2]

STATE	Legislation Permissive/Mandatory	Bilingual Education Certification/Endorsement	Language Proficiency Requirement (English/Target)
Alaska	M		
Arizona[1]	P	C	E&T
California[1]	M	C&E	E
Colorado	M		
Connecticut	M		
Florida	P		
Idaho	P		
Illinois[1]	M	C	E&T
Indiana[1]	P	E	E
Iowa	P		
Kansas	P		
Louisiana	P	C	E&T
Maine	P	C	E&T
Maryland	P		
Massachusetts[1]	M	C	E&T
Michigan[1]	M	E	E
Minnesota	P		
New Hampshire	P	E	
New Jersey[1]	M	E	E&T
New Mexico[1]	P	C	E
New York	P		
Ohio	P		
Oregon	P		
Pennsylvania	P		
Rhode Island[1]	M	E	E
Texas[1]	M	C	E&T
Utah	P		
Washington	P		
Wisconsin	M		

[1] Were found to require language proficiency according to Waggoner (1976)

[2] Information compiled from Waggoner (1976). Gray (1981) and Woellner (1982)

The most brief and general references made concerning language proficiency of bilingual education teachers were found in the regulations of the following states:

Indiana — "The candidate for the bilingual and bicultural endorsement must show oral and written proficiency in the target language."

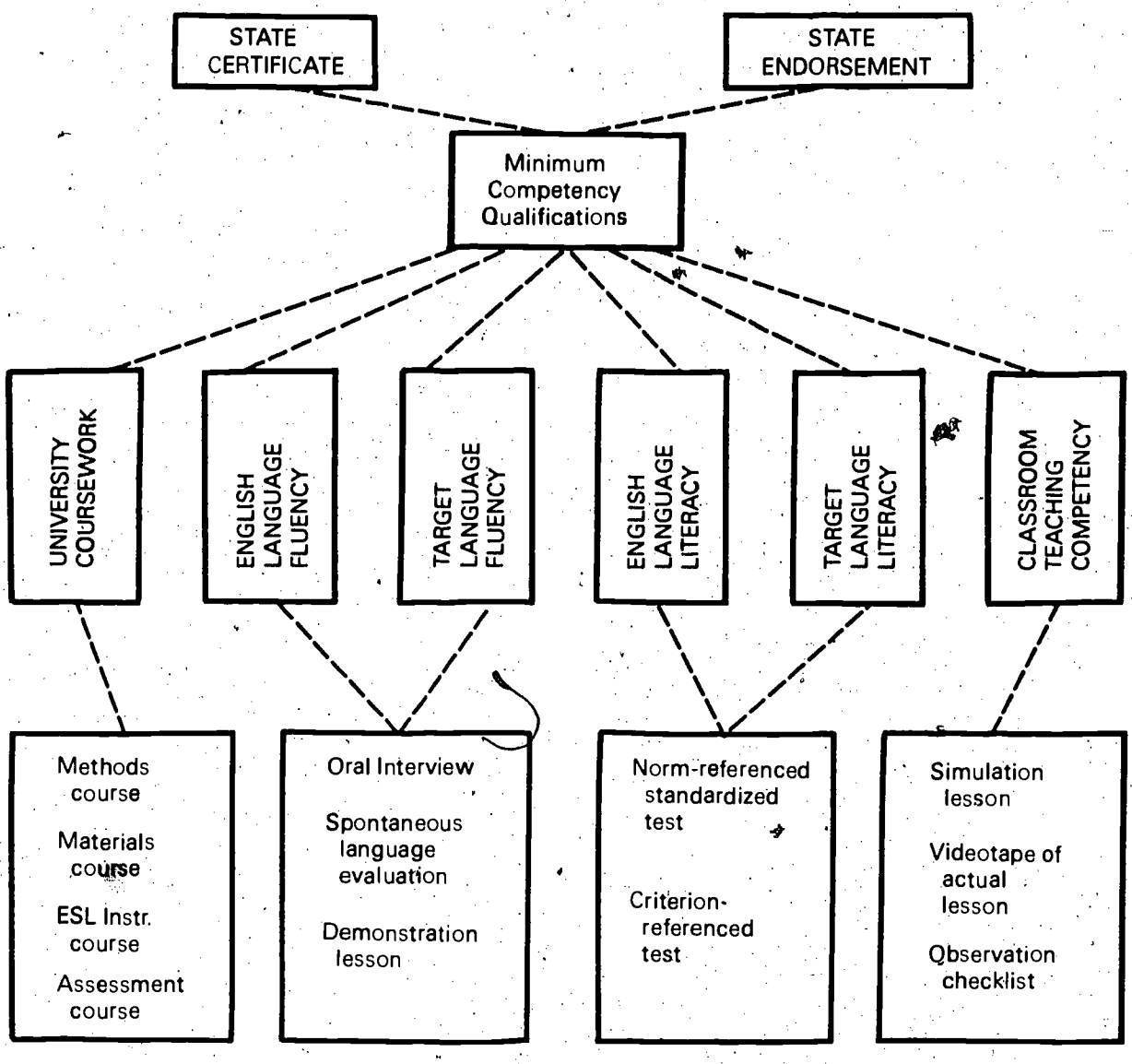
New Jersey — "The requirements are as follows: demonstration of verbal and written proficiency in English and in one other language used as a medium of instruction." (Waggoner, 1976, p. 29)

In the State of Delaware, language proficiency could be demonstrated *through completion* of a "minimum of fifteen semester hours in the language area at or above the third year college level or demonstrate fluency in the language area of assignment as determined by the NTE score at the 50th percentile." (Waggoner, 1976, p. 15)

Rhode Island also required a demonstrated proficiency in speaking, reading and writing in a language other than English. Proficiency would be attested to by the Commissioner of Education at approved colleges or universities in Rhode Island. In Arizona, proficiency in the second language or English must be verified by the language department of a regionally or nationally accredited institution. The New Mexico Department of Education included several specific requirements in its regulations. Candidates in New Mexico must demonstrate fluency in the local dialect as measured by observation in the field as well as an extended functional vocabulary of classroom terminology and literary skills. In Texas, proficiency in English and the target language at the highest level for which certification is taught is required. Proficiency is determined by a public school committee as designated by the Texas Education Agency (Waggoner, 1976).

Figure 1

BILINGUAL TEACHER CERTIFICATION MODEL VARIABLES



Waggoner (1976) describes California's requirements and states that target language proficiency must be verified in oral comprehension, aural comprehension, reading and writing. Proficiency is verified through successful completion of an examination covering each of these areas. The examination may be administered by the Commission for Teacher Preparation and Licensing, a commission-approved California institution or the governing board of a local educational agency or county superintendent of schools.

Since the Waggoner study, several states have introduced bilingual education legislation with provisions related to bilingual teacher certification (see Table 1). Maine, New Hampshire and Louisiana have specified that bilingual teachers must possess skills and demonstrate proficiency above and beyond that which is required of general elementary and secondary school teachers. The three states mentioned above have utilized the bilingual teacher certification models devised prior to 1975 with individual alterations. All of the states currently requiring certification use one or a combination of the variables listed in Figure 1 to evaluate bilingual teacher candidates.

Illinois Perspective

In the State of Illinois bilingual educators and certification specialists have been actively investigating the area of language proficiency of bilingual teachers for the past fifteen years. When the first bilingual programs were initiated, teachers were desperately needed to fill the newly created bilingual positions. Since most of the first programs were created in response to a need to serve limited-English-proficient Spanish-speaking children, teachers were mainly recruited from the ranks of Spanish language departments.

Requirements in the early years included fluency in the target language and residence of two or more years in a country where the target language is used. Another means for satisfying certification requirements was to have been raised in a home where the target language was spoken.

Any attempt to require minimum language proficiency and to determine whether teachers met those minimums was done informally at the district level. More often than not, school districts did not have anyone on staff who could make those judgments concerning the target language. As a result, teachers were hired for bilingual education positions who possessed varying degrees of proficiency in the target language and in English.

No formal testing was conducted in the early years, rather school officials were asked to make use of the advisory council in determining the language qualification of teachers. Later, Office of Education monitors would visit the schools and informally evaluate the teacher's language proficiency (Seelye, 1976).

In 1976 the Illinois State Legislature passed the Transitional Bilingual Education Act (Illinois School Code, 1976). Included in this legislation were provisions for transitional bilingual education certification. This was manifested as a nonrenewable six-year certificate which was required of all teachers employed in mandated transitional bilingual education positions. The mandate required fluency in the target language and communicative competence in English.

Following the adoption of the Illinois bilingual mandate, procedures were developed to conduct language proficiency testing of bilingual teachers. Candidates were to be tested in English if they had graduated from an institution of higher education where a language other than English was used as the medium of instruction. If the candidate had graduated from an instituion where English was used, the candidate was tested in the target language. The assumption made was that graduation from an institution was sufficient evidence that the person was proficient in the language used for instruction in the institutional setting.

In accordance with the *Rules and Regulations for Transitional Bilingual Education* (1976), testing was to be conducted only in the areas of speaking and reading. Educational Testing Services's (ETS) *Foreign Service Institute (FSI) Survey* was selected as the measurement tool to be used in evaluating oral proficiency. Several persons affiliated with the Illinois Office of Education or the Bilingual Education Service Center were trained in conducting and evaluating oral interviews by the ETS. The *Spanish Language Proficiency Test*, published by the Modern Language Association was adopted for evaluating Spanish reading ability and the *Nelson-Denny Reading Test* was chosen to test reading ability in English.

In an effort to determine the minimum proficiency level to be required for demonstrating fluency leading to certification, the State took language samples from employed bilingual teachers and evaluated their language abilities using the FSI Language Proficiency Interview Scale (see Table 2). Teachers were examined in both English and the target language and proficiency levels were found to range from Level Two to Level Five in both languages. Because of the shortage of teachers, minimum proficiency was set at Level Two in English and Level Four in Spanish based on the FSI scale. The lower proficiency level required in English was justified based on the contention that teachers conducting classes all day in the target language need not be highly proficient in English. Therefore, it was decided the "communicative competence in English meant that they (teachers) could just function rudimentarily in the school building" (Seelye, 1976, p. 38).

Table 2

LEVELS OF PROFICIENCY OF THE LANGUAGE PROFICIENCY INTERVIEW
(Educational Testing Service, 1979)

LEVEL	STANDARD
Level 1	Able to satisfy routine travel needs and minimum courtesy requirements
Level 2	Able to satisfy routine social demands and limited work requirements
Level 3	Able to speak the language with sufficient structural accuracy and vocabulary to participate effectively in most formal and informal conversations on practical, social, and professional topics
Level 4	Able to use the language fluently and accurately on all levels normally pertinent to professional needs
Level 5	Speaking proficiency equivalent to that of an educated native speaker

Six institutions have been approved by the Illinois State Board of Education to serve as testing sites which administer the Transitional Bilingual Education Certification Examination. Testing site approval has been consistent with approval for certification programs in bilingual education. The institutions which now serve as testing sites are: Governors State University, Illinois State University, Mundelein College, Northern Illinois University, University of Illinois at Chicago Circle, and Western Illinois University. In the last three years, however, only Governors State University and Northern Illinois University have offered the examination on a regular basis.

Since the establishment of the first testing sites, procedures, instrumentation and standards have changed and/or vary from site to site. One institution requires testing in both English and the target language, while the others only test in one language. Although the MLA and the *Nelson-Denny* are used at most institutions, the *Prueba de Lectúra* and criterion-referenced tests are also used by approved sites. Since most of the testing sites were established several years after the original procedures were introduced, many of the individuals involved in testing did not take part in the training provided by ETS. In 1980, the State Board sponsored a special training session conducted by ETS. However, not all individuals involved in interviewing and evaluating oral proficiency were willing or able to participate. The fact that all test evaluators and interviewers have not received the same training may result in discrepancies and less than perfect reliability in the evaluation results.

The Illinois State Board of Education is presently engaged in the process of changing certification requirements for Transitional Bilingual Education Certification. Included in the changes are requirements in coursework in bilingual education as well as language proficiency. It seems, therefore, that this would be an ideal time to reevaluate the existing language testing procedures and implement changes which would ensure that:

a. certified bilingual teachers are available in sufficient numbers in all languages,

b. the language ability of newly certified teachers is sufficient to enable them to function in bilingual classrooms, and

c. teachers who can function in bilingual classrooms are able to pass the testing criteria.

In addressing these issues, consideration must be given to the questions presented in the opening paragraph of this paper. Each of these issues will now be discussed and suggestions made concerning implementation of changes in language testing procedures.

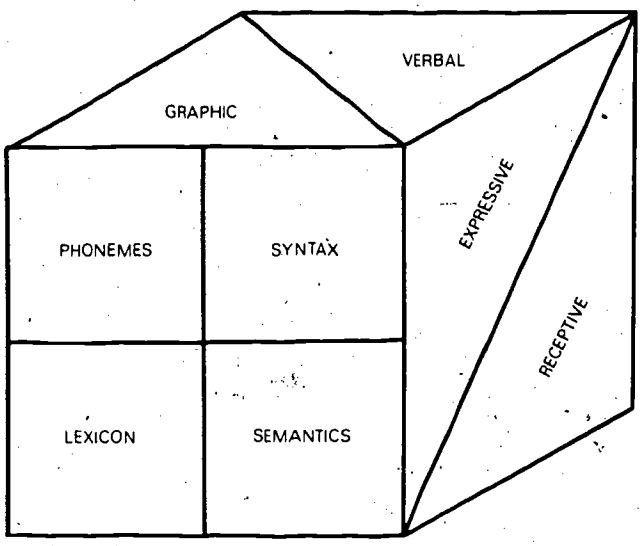
Language Testing Issues and Implications

- LANGUAGE AREAS TO BE MEASURED

In order to convey a sense of the elements of language proficiency testing, Figure 2 includes a graphic representation of the components of language. Theorists conclude that language is composed of four elements. They include: phonemes (sounds), lexicon (vocabulary), syntax (grammar) and semantics (meaning). These elements are communicated through expressive and receptive modes and can be conveyed through vocal or graphic channels. The testing of language proficiency should measure all elements (directly or indirectly) in both modes and both channels. Thus, a bilingual teacher candidate should be tested in the areas of understanding, speaking, reading and writing utilizing instruments which measure the phonemic, syntactic, lexical and semantic elements of language. Since these are usually integrated in language use, separate tests need not be administered for each area. However, instrumentation should be analyzed to insure that all elements are measured in the total test battery. These procedures would insure that successful candidates have an adequate sound system, sufficient vocabulary, proper use of grammar and an understanding and ability to convey meaning when using the language. These characteristics are basic to the communication which is needed to be successful in a bilingual education classroom.

Figure 2

COMPONENTS OF LANGUAGE



- LANGUAGES TO BE TESTED

If a bilingual education teacher is to be successful, s/he must be able to communicate with everyone s/he deals with. Since interactions occur among monolingual English speakers as well as persons who are monolingual in the target language, the bilingual education teacher must be proficient in both languages. Graduation from an institution where a specific language is spoken has not proven to be sufficient evidence of a person's proficiency in the language used in that institution. Similarly, college credits earned in language courses cannot verify a person's language capabilities. The testing of language proficiency in both languages utilizing valid procedures with valid standards appears, at this time, to be the best method of assuring that a potential bilingual education teacher has the skills necessary to be successful in a teaching situation.

- LANGUAGE PROFICIENCY REQUIREMENTS IN ENGLISH AND THE TARGET LANGUAGE

The level of proficiency required in terms of minimum standards should be determined separately in each language. These levels should be based on the language behavior expected in each language. Reading requirements should be based on the difficulty levels of the material the teacher is expected to read in each language. Similarly, the different contexts in which the teacher must speak English and the target language must be taken into consideration when determining the minimum oral proficiency level required. These minimums can be determined separately through the procedures described in a later section of this paper.

- VARIATION OF LANGUAGE PROFICIENCY REQUIREMENTS

The great shortage of bilingual education teachers that existed in the early 1970's no longer exists. Spanish/English bilingual programs for the most part are able to hire qualified, certified personnel. A shortage does, however, exist in some languages such as Hmong and Laotian. The Illinois State Board of Education should maintain an annual record of such shortages and deem special status to those languages for which teachers are scarce or difficult to hire. Special temporary certification could then be granted to qualified persons in these languages. A qualifier to this certification would allow for the teacher to teach only in the target language until passing the regular certificate examination. At that time, individuals who did not previously achieve the required cut-off score or above would receive the appropriate credentials for bilingual teacher certification.

- CRITERIA USED TO DETERMINE LANGUAGE PROFICIENCY

One of the most essential characteristics which all testing procedures must possess is validity. "Validity refers to the extent to which the results of an evaluation procedure serve the particular uses for which they are intended" (Grohlund, 1976, p. 79). If the results are to be used to determine the ability of a person to function linguistically in a bilingual classroom, then the criteria should reflect the specific behaviors which are needed to provide as accurate an estimate of future success as possible. Testing procedures for bilingual education certification purposes should attempt to evaluate the candidates' language proficiency in relation to the following:

1. Ability to read texts and reference manuals;
2. Ability to communicate with students, parents, teachers and administrators;
3. Ability to present oral instruction in a clear and understandable manner;
4. Ability to write to parents, teachers and administrators;
5. Ability to write original content for lessons.

The above criteria should be used in selecting instruments and in establishing the total test battery. Measures should be used to obtain an estimate of the candidate's understanding, speaking, reading and writing ability within the social and educational contexts in which the language will be used. Although actual observation of the candidate within these environments would not be feasible, the content of the instruments must reflect an adequate sample of the universe of language behavior required of bilingual teachers.

- SETTING THE MINIMUM LANGUAGE PROFICIENCY REQUIREMENTS

Livingston (1978) described an eight-step procedure for conducting an empirical standard-setting study. This procedure can be outlined as follows:

1. Identify the behaviors to be evaluated.
2. Identify valid measures of those behaviors.
3. Identify judges qualified to evaluate the performance of an individual.
4. Select a sample from the population of persons for which the standard is to be set.
5. Obtain evaluations from the judges.
6. Analyze the data and establish the probability of error.
7. Consider the seriousness of the two types of errors: passing a candidate whose performance is inadequate and failing a candidate whose performance is adequate.
8. Establish the standard at a point determined by the acceptable probability of reducing the two types of errors listed in #7 above.

Language proficiency requirements in Illinois should be established utilizing these procedures. The behaviors to be evaluated are the language areas mentioned earlier in this paper. Specific tests, including modification of the FSI meeting the criteria listed in THE CRITERIA USED TO DETERMINE LANGUAGE PROFICIENCY section of this paper, should be identified and developed. A group of judges can be identified which are representative of all persons qualified to evaluate

the adequacy of a candidate's language proficiency. These may include bilingual teachers, coordinators, principals, education specialists and teacher trainers fluent and literate in the language to be addressed.

Test scores from the language measures selected can be collected from a sample of potential candidates at one of the present testing sites. The data from this sample can then be analyzed following the acceptability judgments technique described by Livingston (1978). These procedures determine the probability of errors is considered scores. Once the seriousness of the two types of errors is considered and an acceptable probability of error determined, this information can be used to set the minimum cut off scores for the various measures.

Conclusion

The state of the art of language proficiency testing for bilingual teacher candidates is one which can be characterized by diversity, flexibility and fluctuation. The impact of federal and state legislation which ultimately dictates whether bilingual programs will be mandatory or permissive greatly affects the status of bilingual education teacher certification. Fluctuation in public opinion which is reflected by legislative manifestations contribute to the many changes that have been experienced by states since the mid-1970's.

Language proficiency testing for teacher certification is developmental in most states which combine one or more variables to achieve the model best suited to local needs. These variables include university coursework, English and target language fluency, English and target language literacy, and classroom teaching competency.

In one state, Illinois, bilingual teacher certification has changed considerably over the past several years, and combinations of the above variables have been operationalized. Presently, efforts are underway to reexamine certification criteria using a multicriteria approach. Each state must look to its own requirements and resources to determine the best certification plan. However, a plan that inculcates the language testing issues discussed in this paper in response to the questions found in the introduction would have universal application for bilingual education teacher certification.

Bibliography

Acosta, Robert K. and Blanco, George. *Competencies for University Programs in Bilingual Education*. Washington, D. C.: Government Printing Office. HEW OE 78-07903. 1978.

Center for Applied Linguistics. *Guidelines for the Preparation of Teachers of Bilingual Bicultural Education*. Arlington, Virginia: CAL Publication, 1974.

Educational Testing Services. *Training Manual for Language Proficiency Interview Testers*. Princeton, New Jersey: Educational Testing Services, 1979.

Gray, Tracy, et al. *The Current Status of Bilingual Education Legislation: An Update*. Bilingual Education Series #9. Arlington, Virginia: Center for Applied Linguistics, 1981.

Gronlund, Norman G. *Measurement and Evaluation in Teaching*. Third Edition. New York: MacMillan Publishing Co., Inc., 1976.

Illinois State Board of Education. *The School Code of Illinois*. Article 14-C, 1976.

Livingston, Samuel A. "Setting Standards of Speaking Proficiency." In *Direct Testing of Speaking Proficiency: Theory and Applications*. John D. Clark. Editor. Princeton, New Jersey: Educational Testing Services, 1978.

Seelye, H. Ned. "The Illinois Experience." Paper presented at the New York State Conference on Bilingual Education in Colleges and Universities. ERIC Document #ED168337. 1976.

Waggoner, Dorothy. *State Certification Requirements for Teachers for Bilingual Education Programs*. Washington, D.C.: U.S. Government Printing Office. 1976.

CURRICULUM PLANNING AND INSTITUTIONALIZING BILINGUAL EDUCATION: SOME RECENT DATA AND MODELS

Richard E. Baecher
Fordham University at Lincoln Center, New York

While retrenchment and scarcity of resources are rapidly changing the formal higher education landscape, the process of institutionalizing an innovation such as bilingual education within a post-secondary education setting in the years ahead needs careful examination. Bakalis (1981, p. 103) refers to the consequences of economic decline as the "New Austerity" which forces us to ask again an old question: What is the purpose of education? In particular, what is the purpose of teacher and professional education programs within formal higher education? A recent publication from the Office of Bilingual Education and Minority Language Affairs (OBEMLA) entitled *Beyond the Myths: Title VII and Bilingual Education in the United States* poignantly charts the decline of federal support for Title VII bilingual programs. For example, since 1981, Federal support for basic grants, training grants, and support services has plunged by 41 percent.

What appears to be happening today is a scenario captured by Schneider (1976, p. 162) in her summary and interpretation of the 1974 Bilingual Education Act as one of revolution, reaction or reform:

> A reactionary piece of legislation would have been one that totally deleted the federal program or one which adopted the Administration's initial proposal of folding Title VII and other categorical programs into a special revenue sharing approach. In that instance, the onus for bilingual-bicultural education programs would have been removed from the Federal government since the decision on the level of expenditures for bilingual-bicultural content would have been a proposal to meet the problems of the limited English-speaking not with a bilingual-bicultural approach, *but rather with English-as-a-second-language without any mandated cultural content.* (Emphasis ahead)

Although Schneider's analysis discounted this reactionary result in the case of the 1974 Bilingual Education Act, recent events, e.g., the Administration's attempt to convert Title VII into a block grant, the steady drop in federal funds, efforts to amend current bilingual education legislation (Baker and de Kanter, 1981), attest to the major shift in philosophical and pragmatic emphases among important policy and decision makers.

With regard to the topic of this paper — Curriculum Planning and Institutionalizing Bilingual Education — the literature and research are practically nonexistent. Except for scattered papers and symposia at national and regional conferences (See the bibliography for a listing.) and a recent descriptive survey of bilingual teacher training programs within institutions of higher education (IHEs) for 1978-1979 (Binkley et al., 1981), little information of explanatory and useful value is available to address these two crucial questions: (1) What critical factor(s) influence the process of institutionalization within an IHE?, and (2) How can the process of institutionalization endure in the period of the "new austerity"? Based upon a variety of approaches employing literature survey materials, selected concepts borrowed from educational change theory and current conceptual frameworks, my paper will delineate and suggest answers to these important issues pertaining to the past, present, and future efforts at institutionalizing bilingual education within higher education. In this paper, "institutionalization" refers to that socio-educational process whereby committed and knowledgeable members of the formal higher education enterprise loosely couple and systematically integrate selected properties of the field of bilingual education with the academic system of the university or college for the specific purpose of preparing bilingual professionals and specialists to intelligently and sensitively service language minority and majority students.

ISSUE ONE: What critical factor or factors influence the process of institutionalization within IHEs?

This question has been a major concern of OBEMLA, IHEs, and in particular, directors of bilingual programs and faculty involved in bilingual programs since the first federal grants became available to post-secondary institutions under the 1974 Bilingual Education Act. Section 723 of the 1974 Act provided the impetus for the issue:

> "(A) (i) training, carried out in coordination with any other programs training auxiliary educational personnel, designed (I) to prepare personnel to participate in, or for personnel participating in, the conduct of programs of bilingual education, including programs emphasizing opportunities for career development, advancement, and lateral mobility, (II) to train teachers, administrators, paraprofessionals, teacher aides, and parents, and (III) to train persons to teach and counsel such persons, and (ii) special training programs designed (I) to meet individual needs, and (II) *to encourage reform, innovation, and improvement in applicable education curricula in graduate education, in the structure of the academic profession, and in recruitment and retention of higher education and graduate school facilities, as related to bilingual education...* (Public Law 93-380, Sec. 723, cited in Schneider, 1976, (p. 218).

Table 1

INDICATORS OF PROGRAM INSTITUTIONALIZATION
(Adapted from Binkley et al., 1981)

INDICATOR	COMMENT
1. Administrators actively support program.	1. Manifested through allocation of institutional funds, tenure; formation of bilingual advisory committees by deans; "professional regard" in which director and bilingual education faculty are held in the IHE.
2. Most nonbilingual faculty have positive attitudes toward program, or at least do not create obstacles to institutionalization.	2. Although administrator support is more critical to survival, negative attitudes of a large number of nonbilingual faculty can be detrimental; an area of serious concern as formal higher education "downsizes" in the years ahead.
3. Some faculty are supported by institutional funds.	3. & 4. Related to indicator #1; more obvious indicator of institutional commitment.
4. Some faculty are on tenure track.	
5. Program is reported to continue in absence of Title VII funds.	5. "Expressed opinions are held opinions"; administrators/directors report continuation of program without Title VII funds.
6. Program operation involves the efforts of several professional staff and is not dependent on one or two prime movers.	6. This point is debatable, depending upon the size of institution and degree to which program is coupled with existing university structures. However, one or two faculty members must be responsible for the management of the program along with its academic quality.
7. Program is compatible with established institutional priorities.	7. Degree to which program's goals correspond to traditional purposes of a university — instruction, research, service; likewise, state bilingual education certification legitimizes the existence of the program within the School of Education.
8. Enrollment levels would be high enough to sustain the program.	8. Probably the most important indicator of program institutionalization. Level of Title VII support of tuition and financial support at the institution, geographic proximity of competing Bilingual Education programs, and state certification all differentially influence enrollment; another area of consternation as Title VII support for training is reduced.

Table 2

VARIABLES INFLUENCING INSTITUTIONALIZATION
(Carey and Marsh, 1980, pp. 20-21)

Variable	Scale		
	5	3	1
Characteristics of the Project			
1. The extent that the proposal includes the innovation as an aspect of project.	The innovation is an explicit goal for the project as stated in the proposal.		The innovation is not mentioned nor implied in the project proposal.
2. Proposal emphasizes institutionalization of the innovation as goal.	Institutionalization of the innovation is an explicit goal for the project.		Institutionalization is not mentioned nor implied in the proposal.
3. The innovation is highly interconnected with other project activities (in practice).	The innovation is highly interconnected with other project activities.		The innovation is not connected to other project activities.
Characteristics of the Institutionalizer			
1. The institutionalizer is highly affiliated with the project.	The key institutionalizer is a member of the core project staff.		The key institutionalizer was not affiliated with the project.
2. The institutionalizer is a member of the regular faculty.	The key institutionalizer is a tenured faculty member.		The key institutionalizer is a soft-money project person on the tenure track.
3. The amount of project staff time given to implement the innovation within the project.	Extensive staff time was given to implementing the innovation *within the project.*		Very little staff time was given to implementing the innovation *within the project.*
4. The amount of project staff time given to institutionalizing the innovation.	Extensive staff time was given to institutionalizing the innovation.		Very little staff time was given to institutionalizing the innovation.
Characteristics of the Innovation			
1. The innovation is an adaption of innovation developed/created outside the local setting.	The innovation came from outside with no local revision.		The innovation was entirely locally developed.
2. The innovation is a tangible product or program (degree of transportability).	The innovation is a tangible, easily transportable product.		The innovation is an amorphous entity which would be awkward to transport.

[illegible] 2 (continued)

Variable	Scale		
	5	3	1
Characteristics of the Institution			
1. The institutional leadership wants to institutionalize the innovation.	Leadership provided funds, extensive support, and expressed strong interest in the innovation.		Leadership did not provide funding; no expression of interest in innovation (or negative reaction).
2. Colleagues value and want to institutionalize the innovation.	Colleagues highly valued or wanted innovation.		Colleagues were neutral or opposed innovation.
3. The innovation does not violate the turf of colleagues.	No intrusion on existing faculty's turf.		Extensive intrusion on existing faculty's turf.
4. The innovation has the the promise of generating significant levels of revenue.	Innovation more than pays for itself when used by institution.		Innovation cannot pay for its use by institution.
5. The innovation is easily translated into the administrative building blocks of the institution (courses, programs, credentials, departments).	Innovation explicitly framed in terms of IHE courses, program requirements or credentials.		Innovation very difficult to translate in to IHE courses, program requirements or credentials.

In a subsequent study, Carey and Marsh (1980) used a step-wise regression analysis to clarify the relationship between the model's predictor variables, i.e. characteristics of the project and institution, and the achieved level of institutionalization for each Teacher Corps innovation. Their findings indeed shed some light on the issue of critical factors influencing educational change. These four predictor variables, accounting for 70% of the variance, were strongly associated with the achieved level of institutionalization. (They are listed in their order of importance.)

1. Colleagues value and want to institutionalize the innovation.
2. The amount of project staff time given to implementing the innovation with the project is important.
3. The institutionalizer is a member of the regular faculty.
4. The innovation is an adaptation of innovation developed/created outside the local setting. (pp. 38-39)

These results clearly indicate that each category of their model — characteristics of the project, of the institutionalizer, of the innovation, and of the institution — made an independent contribution toward accomplishing successful institutionalization. Among the implications of their studies, these three stand out: (1) the process of institutionalization entails a cluster of factors about the innovation, the institutionalizer, and the institution, respectively, and neither one single variable nor cluster of variables in their findings explained the *total* variance; (2) individuals interested in increasing institutionalization can manipulate these variables with some assurance that levels of institutionalization can be increased; and (3) the importance of organizational features in any effort to innovate.

On the basis of this review of the research and literature pertaining to critical factors that influence institutionalization within an IHE, one can conclude that a combination of factors needs to be identified. Although considerable overlap is evident in the three models presented, they all negate the fallacy of conceiving of institutionalization as a single, valued process. Further research is necessary to empirically verify any aggregate of factors (See Note 1). I recommend that IHEs explore these critical factors along with others not mentioned within their own geographic settings, the extent to which they have institutionalized their bilingual programs, and ways to sustain student enrollments in an era of economic

decline. The latter remark leads directly into the next issue of this paper.

ISSUE TWO: How can the process of institutionalization endure in the period of the "new austerity?"

This issue has many "angles" and "rough edges" to it, e.g., the attitude of institutional decision makers toward the meaning and effectivenss of bilingual education, the beliefs of bilingual faculty concerning their role and future within formal higher education, the extent to which the Federal government will downsize training grants in the next five years, how the State will respond to and advocate the need for qualified bilingual personnel, and the images we have of schools and the types of change that correspond to each (Firestone, 1980). The question also assumes that some form of bilingual education will persist, thereby raising the kinds of strategies that would be most successful, e.g., employing a "cooperative" strategy, consolidating current gains, appealing to private and corporate funding authorities, etc.

While the issue of whether or not bilingual education will survive in post-secondary institutions at all in the absence of Title VII funds is an important one, ultimately, its resolution will depend upon the commitment, capacity, and competence of the individuals who belong to the institution.

What steps, then, might be considered to maintain the process of institutionalization that will be a vital link in the education of language minority and majority students in this country? Since I do not have a crystal ball that spells out the following panacea, Easy Ways to Guarantee Your Professional Occupation in the Lean Years Ahead, I can only describe some of my own thoughts and activities that presently make sense of the organization to which I belong. As Firestone (1981, p. 461) reminds us, one ought not assume a universal strategy for planned change; "projects must be tailored to the specific settings in which they will be used." Hopefully, my ideas will assist readers in their own deliberations. The bibliography can be consulted for other sources such as Arciniega (1980) and Carey and Marsh (1981). I recommend these few steps (no priority given to the numbering):

Step I. View the process of institutionalization as an adaptive socioeducational process combining the best elements of loosely and tightly coupled systems within the academic system of the institution.

In my opinion, this view simply responds to Weick's (1980, p. 1) observation that "parts of the organizations are heavily rationalized, but many parts also prove intractable to analysis through rational assumptions." The image of "loosely coupled systems" captures the latter whereby structures/elements within the institutions, e.g., bilingual education and administration, preserve their own identity, uniqueness, and separateness over a period of time. The image of "tightly coupled systems" connotes the degree of bureaucratic, heavily rationalized linkages that are easily located within an institution, e.g., the bilingual teacher education office, with its directors, staff and other visible organizational elements. Another positive feature of this perspective is that it points out the fallacy that institutionalization can follow traditional models of educational innovation. To this writer, the image of "loosely and tightly coupled systems" expresses the fluidity and structural stability that characterize this process. Since 1975, the bilingual faculty at Fordham has been successful in "coupling" with the College at Lincoln Center to establish an undergraduate bilingual teacher education program; with the Division of Administration, Policy and Urban Education of the School of Education, to provide a professional diploma program for bilingual administrators and supervisors; with special education, to develop a bilingual special education master's degree; with the Division of Psychological and Educational Services, to establish a bilingual urban school psychology program; and finally, within the Division of Curriculum and Teaching, to provide a master's degree in bilingual bicultural education. The concept of "loosely coupled systems" allows one to be on the "lookout" for unexpected and potentially promising bonds. See Appendix A, "Fordham University Higher Education Programs" for more information (Justiz et al., 1980, pp. 157-160).

What is important about Step I is that each "institutionalizer" make explicit his/her own understanding of institutionalization along with "significant others" within the institution.

Step II. Secure approval and certification of your bilingual program(s).

Besides legitimizing the program within the institution, this step just might assure future enrollments. At Fordham, we have recently decided to "tightly couple" with the College at Lincoln Center and therefore to initiate a four year liberal arts program with certification in bilingual elementary education. We've come to realize that bilingual education at the elementary level is absolutely necessary. My own research (Baecher, 1981, 1982) and that of Cummins (1981) pertaining to the cognitive learning styles of bilingual children and the developmental aspects of bilingual proficiency, respectively, point to the need for early bilingual intervention in learning and augmentation of cognitive styles. And one place to begin is at the pre-service level, adequately and forcefully documented in Gaspar (1981) dealing with Hispanics and the U.S. economy, and the *Research Bulletin* of the Hispanic Research Center describing Hispanic American diversity (1981).

Step III. Don't abdicate your role as bilingual educators and leaders within your institution.

Vast transformations have been predicted for the educational system in the years ahead. While these changes are inevitable and possibly beneficial, thereby impacting bilingual programs and their curriculum

Another indication that the IPT ignores characteristics of L2 learners is the fact that language proficiency levels are correlated to grade levels. While it is possible to correlate a child's L1 language development with his/her grade in school, such a correlation in the case of the L2 learner does not reflect the same amount of exposure to the language. A child who is a new arrival and who is beginning to learn English in the 6th grade can not be expected to have the same language skills as a native, English-speaking child who is in the same grade. Furthermore, such a system fails to accurately distinguish among L2 learners in the same grade who may have differing amounts of exposure to English due to their date of arrival in the United States in the case of older immigrant children, or due to sporadic school attendance in the case of migrant children.

Construct Validity

The IPT has even more serious problems of a theoretical nature in that the basic construct which it purports to measure is not clearly defined. While the IPT is labeled as an oral language *proficiency* test, the word *mastery* is used interchangeably with proficiency throughout. The point is made that linguistic competence must be inferred from linguistic performance data and that the constructs which underlie linguistic performance and upon which the IPT is based are the following: a) language is developmental; b) it is incremental; c) it is systematic; d) it is used to communicate in a social context; and e) it involves both receptive and productive skills (Dalton, 1979, p. A-3). It is implied that some of the constructs which underlie linguistic performance, as described by the author, are inherent in the IPT due to its format which places the child in a dyadic social situation with the examiner whereby the child must comprehend and produce progressively more difficult language. From the child's responses, his/her linguistic performance is evaluated and the level of language proficiency is determined. Dalton states, "Validation of the above constructs can be inferred if the data generated through the (construct validity) study demonstrate positive correlation with the age and grade of the student and the IPT level results" (Dalton, 1979, p. A-4).

The construct validity study was conducted on 364 students in a suburban school district in California. An analysis was made of the correlation between age and IPT results on form A and B of the English version of the test. The data were further analyzed after excluding the scores of NEP/LEP students which resulted in a sample which was comprised of native speakers of English or L2 learners who were fluent speakers of English. The same procedures were followed to derive a correlation between IPT results and the students' grade level. Not surprisingly, the study obtained high positive correlations between IPT results and age and grade level with high levels of significance on both forms of the test.

Having shown that a strong positive correlation existed between the age and grade of native speakers of English and language proficiency in their own language, Dalton concludes that these data validate the constructs which state that language is developmental, incremental and systematic. In addition, the data are also said to validate the use of the IPT as a method of assessing oral language proficiency through linguistic performance.

The fact that there is a strong positive correlation between the age and grade in school of native speakers of English hardly seems to be a worthy subject for research. Furthermore, one has to wonder why the construct validity study was conducted and what its applicability as a rationale for a language-proficiency measure for L2 learners is. An even more basic problem with the study is its definition of linguistic performance which is composed of vague generalities which are not accurate. While language may be said to be systematic and incremental in its development, one's linguistic performance is not. Linguistic performance is a speaker's use of his/her linguistic competence (underlying knowledge of the system of a language) at any one time. In a native speaker of a language, linguistic performance may vary from one occasion to the next due to such factors as memory, distractions, shifts of attention, etc. (Maclay & Osgood, 1959, p. 24)

Language proficiency measures attempt to accurately measure one's linguistic competence by inference from linguistic performance data which is variable. Due to this fact, a more germane construct upon which to measure a test for L2 learners is, what is the nature of communicative competence? What knowledge must one possess in order to produce grammatically correct speech in situationally appropriate contexts? Such an approach requires the articulation and validation of the learner's stages of language acquisition in the areas of syntax, morphology, lexicon, phonology, functional language and sociolinguistic skills. A complete theory of communicative competence would also have to order and weight items to indicate which are the most crucial for achieving intelligibility in a language. While such a theory has not been fully developed to date, researchers in the area of language testing are working toward that end. Federal and state regulations which require language proficiency testing for children in bilingual programs have caused great strides to occur in the area in recent years. Unfortunately, the IPT does not reflect recent theory and suggested practices in the field.

Content Validity

Content validity is the extent to which a test covers a representative sample of behavior in the domain to be measured. The IPT purports to measure language mastery in the areas of syntax, lexicon, phonology, morphology, comprehension and oral expression and reports extremely high percentages of items which correspond to and measure each one. This stems from the fact that test items were scored for all of the above categories which could possibly pertain to a response and not to those which are reflected in the IPT scoring system. For example, the item, "Who is this"? which requires the student to supply the lexical item, *teacher*,

Burn, Barbara B. "Strength through Wisdom: The Report of the President's Commission on Foreign Language and International Studies." *Independent School*, 40, 1 (Oct. 1980): 11-17.

Cardenas, Jose A. *Keynote Address: National Conference on the Education of Hispanics*. Intercultural Development Research Association, San Antonio, Tex., 1978. (ED 182066)

Carey, Lou M. and Marsh, David D. *University Roles in Inservice Education: Planning for Change*. The American Association of Colleges for Teacher Education, One Dupont Circle, Washington, D.C., Feb., 1980.

Cummins, James. "The Role of Primary Language Development in Promoting Educational Success for Language Minority Students." In *Schooling Language Minority Students: A Theoretical Framework*, developed by Office of Bilingual Education, California State Department of Education. Los Angeles, CA: Evaluation, Dissemination and Assessment Center, California State U., 1981.

de los Santos, Alfredo G. *Hispanic and Community Colleges. Topical Paper No. 18*. Center for the Study of Higher Education, U. of Arizona, 1415 North Fremond Avenue, Tucson, AZ, 1980. (ED 184615)

Federal Register, Part 123f — Bilingual Education: School of Education Projects Vol. 45, No. 67, Friday, April 4, 1980.

Firestone, William A. "Images of Schools and Patterns of Change." *Am. Journal of Education*, 88 (August 1980): 459-487.

Garcia, Nilda, and Ortega, Jaime., comp. *Selected Papers from the Hispanic Conference (San Antonio, Texas, Feb. 14-16, 1980)*. Dissemination and Assessment Center for Bilingual Education, Austin, Texas, 1980. (ED 202265)

Gaspar, Jeffery. "The Hispanic and the U.S. Economy." In National Hispanic Center for Advanced Studies and Policy Analysis, *The State of Hispanic America*. Babel, Inc., Oakland, CA., 1981.

Hernandez, Norma G. *Institutionalizing Bilingual Education in Schools, Colleges and Departments of Education: Programmatic Concerns at Institutions of Higher Education*. Paper presented at the annual meeting of the AERA, Boston, MA., April 7-11, 1980. (ED 191643)

Hispanic Research Center. "Hispanic American Diversity." *Research Bulletin*, Vol. 4, #2, 3, April-July, 1981, Fordham University, Rose Hill, NY, 10458.

Justiz, Manuel; Resta, Paul; Hopkins, Thomas; Platero, Dillon. *A Directory of Hispanic and American Indian Higher Education Programs*. U. of New Mexico, New Mexico, 1980.

Martinez-Perez, Luis A. *Hispanic Students in Higher Education*. Paper presented at the National Conference on the Education of Hispanics, Alexanderia, VA., Aug. 20-23, 1978. (ED 178257)

Mulhern, John D. "A Dean Views Bilingual Education." *NABE Journal* 1, 2 (May 1977): 75-76.

OBEMLA. "Beyond the Myths: Title VII and Bilingual Education in the United States," Mimeographed, n.d.

Schneider, Susan G. *Revolution, Reaction or Reform: The 1974 Bilingual Education Act*. NY: Las Americas Pub. Co., 1976.

Seidner, Stanley S. "Language Assessment at Post Secondary Institutions." In R. Padilla (ed.) *Ethnoperspectives in Bilingual Education Research: Bilingual Education Technology, Vol. III*. Ypsilanti, MI.: Dept. of Foreign Languages and Bilingual Studies, Eastern Michigan University, 1981.

_____ "Language Assessment Practices and Policies at Colleges and Universities: Final Report of a National Survey of Bilingual Teacher Training Programs." In S. Seidner, *Issues of Language Assessment: Foundations and Research*. Springfield: Illinois State Board of Education, 1982.

Washburn, David E. *Bilingual/Bicultural Education in the United States: Higher Education*. Ethnic Heritage Studies Branch, DHEW, Washington, D.C. 1981. (ED 204325)

Weick, Karl. "Educational Organizations as Loosely Coupled Systems." *Administration Science Quarterly* (Mar. 1976): 1-19.

Weryackwe, Suzanne L. *Bilingual Education at the Graduate Level and Its Relevance to Culturally Pluralistic America*. Paper presented at the Annual Meeting of the AERA, Boston, MA., April 7-11, 1980. (ED 196617)

CURRICULUM PLANNING: IMPLEMENTING PROGRAMS THAT WORK

Anne M. Kiefer
Bilingual Education Service Center

ABSTRACT

The design and implementation of language assessment programs involve professional staff in a complex change process. Recent studies of program design and implementation suggest procedures which would convert that complex change process into a more specific series of steps toward a goal. They also provide a series of recommendations for program planners.

This paper will apply those recommendations to assessment processes and the decision making involved in language programs for limited-English-proficient students at a local school or district level. It will make use of the theories mentioned above as well as concepts of assessment procedures, effective programs, inservice and staff development, in addition to the core theories of language assessment which are the focus of this institute.

The summary will include a series of recommendations for an adaptive implementation process which takes into account: the level of certainty involved in the theories or technology of language assessment; the levels of stability in the environment; the kind of organizational structure (loosely or tightly coupled); the level of conflict among persons involved in the process (conflict over goals and means); and, finally, the scope of the change involved in the new program.

INTRODUCTION

The key to establishing programs of language assessment that work is to establish a flexible change process of implementation before beginning to design the program.

Programs of language assessment that "don't work" are those which are totally designed ahead of time with no provision made for the process of change. They are programs that have not allowed for the uncertainty and ambiguity involved in language learning and language assessment theory and technology. Their planners have not made provision for the conflict resulting from that uncertainty nor from differing ideas on teaching/learning processes, on evaluation and on instruments which are effective for assessment. These programs have acted as if schools and school systems are as tightly structured as are business organizations with each unit operating as a part of the whole. Thus they have not acknowledged educational organizations as "loosely coupled" (Weick, 1982), nor have they made provisions for the resulting best practices. The "not-working" programs have not examined the context of the environment nor its stability or lack of stability. Finally, they have not analyzed, acknowledged or provided for the scope of the change involved in the program being designed and implemented. (See Figure 1, Berman, 1979.)

Using Berman's terms to summarize: they have not provided a planned process of adaptive implementation which would ensure a program that worked.

Figure 1

MATCHING IMPLEMENTATION TO SITUATION

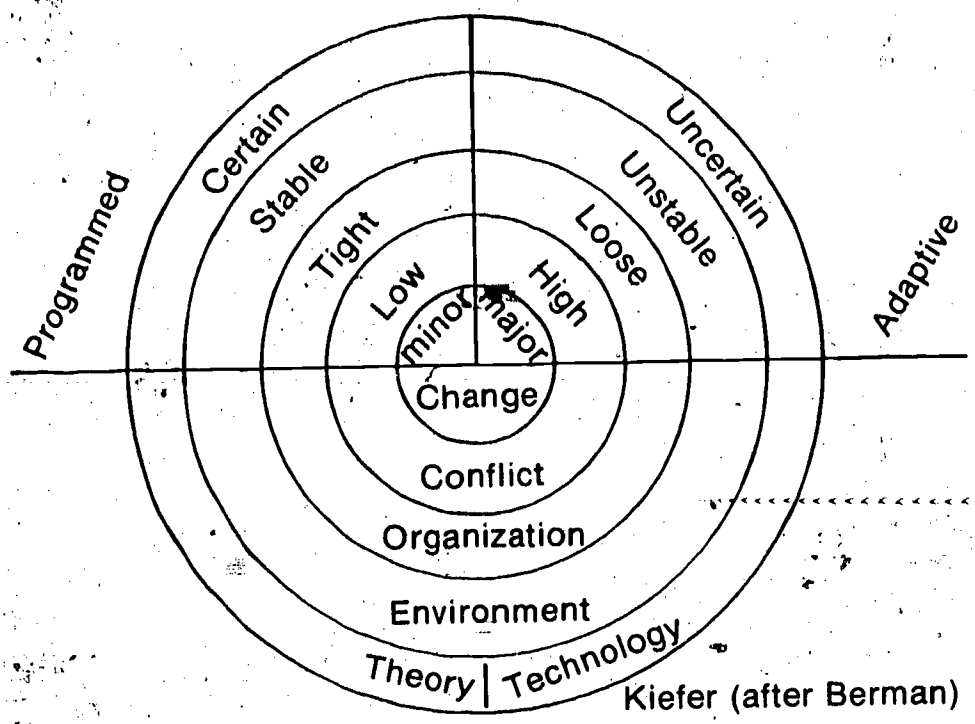
Situational Parameters	Implementation Approach Programmed	Adaptive
1. Scope of change	minor	major
2. Certainty of technology or theories	certain (within risk)	uncertain (high risk)
3. Conflict over goals or means	low	high
4. Structure of the setting	tightly (coupled)	loosely
5. Stability of the environment	stable	unstable

(Berman, 1979)

In too many cases, the programs that did not work were designed by one or two persons in isolation, formulated under time pressures, implemented in a segment of a school system or a part of a school building, given an identity that confused or irritated staff and assigned to persons unfamiliar with the total school environment and the system's environment in the community. There were no provisions made for flexible adaptation to changing theory, technology, values, opinions, resources or methods. If you change your perspective on what was just stated, you will recognize that what was intended was an efficient, carefully planned and precisely implemented program. But because it did not take into account all of the previously mentioned variables, it became an imposed, resisted, attacked invader in the school system.

Since 1968, we have seen mounting conflict centered around education for language minority children. We have seen bilingual education attacked, resisted and condemned at all levels of both education and society, even though not in all situations. In 1968, few persons were yet aware of the scope of the change being asked/demanded of the educational institutions in the United States of America. There was no agreement then, nor is there today, on the character of language acquisition or as[illegible] on the methods and instruments to be [illegible]lict is, of course, tied to many deep cult[illegible]ue differences and to the powe[illegible] issu[illegible] the allocation of educational resourc[illegible]t, however, is only tangential to the pres[illegible]

Figure 2



Kiefer (after Berman)

We are only now beginning to realize how loosely school systems are coupled or tied to each other. We, therefore, do not have formulas for operating with such systems, nor do we have an understanding of all of the implications of that kind of organizational pattern. We are only now learning management strategies to deal with the instability produced by changes in the environment where schools operate.

The remainder of this paper will describe a process of design and implementation that takes the above concepts and uses them as a base for planning and organizing language assessment programs.

PLANNING FOR CHANGE

If we want ours to be one of the language assessment programs that work, then we have a complex change process ahead of us. First, we will have to acknowledge all of the variables involved. We cannot simply choose a test and tell teachers to give it. Second, we will need to collect information on the environment in which the program will be implemented and the persons in that environment. Third, we will need both courage and patience to live with uncertainty, keeping the assurance that we will become more certain. Uncertainty and ambiguity are both forces which can paralyze us into inaction and/or force us into arbitrary pronouncements that we will regret.

So: there are certain things that we can do to implement language assessment programs that work. These "things" are related to the following questions and are a part of planning change. They establish the environment and start the planning.

1. Do we know the present status of assessment in the system?
   Do we know who directs assessment and what the consensus is on assessment theory?
   What instruments are used?

2. Do we know how decisions on assessment are made?
   (A person? a committee? with or without recommendations?)

3. Is language proficiency assessment a major or a minor change in this assessment program?
   Is there a form of language assessment being used for another program? (Speech and Language Program? Special Education? Reading? Foreign Language?)

4. How can we relate to and make use of the district program in our own planning?

5. Who is involved in language teaching/assessment for other programs?

6. What theories of language learning are generally accepted?

7. How familiar are staff members with these theories? With kinds of tests? With aspects of language that are assessed? With using test data?

8. How familiar are staff members with language and reading theories?

9. How much conflict is there among these "language people" about the theories of language?

10. What are the various segments or units of the school or system which would be involved with the language assessment procedures?

11. How are these units related and how do they communicate?

12. Is there a central committee to which new programs need to be submitted?

13. What projections have been made on students, languages, systems growth or decline, etc.? In other words, what other changes are in process or being projected?

14. What resources are available? How much money can be spent for tests? For training personnel to administer them? For scoring, grading, reporting and using the results of tests?

It is possible for one person to sit alone in an office and work out answers to these questions, then seek information on the areas which are unclear. It is also possible and much more effective for a core group to meet and analyze these questions, clarify and discover common understanding of these concepts and then proceed to the next step: formulating a description of the situation with a clear policy statement that directs program development efforts.

Why wait until a program is in operation to discover that you have a different concept of assessment than I do; that you perceive different purposes for an assessment program than I perceive; that you know more about the district assessment program than I do and that neither of us has a clear picture of the staff available to do the "assessing"? Even if you sit alone to formulate a beginning to focus the discussion, you [redacted] need to verify your perceptions and check with ou[redacted] persons on their perceptions and concepts. A progra[redacted] that works will have a core discussion to initiate the process of program design and implementation.

In the second step, this group can formulate a list of further information needed as well as sources for that information. The members of the group can establish the process and time line for collecting this information and fitting it into the total design process.

It should be obvious that one necessity is time: time for talking and clarifying, time for research, time for analyzing, time for deciding and time for training. One of the most effective ways to use time is to spend it at the outset to make the process clear. We may not know what we are going to do, but we really are clear on how we are going to do it.

THE PLANNING PROCESS

The key to effective implementation is a process that allows for and encourages an adaptive process of change. The first step in that process is designing the process itself. The chart on the following page (Figure 3) outlines the steps which should follow that initial discussion. Each of the steps will be discussed and then applied to a sample development of a language assessment program. (It can be mentioned here that this is a developmental process that in effect has no end. A policy may develop which is the opposite of the first one considered. Yet this new policy remains a step toward the implementation and refinement of the original program.)

Change needs to be considered as an ongoing developmental process of increasing precision and refinement in policy and program.

Policy Development

A policy statement is a formulation of a consensus of the group on a particular response to a problem or condition.

Example:

Problem: We do not know how nor when to initiate reading instruction for language minority students.

Policy: All incoming first grade language minority students will be given an assessment of oral language proficiency to determine which language provides a skills base for the reading process.

Figure 3

PROGRAM MANAGEMENT FOR CHANGE

WHAT	WHO	HOW
1. Policy Statement	Total Staff	Consensus
2. Program Design	Committee with Staff Input	Task Force
3. Implementation	Manager	Following Design
4. Evaluation	Total Staff	Following Design
5. Conclusions	Total Staff	Consensus Decision
6. Revised Policy	Total Staff	Consensus Decision

(WHEN — WHERE)

At this stage we acknowledge that we do not have the experts who can tell us with certainty that we have made a *perfect* choice of response. We cannot even be sure that we have made the *best* choice. We know that there is controversy over the relationship between oral skills and literacy skills. We know there is no certainty that we will not change the policy when we have accumulated more information on the students and more skill in assessing language and predicting reading success. We can prevent sudden and arbitrary changes of policy and program by establishing regular times for evaluation. We can be sure that we have agreed upon a course of action that seems to start us in the direction of clear and careful planning for the instruction of children. We also know that the policy will be reviewed after a stated length of time and after carefully collecting facts about what happens to the students.

The review will enable us to re-decide with more precision. This is in no way an experiment. An experiment is designed to see if something works or how it works. What we are doing is acknowledging that information, on language is exploding along with all other categories of information. It is no longer possible to know all there is to know before we make a decision. (One of the current problems is that language research was not available to today's teachers at the time they were being trained. Since 1950, and especially in the last ten years, we have profited from many persons' efforts to learn exactly what language is and how we acquire it.)

A consensus, therefore, on what course will be followed is the first step in the process of seeing how a program that will enable us to teach/learn more effectively can be developed. We can be paralyzed in our efforts if we expect to know beforehand everything that we will know after the program is implemented.

(A side note: we have to trust ourselves as professionals to discover what the questions are and to seek from the "experts" answers for what *we* know are *our* questions.)

Policy determination is best done after clarifying all the concepts and coming to a consensus decision of the persons involved in the program.

Design

This process, however, seldom works for the next task. The discussion and consensus-building, decision-making process flounders when a group tries to design the implementation and evaluation steps to see that a policy becomes operative. To have a task force of about five persons receive directions from the group (A list of "watch out for," "please include," "don't," etc.) before beginning the plan will direct their efforts.

Back to the sample: a staff might give the following tasks to the group:

Do—Plan for multiple language groups, not just Spanish.

Do—Provide both pre- and post-tests.

Don't—Use norm-referenced tests.

Do—Provide inservice to all staff.

Do—Check with the Director of Pupil Personnel Services for a representative to review the design.

Do—Be finished by July 1.

etc.

This way the task force hears ahead of time what the group will be looking for in the final review. (Few groups seem to be able to identify all these points ahead of time, but there will usually be a sufficient number of "caveats" to provide direction for the task force.) Projects that "don't work" find that the long hours of designing a program can be wasted, and the staff frustrated, if their project is severely changed by the committee review. All members know that they need to be flexible and respond to criticisms and suggestions, but the "do's and don't's" can focus their time and efforts to prevent some of the frustrations.

The sample again: A design has all the information: Who is going to do what to whom, when, where and with what. It will list the personnel, the resources and materials, the process, the persons involved, the dates, the person or persons responsible, the locations, and the questions which will direct the collection of information for the evaluation. It will, most of all, be aimed at providing a process to solve the original problem: Will this design for implementation and evaluation provide us with a way to decide on the time and the language in which to initiate reading instruction for language minority children?

DATE	TIME	GROUP	STAFF RESPONSIBLE
July 30	10:00 a.m.	Inservice Staff	—Program Manager
Aug. 15	10:00 a.m.	Inservice Testers	—Program Manager
Sept. 2	2:00 p.m.	Awareness Session	—Manager & Staff —All Secondary Staff
Sept. 10	9:00 a.m.	Begin speech and language screening.	
Sept. 20		Evaluation report on screening: how many screened, how many identified as needing testing, etc.	

The complete time line/action plan will lead to a formulation of all the resources, both personnel and materials, that will be needed to implement and evaluate the program. The design, therefore, needs to anticipate problems in all aspects of implementation and evaluation. When the design is completed to the satisfaction of the task force, following the specifications of the staff, it goes back to that staff for finalizing, then to the program manager to begin to make it happen.

Implementation and Evaluation

As a transition to step three, implementation, one manager told me that he tried to use the exact language of the design and policy for training, for awareness sessions, and even for the necessary reports. Thus, feedback could be coded to certain sections of the design and used for both formative and summative evaluation of the program.

Remembering the sample again: We aim the implementation at the original problem; decisions need to be made on initial reading instruction. Our program of evaluation, our choice of test instruments, our training of the testers, our analysis of the test results: all should provide us with facts that can clarify the decision on reading. Implementation (operationalizing and institutionalizing the design) also operationalizes the evaluation process. For, even though implementation is listed as step three and evaluation as step four, they are concurrent processes. In addition, they are both now a matter of taking the design and seeing that it happens.

Our evaluation will begin with the question: Do we now have sufficient data for decisions? If the answer is yes, continue and refine the process. If no, why not? What have we learned about the instruments, the aides giving tests, the reading series, the opinions, values and attitudes of the personnel that will enable us to refine the original policy or the design? The sample again: we might find that:

- Those monolingual in another language may not have a sufficient language base for reading in either language.
- The tests we have chosen do not provide what we need.
- Some native speakers of English do not have a sufficient language base for reading.
- We did not sufficiently train those who administered the tests.

In other words, we will be able to pinpoint the reasons why we still need more information on students to adapt the design or the policy. We now have an alternative to scrapping the procedure and either returning to the former arbitrary decisions or floundering with random attempts, to discover "what went wrong."

We may discover that the policy is correct and implementation ineffective or that the implementation went well but the evaluation information was not systematically collected, recorded and analyzed. In short, we should get enough information in a useable format to determine how to adjust to the results of the evaluation.

Summary

To "make programs work," we need:

- to anticipate,
- to plan,
- to investigate,
- to evaluate,

and to be flexible in all the processes without being arbitrary and/or casual in our decisions. We should have a program that works in our own terms: we know why we are doing what we are doing; how we are doing it; who did or did not do what; and the opinions and reactions of those involved.

We will be able to describe:

- the scope of the change we projected and the change we accomplished,
- the amount of congruence between goals and means,
- the organizational context or setting in which the program exists,
- the stability of the setting,
- the certainty or uncertainty of the theories and technology involved.

This collection of papers certainly adds to your general understanding of the state of those theories, as uncertain to say the least and conflicting to say a little more. Perhaps knowing *how* we will work can provide a compensation for the uncertainty involved in the exact content of *what* we need to do. Perhaps a definite, but flexible, change process can direct our efforts to determine and refine what needs doing.

APPENDIX

Appended to this article are two drafts of the results of the process of planning. Patricia Chamberlain of the Illinois Resource Center in Arlington Heights shared what is being developed by a staff in a local district. The charts are examples of the increasing precision that can be obtained in describing exactly what will be done and how. The samples provide a demonstration of how the group understanding grows and how the myriads of details involved in a program begin to organize themselves in a way that supports and encourages implementation.

There may be a third draft and a fourth draft as well, as the knowledge that, subject to both formative and summative evaluation, the "final form" need never and may never be developed.

APPENDIX
DRAFT #1
July, 1982

ENTRY/EXIT PROCEDURE

POLICY	DESIGN	IMPLEMENTATION	EVALUATION
Students will be identified by the Home Language Committee.	All entering students will complete the Home Language Survey form.	Students that answer a language other than English to questions 1, 2 or 3 will be screened further via testing and observation. — All pre-school achievement data will be reported to the receiving teacher.	
Children should be given an English language proficiency test.	Children will be given the Language Assessment Scales in English (1-8). — Children (4-6) will be given either the LAS or a language-sample rating system.	Students must achieve a level 4 proficiency to exit the bilingual program.	
Children should be given an achievement test in English.	— Children will take the LABS. — Children will take the SAT (where possible).	Students must achieve the following on English achievement tests to exit the bilingual program: 70% or above on LABS or 2/3 at or above district norms on the SAT.	A follow-up evaluation will be conducted twice a year as part of teacher/ principal meetings.
Placement decisions will be made by a building committee.	The Committee will consist of the bilingual teacher, principal, the receiving teacher, parents and other informed participants.	Placement may include bilingual program, all English program, ESL only, Title I, Special Education referrals, others.	
All children in bilingual programs should be tested in native language.	Children (1-8) will be tested using the Language Assessment Scales in Spanish, if appropriate. — All children will be tested using the California Test of Basic Skills (CTBS) Spanish. — Children will not be tested using the Boehm test of basic concepts. — Children (1-3) will be tested using the Santillana Reading tests.	Children will be tested in the spring of each academic year.	
Preschool students will be tested.			

POLICY	DESIGN	IMPLEMENTATION	EVALUATION
Lincoln, Paul Revere, Whittier, Greenwood, Greenbriar, and Kerr Schools will house bilingual programs. Pull-out: Bilingual teachers will receive an aide if they teach more than 3 grade levels.	Lincoln will adopt an integrated model. All other schools will adopt a pull-out model. A job description for bilingual teachers will be written.	(For pull-out programs) Schedules will be developed cooperatively by the bilingual teacher, classroom teachers, the building principal and the administrator of bilingual program. Teachers in both the integrated and the pull-out models will have planning time as designated by the Professional Negotiations Agreement. Principal may provide up to 60 minutes per week release time for pull-out teachers for testing, conferencing and translating.	
Level 1 & 2 bilingual students (1-8) will receive 150 minutes per day of instructional time in the bilingual program.	This includes native language in reading, language arts, science, social studies, math, and ESL instruction.	The Santillana *Reading in Two Languages* Series in Spanish will be adopted for grades K-4 for reading. The Santillana *EGB Senda* Series will be adopted for grades 5-8 for reading and language arts.	
Other eligible students will receive a minimum of 90 minutes per day of instructional time in the bilingual program.	Content areas will be taught in native language based on individual needs.	The selection of supplementary materials in science and social studies will be made at the fall bilingual monthly staff meeting. —Addison-Wesley's *Math in Our World* will be adopted in Spanish for grades 1-8 for math. —Scott Foresman *I Like English* will be adopted for ESL instruction in grades K-5. —The selection of supplementary ESL materials for grades K-8 will be made during the monthly bilingual meeting. —Additional money should be made available for language/experience supplies and activities for grades K-8.	
Formal reading instruction will be given in the student's stronger language cognitive.			

POLICY	DESIGN	IMPLEMENTATION	EVALUATION
New bilingual staff will be trained to work with the limited-English-proficient students.	Training will be provided in the following areas: bilingual/bicultural education, district and state policies and procedures, and district bilingual curriculum.	Training will take place two days prior to the opening of the school.	
The bilingual staff will meet once a month during the school year.	The bilingual administrator, bilingual staff and principals of bilingual buildings will be invited.		
A basic skills curriculum in Spanish language arts will be developed.		October 4 Institute Day will be designated as a work session for bilingual teachers to begin curriculum development.	
An ESL curriculum will be developed.			
Report cards will be revised.	Two sections will be added to the existing report card: Spanish language arts and ESL		

NONBILINGUAL PERSONNEL

POLICY	DESIGN	IMPLEMENTATION	EVALUATION
Nonbilingual personnel will be provided with resources and information regarding the bilingual program and student.		Two copies of Handbook for Classroom Teachers of LEP Students will be disseminated to each building with LEP students by the project administrator.	
	Materials for LEP students to use in the nonbilingual classroom will be made available to all staff members.	The selection of resource materials will be made at the fall bilingual staff meeting. The materials will be housed in the Learning Media Centers at each building.	
	Administrators and nonbilingual staff will be provided inservice on bilingual policies and procedures as soon as possible during the 82-83 school year.	Bilingual materials available in individual buildings will be combined and housed in the LMC.	

POLICY	DESIGN	IMPLEMENTATION	EVALUATION
	Lincoln, Paul Revere, Whittier, Greenwood, Greenbriar, and Kerr Schools will house bilingual programs. Lincoln will adopt an integrated model. All other schools will adopt a pull-out model. Pull-out: Bilingual teachers will receive an aide if they teach more than three grade levels. A job description for bilingual teachers will be written.	(For pull-out programs) Teachers in both the integrated and the pull-out models will have planning time as designated by the Professional Negotiations Agreement. The bilingual program administrator will monitor compliance. Schedules will be developed cooperatively by the bilingual teacher, classroom teachers, the building principal and the administrator of the bilingual program by time. Principal may provide up to 60 minutes per week release time for pull out teachers for testing, conferencing and translating.	
Formal reading instruction will be given in the student's stronger language cognitive.	Level 1 & 2 bilingual students (grades 1-8) will receive 150 minutes per day of instructional time in the bilingual program.	The Santillana *Reading in Two Language* Series in Spanish will be adopted for grades K-4 for reading. The Santillana *EGB Senda* Series will be adopted for grades 5-8 for reading and language arts.	
Students will receive first language instruction based on need in order to provide for cognitive growth.	Instruction includes native language reading, language arts, science, social studies, math, and ESL. Other eligible students will receive a minimum of 90 minutes per day of instructional time in the bilingual program.	The selection of supplementary materials in science and social studies will be made at the fall bilingual monthly staff meeting. —Addison-Wesley's *Math in Our World* will be adopted in Spanish for grades 1-8 for math. —Scott Foresman *I Like English* will be adopted for ESL instruction in grades K-5. —The selection of supplementary ESL materials for grades K-8 will be made during the monthly bilingual meeting. —Additional money should be made available for language experience supplies and activities for grades K-8.	

POLICY	DESIGN	IMPLEMENTATION	EVALUATION
New bilingual staff will be trained to work with the limited-English-proficient students.	Training will be provided in the following areas: bilingual/bicultural education, district and state policies and procedures, and district bilingual curriculum.	Training will take place two days prior to the opening of the school and be provided by *(person)*.	
The bilingual staff will meet once a month during the school year.	The bilingual administrator, bilingual staff and principals of bilingual buildings will be invited.		
A parallel bilingual curriculum will be developed to ensure a smooth transition from the bilingual program.	A basic skills curriculum in Spanish language arts will be developed.	October 4 Institute Day will be designated as a work session for bilingual teachers to begin curriculum development.	
	An ESL curriculum will be developed. (See attached sheet.)		
	Report cards will reflect the bilingual curriculum. Two sections will be added to the existing report card: Spanish language arts and ESL.	The additions will be made by *(person)* by *(time)*.	

STEPS TO DEVELOP AN ESL CURRICULUM

1. Clearly define levels:

   Example Level 1
   1. Understand very little spoken English, but can distinguish some words/phrases.
   2. Can label certain concepts, i.e., book, door, etc.

2. Assign percentage of time devoted to four skill areas by level.

   Example

	Level I Beg. — End	Level II Beg. — End	Level III Beg. — End	Level IV Beg. — End	Level V Beg. — End
Listening	90% — 50%	70% — 50%	60% — 40%	40% — 20%	30% — 15%
Speaking	10% — 50%	30% — 50%	40% — 60%	40% — 60%	
Reading	10% — 30%	15% — 35%	20% — 40%	25% — 45%	
Writing		10% — 25%	15% — 30%	20% — 40%	

Levels I and II should spend more time on listening and speaking.
Levels III and IV should spend a little more time on reading and writing.

3. Brainstorm for topics (survival) and then think of one language experience activity for each topic.

4. Brainstorm and assign topics to different levels — ensure that the topics are spiraled throughout levels (when appropriate).

   For example: Level I: Greetings and Leave Taking Language Experience = have kids line up outside of room and walk in one at a time. Each one should say hi, hello, how are you, etc.

5. Go back, brainstorm and think of appropriate language that fits the situations (topics) and functions.

POLICY	DESIGN	IMPLEMENTATION	EVALUATION
Nonbilingual personnel will be provided with *resources and information* regarding the bilingual program and students.	Resource materials will be available in every bilingual *building*.	Two copies of *Handbook for Classroom Teachers of LEP Students* will be disseminated to each building with LEP students by the project administrator by *(date)*.	
	Materials for LEP students to use in the nonbilingual classroom will be made available to all staff members.	The selection of resource materials will be made at the fall bilingual staff meeting. The materials will be housed in the Learning Media Centers at each building. *(Person)* will be responsible for implementation.	
		Bilingual materials currently available in individual buildings will be combined by *(person)* and housed in the LMC by *(time)*.	
	Administrators and nonbilingual staff will be provided inservice on bilingual policies and procedures as soon as possible during the 82-83 school year.		
A committee of staff representatives will be formed to review the bilingual program policies, procedures, design and implementation.	The committee should have representatives from the bilingual program, the nonbilingual staff, the administration and the preschool program.	The committee will meet one week in June. It will reconvene in November and May. A bilingual program manual will be developed by September of 1982 by *(people)*.	
Parents are critical to students' school success.	The Parent Advisory Council will meet publicly a minimum of three times per year.	District personnel as well as parents should be informed of meeting dates.	
Bilingual records will be accessible to authorized personnel.	Bilingual Census data will be housed in each bilingual building.	The bilingual teachers will keep the information. Data will be collected *(time)* by *(person)*.	
	A bilingual student information sheet will be maintained in the temporary cumulative record file for each student considered for or placed in the bilingual program. (See attached sheet.)	Each teacher will maintain up-to-date records on the students in his/her class. The program administrator will monitor compliance.	
A part-time bilingual project director should be assigned/hired for supervision of the program.			
Parents should be inserviced on a regular basis.	Topics for parent education should include: Bilingual Education in Two Languages, The District Bilingual Program, and How to Help Your Children at Home.	Meetings should be held quarterly.	

SCHOOL DISTRICT #130

Bilingual Student Information
(to be filed in student's temporary record)

Name _____ Birthdate _____ Referral date _____

Address _____ Phone _____

Language Survey Parent _____ Date _____

Teacher _____ Date _____

Kindergarten Assessment/Evaluation Information _____

YEAR _____

GRADE 1 2 3 4 5 6 7 8

SAT Total Reading/GE

SAT Total Math/GE

SAT Total Auditory/GE

LABS (% C/M) or
LABS Survey Test %

English Language Proficiency Test

Test _____ Proficiency Level _____ Date _____

Test _____ Proficiency Level _____ Date _____

Test _____ Proficiency Level _____ Date _____

Native Language Academic Testing:

Date

Pretest 1 2 3 4 5 6 7 8

Reading — Test

GE/
Score

Math — Test

GE/
Score

Assessment Committee Conferences

Date

In Attendance

Recommendations:

Date

In Attendance

Recommendations:

Date

In Attendance

Recommendations:

ENTRY/EXIT PROCEDURE

POLICY	DESIGN	IMPLEMENTATION	EVALUATION
Student's home language will be identified upon entry.	All students will complete the Home Language Survey (HLS) form upon entrance.	HLS will be reviewed by *(person)*. The *(person)* will assure the completion of the form by *(date)*. Students that answer a language other than English to questions 1, 2, or 3 will be screened further via testing and observation within *(time)*.	
Students should be given tests in English to determine their level of functioning in an all-English program.	Students should be given an English language proficiency test. Students will be given the Language Assessment Scales in English (grades 1-8). Students (ages 4-6) will be given either the LAS or a language sample rating system.	Students must achieve a level 4 proficiency to exit the bilingual program. Level 4 proficiency alone is not sufficient for program exit. Students will be tested by *(person)* within *(time)*.	
	Students should be given an achievement test in English. Students will take the LABS. Students will take the SAT, where possible.	The tests will be administered by *(person)* on *(date)*. Students must achieve the following on English achievement tests to exit the bilingual program: 70% or above on LABS or 2/3 at or above district norms on the SAT. Test data will be reviewed by the committee within *(time)*.	A follow-up evaluation will be conducted twice a year as part of teacher/principal meetings. Is the student functioning at or above grade level in English?
A committee should be formed to review all placements to ensure educational appropriateness.	Placement decisions will be made by a building committee.	The committee will consist of the bilingual teacher, the receiving teacher, principal, parents and other informed participants. Placement may include bilingual program, all-English program, ESL only, Title I, Special Education referrals, others. The committee will meet *(when)*.	
All students in bilingual programs should be tested in native language in order to have diagnostic information regarding conceptual development.	Students grades (1-8) will be tested using the Language Assessment Scales in Spanish, if appropriate. All students will be tested using the California Test of Basic Skills (CTBS) in Spanish. Students will not be tested using the Boehm Test of Basic Concepts. Students (grades 1-3) will be tested using the Santillana Reading tests.	Bilingual teachers will test program students in the spring of each academic school year. New students will be tested by *(person)* within *(time)* of entry.	

POLICY	DESIGN	IMPLEMENTATION	EVALUATION
	- Preschool students will be tested using the *(test)* *(when)*	All preschool achievement data will be reported to the receiving teacher by the sending teacher (at the end of the school year) The preschool teacher will administer the test and record the data on the student information sheet *(when)*.	

Bibliography

Beal, G.M. and Bohlen, J.M. "The Diffusion Process." *Models for Educational Change*. Edited by A.L. Bertrand and R. Von Brock. Austin, TX: Southwest Educational Development Laboratory, 1968. (ERIC ED 025 361.)

Berman, P.; Greenwood, P.W.; McLaughlin, M.W.; and Pincus, J. *Federal Programs Supporting Educational Change*, Vol. IV, (Abr.): A Summary of the Findings in Review. Santa Monica, CA: Rand Corporation, 1975.

Berman, P. "Thinking about Implementation Design: Matching Strategies to Situations." unpublished manuscript, June, 1979. (to be published in Mann, Dean and Ingram, Helen, eds., *Why Policies Succeed and Fail*.)

Bogdan, R. and Taylor, S.J., *Introduction to Qualitative Research Methods*. New York: John Wiley and Sons, 1975.

Corwin, R.G. *Reform and Organizational Survival: The Teacher Corps as an Instrument of Educational Change*. New York: John Wiley and Sons, 1973.

Dalin, P. *Case Studies of Educational Innovation*. Paris: Organization for Economic Cooperation and Development, 1973.

Danoff, M. et al. *Evaluation of the Impact of ESEA Title VII Spanish-English Bilingual Education Projects*. Palo Alto, CA: American Institute of Research, 1977.

Eisner, E.W. *The Educational Imagination: On the Design and Evaluation of School Programs*. New York: Macmillan and Company, 1979.

Gaddis, M.T. *Organizational and Personal Constraints on the Successful Institutionalization of Individually Guided Education*. Technical Report No. 447. Research and Development Center for Cognitive Learning. Madison, WI: University of Wisconsin, 1978. (ERIC ED 151 932)

Glassman, R.B. "Persistence and Loose Coupling in Living Systems." *Behavorial Science*, 18 (1973): 83-98.

Hall, G.E.; Loucks, S.F.; Rutherford, W.; and Newlone, B.W. "Levels of Uses of the Innovation: A Framework for Analyzing Innovation Adoption." *Journal of Teacher Education*, 26, 1 (1975): 52-6. (ERIC EJ 115 168)

Hall, G.E. and Loucks, S.F. "A Developmental Model for Determining Whether the Treatment Is Actually Implemented." *American Educational Research Journal*, 14, 3 (1977): 263-76. (ERIC EJ 180 513)

Havelock, R.G. *The Change Agent's Guide to Innovation in Education*. Englewood Cliffs, NJ: Education Technology Publications, 1973.

_____. *A Linkage Perspective on Educational Change*. Ithaca, NY: Cornell University, Institute for Occupational Education, 1978.

Herriot, R.E. and Gross, N., eds. *The Dynamics of Planned Educational Change*. Berkeley, CA: McCutchan, 1979.

Mann, D. "The Politics of Training Teachers in Schools." *Teachers College Record*, 77, 3 (1976): 336-37. (ERIC EJ 135 254)

Miles, M.B., ed. *Innovations in Education*. New York: Teachers College, Columbia University, 1964.

_____. "Thinking About How to Do It: Alternative Models of Planning and Implementation of New Schools." Paper presented at the Annual Meeting of the American Educational Research Association, 19-23 April, 1976. San Francisco. (ERIC ED 120 898)

Pankratz, R. and Martray, C. "Making Planned Change: A Combination of Staff Development and Political Support." Paper presented at the Annual Meeting of the Association of Teacher Educators, 16-18 February, 1981. Dallas, TX. Mimeographed. Bowling Green: Western Kentucky University, College of Education.

Pankratz, R.; Tanner, J.; Leeke, W.; Moore, B. *Planning for Institutionalization: The Continuation of New Programs and Practices.* Omaha, NB: Teacher Corps Developmental Training Activities, Center for Urban Education, 1980.

Robert, J.M.E. "Implementation of Innovations in Educational Organization and Instruction." Working Paper. Philadelphia, PA: Research for Better Schools, June 1978. (ERIC ED 180 072.)

Runkel, P.J.; Schmuck, R.A.; Arends, J.H.; and Francisco, R.P. *Transforming the School's Capacity for Problem Solving.* Eugene, OR: Center for Educational Policy and Management, University of Oregon, 1978. (ERIC ED 168 206.)

Taylor, B.L.; Sullivan, E.W.; and Dollar, B. *Mapping Teacher Corps Projects: A Planning Resource Book.* New York: Center for Policy Research, 1978.

Weick, K.E. "Educational Organizations as Loosely Coupled Systems." *Administrative Science Quarterly,* 21, (March 1976) 1.

_____. "Administering Education in Loosely Coupled Schools." *Phi Delta Kappan,* 63, (June 1982) 10.

Zaltman, G.; Florio, D.; and Sikorski, L.A. *Dynamic Educational Change.* New York: The Free Press (Macmillan), 1977.

NUMBERS AND NEEDS: ESTIMATING BILINGUAL EDUCATION POPULATIONS

Dorothy Waggoner, Ph.D.
Waggoner and Associates
Washington, D.C.

Introduction

A variety of estimates of the numbers of language minority children who need or might benefit from bilingual education programs has been put forward in the years since the first federal bilingual education legislation was passed in 1968. These estimates have varying sources and varying degrees of reliability. They describe populations defined in various ways. They respond to different concepts of the benefits of bilingual education. This paper discusses some of these estimates. It describes, in particular, the issues underlying the 3.6 million estimate resulting from the 1978 Children's English and Services Study (CESS), the minimum of 3.4 million based upon findings from the CESS applied to the 1980 Census results — both responding to the definition in the Bilingual Education Act — and the less than a million to 1.5 million estimate of the Department of Education's Office of Planning, Budget and Evaluation (OPBE), originally developed in connection with the Department's Proposed Regulations for compliance with national origin nondiscrimination — *Lau* compliance. It sets forth the implications of the published estimates on language minority children from the 1980 Census and the relationship of those estimates to the limited-English-proficient children defined in the Bilingual Education Act. It affirms the need to support the present definition of the target group in the Bilingual Education Act so that all limited-English-proficient children may continue to be served and so that the advantages of integrated programs and of bilingualism for minority and majority children in the wider society may be realized.

> There were in 1960 about five million persons of school age (6-18) in the United States who had a non-English mother tongue. It is reliably estimated that over three million of this group did in fact retain the use of that tongue . . . The situation is not known to have changed notably since 1960. These are the children we are concerned with; plus another million or so in the same category under six years of age and soon to enter the schools. They are necessarily and unavoidably bilingual children.
>
> Bruce Gaarder, U.S. Office of Education, testifying before Senator Yarborough's Special Subcommittee on Bilingual Education, May 18, 1967 (Andersson and Boyer 1970:I, 49-50)

According to the 1970 Census, about five million youngsters in this country come from homes in which the language generally spoken is something other than English. A number of these youngsters speak English, of course, but estimates based on recent samplings in several States are that between 1.8 and 2.5 million others need bilingual education.

> Article in *American Education*, July 1974 (U.S. Office of Education, DHEW)

An estimated 2.4 million children with limited-English language proficiency aged 5 to 14 were living in the United States in the spring of 1978. This number represents 63 percent of all children aged 5 to 14 years living in households where a language other than English was spoken. In addition, there were estimated to be as many as 1.2 million children of limited-English proficiency who were younger or older than the study group but who were also of school age.

> Josue González announcing the results of the Children's English and Services Study at the 1979 NABE conference (NABE News, June 1979:1)

In these proposed rules, the Education Department addresses one of the most serious barriers to equal opportunity in education. It proposes standards for teaching students whose primary language is not English and who have limited proficiency in English. The number of these students is large and growing. Estimates by the National Institute of Education and the National Center for Education Statistics place the number of limited-English-proficient school-age children at over three and a half million.

> Proposed *Lau* Regulations, U.S. Department of Education, August 5, 1980 (*Federal Register*: 52052-53)

We believe that the number of children whose opportunities to benefit from education are curtailed by dependence on a non-English language is almost certainly not more than 1.5 million and possibly less than 1 million.

> Robert E. Barnes, Office of Planning, Budget, and Evaluation, Department of Education, in the draft final report "Size of the Eligible Language Minority Population" (September 25, 1981)

In 1980, one out of every ten persons (5 years old and over) reported speaking a language other than English at home . . . Well over one-half million school-age children . . . who speak a foreign language at home characterized their ability to speak English as "not well" or "not at all."

Bureau of the Census press release on the provisional estimates from the 1980 Census Sample Census 1982)

The Reagan Administration already proposes to use the low estimate (1.4 million) of the number of students in need of services under the Bilingual Education Act. Education Department officials argue that students whose proficiency is limited in both their native language and English would not be effectively served by native language instruction

Article in *Education Week*, June 9, 1982 (Foster and Matzke 1982 6)

Back of these differences in estimates are a number of issues and implications. There is the issue of the definition of the group whose size is being estimated. There is the issue of who needs special programs, such as bilingual education, and who can benefit from them. There are sociopolitical implications in the size of the estimates. The smaller the size of the group which must be served, is "eligible" for services, or can benefit from them, the lower the potential cost to the Federal Government, the states or the school districts. The smaller the number, the more difficult it will be for concerned minorities to obtain a hearing. The smaller the number, the easier for opponents to dismiss the need for special programs and to comfort themselves that it will eventually go away.

The issue of the definition of the group whose size is being estimated developed in connection with the issuance of the Proposed *Lau* Regulations by the Department of Education on August 5, 1980. The original Bilingual Education Act — Title VII of the Elementary and Secondary Education Act of 1965 as amended in 1968 — defined the target group as children of limited-English-speaking ability "who come from environments where the dominant language is other than English" (Andersson and Boyer, 1970 II, 1). The 1974 and 1978 amendments to the Act changed, and refined the definition of the target group and modified the terminology. These amendments recognized that not all language-minority children have difficulty with English — an implication in the original definition. Both specified that "the language normally used by the parents of the child" should be considered the native language of the child. Neither placed any restriction on inclusion in the target group of the limited-English speaking (1974) of limited-English proficient (1978) on the basis of the child's language usage. The 1978 amendments further specified that, in the case of Native American children, the Indian language need only to have had a "significant impact on their level of English language proficiency" (Bilingual Education Act, Sec. 703(a) (1) and (2)). It did not even have to be dominant in their current environment.

The estimate of 2.4 million children, aged 5 to 14, or 3.6 million total of school age, represents the group which meets the Bilingual Education Act definition as presently enacted. This number comes from the 1978 Children's English and Services Study. It was produced in response to the Congressional mandate to count limited English proficient children and adults in the Bilingual Education Act as amended in 1974 and 1978. The CESS estimates were the basis for the Prospective Study transmitted to the Congress last year as a partial response to the mandate. These estimates were and still are, the only available estimates based upon a national sample of language minority children who were tested in their homes with an objective measure of school-related English skills especially developed for this purpose. These estimates will be superseded later this year by the findings from a new study which will be discussed at the conclusion of this article.

The Proposed *Lau* Regulations which were issued by the Department of Education on August 5, 1980, used the 3.6 million estimate from the CESS. However, they employed a complicated set of definitions of the children for whom school districts would be expected to provide services to comply with national origin nondiscrimination as upheld by the Supreme Court in *Lau v Nichols*. In the Proposed Rules, children with a primary language other than English, who were limited-English proficient (LEP) were to be classified according to their relative proficiency in English and their primary language. Depending upon their relative proficiency, they were to be classified as English-superior, comparably limited, or primary language-superior. The primary language was defined as the first language learned or the language normally used by the student. Different treatments were to be based upon these classifications, including right of equal access to compensatory programs for LEP students classified as "English-superior" (Proposed *Lau* Regulations, 1980 pars. 100.33 and 100.39). The CESS findings did not address these classifications except indirectly. The Bilingual Education Act does not contain them.

The Regulatory Analysis Review Group (RARG) of the Council on Wage and Price Stability of the Executive Office of the President has the authority under legislation passed in 1974 to intervene in governmental rulemaking. It did so in the case of the Proposed *Lau* Regulations shortly after they were issued. It explained that it took this action because these were "the first, important ones (regulations) to be issued by the Department (of Education) since it was established" (U.S. Council 1980). To meet one of the requirements of the RARG, personnel in the Department of Education's Office of Planning, Budget, and Evaluation (OPBE) began attempting to estimate the numbers of students who would be affected by the Proposed Regulations according to the relative proficiency classifications. Although all LEP children, regardless of relative proficiency classification, were covered by the Proposed Rules, OPBE determined to reduce the numbers to those who might be classified as primary language-superior. The estimate of less than a million to no more than 1.5 million children comes from the OPBE draft final report, "Size of the Eligible Language Minority Population." This report was issued with a disclaimer that it "does not represent the official position of the U.S. Department of Education." It was issued a year after the RARG's review of the Proposed

Regulations and six months after the inauguration of the Reagan Administration and the withdrawal of the regulations which triggered the effort. This latter point is not unimportant. Having embarked upon revising the results of the Department's effort to meet the Bilingual Education Act mandate, although for *Lau* compliance purposes; OPBE appeared to be reluctant to discontinue the effort even when the reason for it had disappeared. The estimate was, accordingly, proposed to be an estimate of the Bilingual Education Act target group. The latter was redefined to consist only of those considered to be "eligible" for bilingual services or "entitled" to receive them by virtue of "dependence on a non-English Language" (Barnes 1981). The Bilingual Education Act program is not an entitlement program. Instead, school districts apply for grants to serve LEP children they believe will benefit from them. The Department's General Counsel disallowed OPBE's limitation of the target group to those in some sense dependent on a non-English language. The GC pointed out that dependence on a language other than English is not a criterion for limited-English proficiency as defined in the Act. The GC's opinion led, in turn, to the Administration's proposed amendment to the Act to give a funding priority to projects which would serve LEP children whose usual language is other than English.

Back of this determination to restrict participation in bilingual education programs with resulting reduction in the numbers is a pedagogical issue concerning the benefits to be derived from these programs. Consider these quotations:

> Since these limited-English proficient students' stronger language is English (the English-superior group), it follows that they would suffer less of an academic handicap if the academic curriculum were taught in English, rather than in their home language.
>
> . . . We would recommend that equally limited bilinguals who speak *English* at home would best be instructed in *English*, while those who speak another language at home might benefit more from instruction through that language. (Dulay and Burt, 1980:16, 19).
>
> It is our contention that a child who is not dependent on a language other than English, regardless of the child's limitations in English, should not require and would not benefit from services provided in the non-English language. (Barnes, 1981:17).
>
> The research evidence . . . strongly suggests that programs that aim to develop a high level of proficiency in two languages provide greater potential for academic development for *all* children than education through the medium of only one language . . . Research has failed to identify any category of student for whom a bilingual education would be less suitable than a monolingual education . . . (There is) no educational support, either empirical or theoretical, (for) limiting access to bilingual education by Limited-English-Proficient (LEP) students (on the basis of "English-superiority" or use of English at home.) (Cummins, 1981:42)

Bilingual education in the United States, as in many countries in the world, was originally justified by citing the 1953 UNESCO axiom that the best medium for teaching children is their mother tongue. Reinterpreted as meaning that children should be taught in their stronger language, this theory leads directly to the notions of "English-superior," "primary-language-superior," etc., with different treatments for each, embodied in the Proposed *Lau* Rules. It is exemplified in the quotations from Dulay and Burt and from the OPBE report cited above. It is not, however, embodied in the Bilingual Education Act definition which, as indicated, emphasizes the limited-English proficiency of language-minority children as the crucial characteristic. As presently enacted, the legislation enables school districts to apply for funds to develop and implement bilingual programs for all LEP children whom they believe will benefit from them. As presently enacted, the legislation permits school districts to apply for funds to develop and implement programs in which as many as forty percent of the participating children are English-speaking majority children. If, as affirmed at the 1983 National Association for Bilingual Education conference, bilingualism is in the national interest, then these aspects of the present Bilingual Education Act must be defended. All LEP children who can benefit from bilingual programs must continue to be able to participate in them. Majority children must continue to be able to participate up to the limit of forty percent. Only in this way will the advantages which follow from integrated programs for minority and majority children and for the wider society be realized.

The estimates from the 1980 Census published by the Bureau of the Census have the most serious sociopolitical implications because of the danger of misinterpretation and because Census data have a way of becoming the final word. The national number, about 650,000, and equivalent estimates for states and smaller geographic units represent children, aged 5 to 17, *reported* to speak languages other than English at home and not to speak English well or not to speak it at all. The 650,000 is a provisional estimate from a sample of the 1980 Census sample which was published last spring. The estimate from the whole sample will eventually be published. Final estimates from the entire sample are already being published for the states. In the CESS, there was a considerable difference between the ratings of children's English-speaking skills by their parents or guardians and objective testing of the school-related English skills, including reading and writing, of those same children. Applying the limited-English-proficiency (LEP) rates from the CESS to the total of 4.5 million school-age children who were reported in the Census to speak non-English languages at home yields about 2.9 million children with limited-English proficiency as defined in the Bilingual Education Act. However, these are *only* the LEP children reported to speak non-English languages at home.

Language minority children who speak English at home were not included. Information on their English-speaking skills was not gathered in the Census. Again applying rates from the CESS, it is estimated that there are at least 544,000 other language-minority children who only speak English, but who lack sufficient proficiency in the school-related English skills to succeed without special help in the English-medium school system. Thus, a minimum of 314 million children in the United States in 1980 meet the definition in the Bilingual Education Act. A minimum of 3.4 million children have insufficient English-language skills to succeed without assistance in the English-medium school. Moreover, because the precise interpretation of the language question in the Census is presently unknown, it is possible that there are considerably more school-age LEP children. The 3.4 million is a difference of 2.8 million from the Census's published estimate of those reported by their households to speak languages other than English and to lack English-speaking skills. It is a difference in the proportion of school-age children in the United States in 1980 of nearly 6 percent. Those reported to speak languages other than English at home and to lack English-speaking skills constituted 1.4 percent of the total school-age population in 1980. Those estimated to be LEP constituted a minimum of 7.2 percent of the total school-age population. The differences are much more dramatic in states and localities with high concentrations of language-minority children.

As stated earlier in this paper, the CESS estimates are the only ones now available. Similarly, the rates which were used to extrapolate the estimates of LEP children from the 1980 Census data are the only ones currently available. The CESS estimates and LEP rates are national ones. They do not reflect differences related to concentrations, much less differences which will be revealed when the data are analyzed by language group. To remedy these problems and to obtain LEP data based on an actual sample of 1980 Census respondents, the Bureau of the Census conducted a special study last summer for the Department of Education. The test used in the CESS was administered to children and an adult test, developed for this purpose, to language minority adults in a national sample. Results of this study will be available later this year. They will make possible estimates of LEP children and adults, based on the 1980 Census, for a number of states and different language groups. They will make possible new national estimates. The CESS estimates and the extrapolations from the 1980 Census based on the CESS findings will serve as bench marks for comparison with the findings of the new study.

The estimates of the size of the population in need will play an important role in the public perception of the importance of bilingual education and other programs and services for language minorities. They will play an important role in continued public support for these programs and services in 1980's. The understanding of the potential benefits of these programs for minority children and for majority children will greatly affect the extent to which bilingual education is continued and expanded. If bilingualism is in the national interest for all the population, then the varying estimates must be understood for what they are. Narrow concepts of benefit must not be allowed to obscure the extent of need among minority children. They must not be allowed to obscure the possibilities for minority and majority children to interact in mutually rewarding language-learning experiences.

References

Andersson, Theodore and Mildred Boyer. *Bilingual Schooling in the United States*. Washington, D.C.: U.S. Government Printing Office, 1970.

Barnes, Robert E. Draft final report: "Size of the Eligible Language Minority Population." U.S. Department of Education: Office of Planning, Budget, and Evaluation, September 25, 1981.

Bilingual Education Act. August 21, 1974. P.L. 93-380, 88 Stat. 503.

Bilingual Education Act. November 1, 1978. P.L. 95-561, 92 Stat. 2270.

Cummins, James. "The Role of Primary Language Development in Promoting Educational Success for Language Minority Students." In Office of Bilingual Bicultural Education, California State Department of Education, *Schooling and Language Minority Students: A Theoretical Framework*, 3-49. Los Angeles, California: Evaluation, Dissemination and Assessment Center, 1981.

Dulay, Heidi and Marina Burt. "The Relative Proficiency of Limited English Proficient Students." *NABE Journal*, IV.3, 1980, pp. 1-23.

Foster, Susan G. and Martha Matzke. "U.S. Study Reports Shortage of Qualified Bilingual-Ed. Teachers." *Education Week*, June 9, 1982, p. 6.

González, Josué. "Josué González Opens NABE Convention." *NABE News* II, 5 (June 1979) p. 1.

Proposed *Lau* Regulations, 34 CFR Part 100, U.S. Department of Education. August 5, 1980. *Federal Register*, 45, 152:52052-76.

U.S. Bureau of the Census. March 1982. Press release, "Population Highlights from the Provisional Sample."

U.S. Council on Wage and Price Stability, Executive Office of the President. October 21, 1980. Press release.

U.S. Office of Education, DHEW. July 1974. "DHEW Expenditures on Bilingual Education." *American Education.* In Center for Law and Education, *Bilingual-Bicultural Education: A Handbook for Attorneys and Community Workers,* 245. Cambridge, Massachusetts, December 1975.

207

ACKNOWLEDGEMENT

The Illinois State Board of Education extends its appreciation to Stanley S. Seidner, Ph.D. who contributed his services as editor of this publication.