DOCUMENT RESUME

ED 221 023 FL 013 112

AUTHOR Seidner, Stanley S., Ed.
TITLE Issues of Language Assessment: Foundations and Research. Proceedings of the Annual Language Assessment Institute (1st, Evanston, Illinois, June 17-20, 1981).
INSTITUTION Illinois State Board of Education, Springfield.
SPONS AGENCY Department of Education, Washington, DC.
PUB DATE 82
NOTE 138p.; Some faint type. For "Psycholinguistic Foundations of Language Assessment," see ED 206 165. For "Improving Cognition: A Multicultural Approach," see ED 209 078. For related documents, see FL 013 113-122.

EDRS PRICE MF01/PC06 Plus Postage.
DESCRIPTORS *Bilingual Education; Communicative Competence (Languages); Educational Diagnosis; *Language Proficiency; *Language Research; Language Tests; Limited English Speaking; *Second Language Instruction; Sociolinguistics; *Testing

ABSTRACT

The following papers are included in this volume dealing with issues and research in language assessment: (1) "Sociolinguistic Foundations of Language Assessment," by J. L. Ornstein-Galicia; (2) "Language Proficiency Assessment: Research Findings and Their Application," by C. Rivera and C. Simich; (3) "The Role of Grammar in a Communicative Approach to Second Language Teaching and Testing," by M. Swain and M. Canale; (4) "Dilemmas in Diagnosis," by R. L. Thorndike; (5) "Language Proficiency Assessment: Issues and Definitions," by R. E. Baecher; (6) "Foreign Language and Bilingual Assessment: Issues Approaches," by P. Woodford; (7) "Integrating Language Assessment with Teaching Performance in Subject Areas," by G. D. Keller; (8) "On Assessing the Oral Language Ability of Limited-English Proficient Students: The Linguistic Bases of the Noncomparability of Different Language Proficiency Assessment Measures," by B. Wald; (9) "Language Assessment Practices and Policies at Colleges and Universities," by S. Seidner; and (10) "Data Collection and Language Assessment Policy," by S. Duron. (AMH)

ISSUES OF LANGUAGE ASSESSMENT:
FOUNDATIONS AND RESEARCH

Edited by

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

Proceedings of the First Annual Language Assessment Institute

June 17-20, 1981

Conducted at
National College of Education
Evanston, Illinois

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

INTRODUCTION

If we define a field in terms of scholarly endeavor, language assessment is a fairly recent development. Concepts of measurement remained in isolation from the study of language acquisition through the nineteenth century. Thomas Simpson's innovations of observational error during the eighteenth century, for instance, remained confined to the study of mathematics (only a few of his contemporaries, particularly the Marquis de La Place, and the Comte de La Grange ventured to apply these mathematical concepts to astronomy). So fared the probability and antecedental formulations to the "normal curve", as well, by another of Simpson's contemporaries, the Huguenot mathematician, Abraham De Moivre. Within one hundred years, Carl Friedrich Gauss would proclaim mathematics, "the Queen of Sciences," and introduce the application of the "normal curve" to other disciplines. The chasm between measurement and language appeared to have been inadverdently and tenuously bridged through the Wundtian school, as well as through the work of the eugenicist, Sir Francis Galton. Wilhelm Wundt's "relativity theory" formed a framework for his devoted followers to explain the psycholinguistic relationships between "Gliederung" and "Gesamtvorsfellung." It was Galton in his experiments, who had attempted to apply the theories of his cousin, Charles Darwin, in order to prove the inheritability of intelligence. Although ensconced in psychology, the important linkages between language and measurement seem to have become apparent, albeit grudgingly, to such modern day inheritors of "Galtonian" tradition as Arthur Jensen.

Perhaps researchers will declare, within forthcoming decades, that the emergence of language assessment as an independent entity resulted in large part from developments in Civil Rights movements and a quest for social justice. Some modicum of evidence can be brought forth in support of this contention, when one views the historical rise of bilingual education and the plight of limited-English proficient populations in the United States. New terms would appear in educational vocabulary, such as language "dominance" and "proficiency," indicating additional frontiers for exploration and pardoxes for resolution. We can exemplify this last point by suggesting Roger Shuy's innovative concept of the "iceberg effect." Introduced in 1976, the concept posed the dilemma of reconciling statistical precision in quantification with true and complete representation of observable data. The closer one arrives to the tip of the "iceberg," the more statistically accurate are representations of surface structures (i.e., phonology, vocabulary, and grammar). The further away we go below the surface-line toward deep structures, the more accurate we are in representing semantic meaning. On one level, the precision/representation dilemma is couched in terms of discrete-point and global measures. On another level, the problem is indicative of reconciling linguistics and psychometry in language assessment. It also suggests the cross-fertilization of language assessment by developments from a myriad of disciplines, which continue to increase respective realms of knowledge at a geometric rate. With the developments will arise new issues and paradoxes for resolution, as the alchemists of today give way to the innovators of tomorrow. For indeed, to fail to claim that one collection of ideas holds sway over one historical period would be tantamount to measuring one chronological era against the standards of another.

It was with the thought of bringing current issues to the forefront, that the First Annual Language Assessment Institute was conducted at National College of Education. This volume was conceived as the first 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 Service; 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. Papers in this volume are grouped into three succinct categories: Foundation, Assessment Approaches, Research and Policy.

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, and to Lynn Rhoades, 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 Support and Printing Section, and Bernie Neff, Assistant Manager, Printing and Graphics were 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 for her professionalism and patient, loving support and to my infant daughter, Jacqueline Elinor, for prolonged and welcome moments of quiet in completing this volume. Finally, my appreciation goes to the contributors of this volume for their excellent presentations and papers.

Editor
Dr. Stanley S. Seidner
Assistant Dean for
Research and Development
National College of Education

TABLE OF CONTENTS

Part I
Foundations

ZENO'S PARADOX AND LANGUAGE ASSESSMENT

Rudolph C. Troike
University of Illinois at Urbana-Champaign

My first experience with language assessment was over twenty years and two linguistic paradigms ago, when I joined the Georgetown University English Language Program in Turkey, and found myself involved in testing AID participants using the Foreign Service Institute interview approach, and testing Turkish students in English classes using discrete point tests. In that dimly remembered and distant past, we were all confident positivists and knew all there was to know about language and language assessment — or at least how to find out what we didn't know. But as Robert Lado once stated, "the older you get the more you find you have to live down." Or, as the Pennsylvania Dutch saying goes, "How come you're too soon old and too late smart?"

Today we have reached the point — or at least some of us have — of true transcendent wisdom. We now know that what we thought we knew in the past was wrong, or incomplete, and we have begun to understand enough about language to realize how little we now know. As we learn more and more, we recognize we know less and less. Thus we are gradually approaching the achievement of the ultimate truth, which will be to know everything about nothing and nothing about everything.

One consolation in our present state of wisdom is that at least no one will try to fault us for being prematurely overconfident. Those who remember back to the heady days of yesteryear may regret one thing which was lost along with our youthful innocence — the sublime satisfaction of having all the answers whenever someone asked.

Unfortunately, people are still asking questions and expecting answers, and unfortunately now that we have achieved wisdom, we no longer have such ready answers. The price of wisdom is the awareness of ignorance, which is not nearly as satisfying as knowing everything. Indeed our inability or hesitation to answer questions can be very frustrating, both to questioners and to language assessment specialists alike. Of course, there are those who feel that when ignorance is bliss, 'tis folly to be wise, and confidently stand ready to dispense quick answers when asked. The dilemma is that there are massive national needs calling for solution, and school administrators and teachers who desperately need answers. As specialists knowledgeable about the state of language assessment, we would be derelict in our responsibilities if we refused to respond or turned them away empty-handed.

Which brings us to Zeno's paradox, which I believe is relevant here because the Greeks were always interested in truth. I have chosen Zeno's paradox because I think it appropriately symbolizes the situation in which we find ourselves. In its simplest form, it deals with the problem of attempting to reach a goal — say, walking from the front to the back of a room. In doing so, you will progressively cover half of the distance, and then half of the remaining distance, and so on through an infinite series of successive approximations. In theory, no matter how long you go, you can never reach the final point. In reality, of course, we do in fact reach the point we set out for. There is a moral here, I believe, for language assessment, for our pursuit of understanding is infinite, and we know we can never arrive. Nevertheless, we have practical issues and problems in the real world to deal with, and we must do what we can in the interim with the tools and knowledge provided by our latest successive approximation.

I should now like to turn to the subject of *language assessment*, and consider what it is we are about and why. There is an interesting interaction between the two component elements of the name itself, which strongly affects what language assessment actually consists of. First, the conception of *language* which is held will partly determine *what* is assessed and *how* it is done, and secondly, the view of appropriate assessment procedures will affect *the choice of what is measured*, which for the purposes of assessment becomes a de facto definition of language, i.e., language is what the assessment procedure/instrument measures. Both aspects in turn are strongly affected by the *purposes of assessment*. A definition of language or a specification of assessment scope or procedures cannot be specified in absolute terms independently of *why* an assessment is to be made.

Several examples will serve as a caveat. Twenty years ago the prevailing linguistic paradigm viewed phonology as the most important aspect of language, grammar as next in importance, and vocabulary or meaning as least important. The transformational paradigm rejected this conception and placed grammar on top, only to be overturned by Fillmore's case grammar and other related models which gave primacy to meaning, with grammar next, and phonology least in importance — a complete inversion of the earlier "structuralist" view.

Sociolinguistic and anthropological (not to mention psychological) perspectives have added additional complications to the concept of language by removing it from the antiseptic decontextualized ideal realm to real life settings, where domains, repertoires, skills, roles, statuses, attitudes, settings, topics, knowledge, personality, and individual relationships interact to affect linguistic competence and performance in various situations.

The older view of language as primarily oral caused reading and writing (as well as manual signing) to be ignored for testing, a serious bias which unfortunately continues to the present. A caveat that I would raise at this time is that the bright new hope of *communicative competency testing* may, depending on our purposes,

lead us still further astray. We must be careful that we do not find ourselves in Zeno's paradox in reverse, with each new step leading us halfway again away from our goal.

As stated earlier, what we assess and how we assess it is intimately associated with the purpose of assessment. As Bernard Spolsky noted at the Language Proficiency Assessment Symposium in Warrenton, Virginia, perhaps the most efficient test on record was the single-item discrete point criterion-referenced entry-exit test mentioned in the Book of Judges in the Old Testament: when the Ephraimites sought to cross the Jordan, the Gileadites asked them to pronounce the word for fish; since there was a š/s isogloss between the two tribal varieties of Hebrew, if they said *sibboleth* instead of *shibboleth*, they were deemed to have failed the test and instantly exited from existence. Test statistics state that 42,000 Ephraimites failed the test that day.

At the other end of the spectrum, we have the recent proposal by Michael Canale of OISE for an extensive and sociolinguistically sophisticated assessment of language proficiency, essentially by making a detailed inventory of communicative competence. But here again let me raise another caveat regarding *purpose*. We need to determine if what we are measuring bears a meaningful relation to the reasons for making the measurements. For those concerned with bilingual education and the school achievement of children from non-English language backgrounds (NELB), I would like to propose the perhaps startling idea that *we may be making a serious mistake in assessing language at all as a criterion for program exit*.

Lest I be misunderstood, let me hasten to explain. When bilingual education was getting started in this country a little over a decade ago, many of us argued for its necessity because of the language barrier, an argument adopted by the Supreme Court in the *Lau vs. Nichols* decision: if the child could not understand English, he or she was obviously excluded from effective participation in the classroom. Just as obviously, acquisition of sufficient English to follow instruction seemed to remove the barrier, so that this became the touchstone for most state and federal legislation, and for implementation of the *Lau* decision itself. I have previously suggested elsewhere that his argument may have been mistaken. Especially since the work of Skutnabb-Kangas and Toukomaa on Finnish immigrants in Sweden, James Cummins and I have argued on somewhat separate grounds that English language proficiency, as ordinarily conceived, is, beyond a certain minimum level, only a small and perhaps largely irrelevant factor in the school achievement of NELB students. To take a simple example: various American Indian tribes have given up their native languages and have adopted English instead, but they have not improved their school achievement as might have been expected if language proficiency were the main factor inhibiting achievement.

The Finnish researchers, Cummins, and John Oller as well have argued in the direction that there are deeper cognitive skills which underly the use of language and which may develop through the use of language. The direction of my own argument has been that extrinsic and intrinsic social and cultural factors (including such things as the Pygmalion effect of self-fulfilling teacher expectations) may be of overriding importance vis a vis school achievement, as evidenced by the fact that Blacks in Texas and parts of California fare worse in school than do Hispanics. In this view, while the use of the student's language as a medium in a bilingual program may indeed promote the development of cognitive abilities, its chief value may be symbolic, by helping to legitimize the positive valuation of the student's culture, social group, and self. Otherwise we would be at a loss to account for the success of middle class majority students in second-language immersion programs in Canada and elsewhere, and the massive failure of minority language students in regular all-English (submersion) programs.

Oller meanwhile has proposed that language ability per se is at the root of the results of I.Q. measures, and that what such measures reflect is basically a general language proficiency factor. Muriel Saville-Troike, on the other hand, has suggested that certain types of successful language proficiency, such as in interpersonal communication skills, may for primarily social reasons have a *negative* effect on the development of proficiency in handling decontextualized academic prose, which is essential for academic achievement.

The problem is obviously complex and will require much more research to resolve. Whichever way one goes, it is clear, as Cummins has pointed out, that rather *superficial measures of English fluency* (with no attention to competence in the native language) may be completely misleading with regard to the level of language control needed for more abstract decontextualized cognitive functions. Cummins has reported on children who sound like fluent native English speakers but who are yet having severe problems in school because of their limited ability to manipulate the language at a level necessary to successfully perform academic-type tasks. He suggests that it may take up to five years to acquire that ability, an observation that fits well with results from several bilingual programs which show students reaching national norms in their fifth and sixth years.

The difficulty we find ourselves in with regard to language assessment is that our original rationale for bilingual education is coming back to haunt us. Having successfully argued for bilingual education as a means to overcome the language barrier, we not suprisingly see English proficiency defined as the sole criterion for program entry and exit, even though we now realize that this may be a third or fourth-order factor or even completely misleading. Nevertheless, the parameters for language assessment are being determined primarily by legislators and administrators, rather than by specialists in bilingual education or language assessment, and we face the danger that people will uncritically accept and respond to these parameters. The statement may sound heretical coming from a linguist, but I believe we may be making a mistake in attempting to assess language at all. The purpose of bilingual education should be *educational*, not just the development of English pro-

ficiency. Consequently, *language assessment may very well be an inappropriate basis for educational decision making for limited-English proficient students.*

Since the demands for language assessment are there, mandated by law and court decisions, what can we do? We can, as many already do, try to take other factors such as achievement scores into account, and avoid, if possible, using a language score as the sole criterion for program exit. Meanwhile, we need additional research, of course, but more importantly, we need to work to convince legislators, administrators, and others that *academic achievement*, rather than English proficiency, should be the basis for exiting from a bilingual program. Better than that, we should seek to remove the compensatory label from bilingual education and work to institutionalize it as a permanent program option which is available to all students and in which requirements for program exit would not exist.

SOCIOLINGUISTIC FOUNDATIONS OF LANGUAGE ASSESSMENT

by
Jacob L. Ornstein-Galicia
Department of Linguistics
University of Texas
El Paso, Texas

A. Prefatory Note

Does sociolinguistics have anything to offer the art of language assessment? That, quite simply, is the issue that I intend to address in this paper. It is characteristic of Americans, and indeed it is an an admirable feature, that they are ever interested in a new body of knowledge which will, hopefully, bring great breakthroughs and which will advance us somewhat further to a truly Golden Age. I wish that I could sit here today and promise all of you interested in bilingual education and in assessment, that low and behold the content of sociolinguistics is the kind of study that will solve all your assessment problems and mine. Unfortunately, paraphrasing Churchhill of World War II, I can offer you little but blood, sweat — and at the worst — eyestrain from perusing the voluminous literature that has accumulated in sociolinguistics (hence SOL) both here and abroad.

In this connection, it is heartening that the proponents of SOL have not fallen into the same error as those in succession, of Bloomfieldian structural linguistics, then Chomskian transformational grammar, then mathematical linguistics. So great is the charisma of psychology that it was likewise thought that the marriage of linguistics and the former discipline could not but lead to unprecedented breakthroughs providing answers to all the puzzles of applied and theoretical linguistics. With all due respect to the progress made in all these subfields, we now realize that there is no quick way to find solutions in the language field. I'm afraid that is also the plight of bilingual assessment and language proficiency testing in general, which does not mean that we can stop trying.

After such a pessimistic side-trip, I must return to my topic and sound a more optimistic note. Yes, I do feel that SOL has a great deal to offer, and will attempt to so demonstrate. In the forthcoming section, I will trace the development of SOL, followed by discussions of possible relationships to language assessment (hense: LA). If all that is seen in the following pages is not precious metal, hopefully it will not be fool's gold.

B. Development of Sociolinguistics (SOL)

Sociolinguistics has always existed, to the extent that those who described languages also included information on the location and characteristics of its speakers, as well as in what contexts they employed the languages of a community or a nation. In saying this, we have really furnished a simplified definition of this study, which attempts to combine both the linguistic and the social dimension. In more technical terms, we might also describe it as a study which addresses itself to language behavior and its societal concomitant.

Following World War II and America's much-broadened geopolitical position, it began to be realized that our parochial and myopic stance toward language and area studies was out-dated. These studies blossomed up to the late 1960's when the revolt against school requirements of any sort almost obliterated them. At any rate, linguistic scientists were also dissatisfied with the narrow perspective of language study and analysis, and decided to join hands with social scientists, especially sociolinguistics, with or without hyphen. Actually the term had first been used in 1946 by Haver and Eva Curie, the former a social scientist, the second a speech teacher (Curie and Curie, 1952). Scholars from linguistics and the social sciences, particularly sociology, organized two meetings, both occurring in 1964 with the aid of the Social Science Research Council, and taking place at UCLA (Bright, 1966) and at Indiana University (Lieberson, 1967). No sub-branch of linguistics or sociology has shown such phenomenal growth in the last two decades as regards both college teaching and research, and apparently today it is stronger than ever.

It is, however, erroneous to give the impression that SOL was "invented" by the two groups of American scholars convening at Los Angeles and Bloomington. Many Europeans continue to be amused at, if not irritated by, what they consider American "faddishness" and presumption, as it were, of discovering all that is modern and good in linguistic science. Indeed, it has been a general tendency of both British and European schools of linguistics that they have traditionally incorporated much more extra-linguistic information than has been the case here. By way of example, in Great Britain, Firth, under the influence of anthropologist Malinowski, accorded the cultural component a significant role in his theories. Halliday, a contemporary Briton, has been greatly concerned with the compartmentalized roles that language fulfills in society, as he developed his version of functional grammar. Basil Bernstein (1971) at the University of London, Institute of Education, originally basing himself on research carried out on middle- and lower-class 10-12-year-old schoolboys, developed his own model of SOL which, in brief, holds that the speech of working class youngsters consists mostly of juxtaposed simple sentences representing a "restricted code."

By contrast, Bernstein claims that middle-class pupils produced an "elaborate code," marked by complex sentences showing embedding, subordination, and the like. Bernstein's theory in its varied versions has exercised a powerful influence among those working on intervention programs, but there is continuing controversy over his views as "elitist."

In Europe it would be impossible to enumerate those who have embodied sociocultural concern in their approaches to language science. The Prague School before World War II was careful to regard language in a social context, a trend which present-day practitioners such as Jakobson and Garvin have continued in the United States. In France, as early as 1956, Marcel Cohen had published his *Pour une sociologie du language (For a Sociology of Language)*, but his clarion call was not heeded. Meanwhile, in the United States, there were scholars like Einar Haugen, author of *Bilingualism in the Americas* (1956), a guide considered by many unsurpassed in its conciseness, as well as the *Norwegian Language in America* (1958 and 1969). Both these works are unquestionably sociolinguisitic in scope and purpose, while the latter remains to this day one of the best, if not the best, descriptions of an ethnic dialect in America. Finally, the late Uriel Weinreich's monumental *Languages in Contact* (1953), again addressed the social context throughout its rich material not only on the Romantsch of Switzerland, in contact with German, French and Italian, but also in the other numerous contact situations described by him. Ironically, however, the works of American sociologist, Joyce Hertzler, including both his article "Toward a Sociology of Language" (1953) and his *A Sociology of Language* (1964) failed to create a following despite his sound ideas that both sociology and linguistics ought to join forces, for they are both quintessentially generalizing and categorizing disciplines, more than almost any others. All this is by way of avoiding an over-simplistic picture of the development of SOL.

It would be difficult to understand the foundations of American SOL without referring to the status of linguistic science up to and at the time of its explicit development, beginning mostly in the 1960's. Dissatisfied with what they considered the static nature of philology, with its predilection for written texts, a small group of scholars around Leonard Bloomfield at Yale, then Chicago, founded in 1925 the Linguistic Society of America, determined to make "linguistics" a science no less rigorous than the natural sciences and devoted to spoken language. American structural linguistics as it came to be called, was concerned with the structures of sound and grammar, described in a special framework and with symbols intended to bring uniformity to linguistic discussions. A number of exotic languages were investigated through native speakers, including many Amerindian languages, some of them known only to a few survivors. When Pearl Harbor occurred, it was natural that the Armed Services turned to the "new linguists" to set up programs in over 40 languages, known as the ASTP (Army Specialized Training Program), for some of the 11 million service personnel dispatched overseas.

Bloomfield's structuralism reigned supreme for two decades following 1945, the war's end. Although it brought a breath of fresh air and many innovations still valid today, Bloomfield's scientific aspirations created a bugaboo in "mentalism" or the so-called attempt to apply psychological suppositions, or extra-linguistic phenomena in interpreting linguistic ones. This was fully challenged by Noam Chomsky in his *Syntactic Structures* (1957) which became the manifesto of generative-transformational grammar or "TG," a theory which in its turn is questioned today, but which is still very powerful. TG at first turned its back on the phonological fixation, as it viewed it, and took a dim view of "phonemics," preferring to concern itself with syntax or grammar. Over-simplifying, Chomsky considered language at several levels, and not in the form of linear strings or symbols with meaning or sounds with meaning. At the base was "deep structure," roughly the semantic component or the message. A series of transformations expressed in a medium borrowing from symbolic logic mediated between the deep and the "surface" structure, the latter being what the speakers say or write. Language is, then, a two-way process in which messages are first encoded, then decoded. Chomsky has upon occasion stated that linguistics is, or ought to be, a branch of cognitive psychology, a stance which is anathema to what might be called "anthropological linguists," those following somewhat in the descriptive tradition of Boas, Sapir, Bloomfield, and a group of brilliant anthropologists aggressively concerned with language.

With early or vintage Chomsky at least, the social context of language, for differing reasons, fared even more poorly than with Bloomfield, because structural linguists could at least take refuge in anthropological linguistics and thus not be accused of the error of mentalism. In his turn, Chomsky wrote in *Aspects of Language* and elsewhere, that the business of linguistics was to study the speech of an "ideal speaker-listener" in a completely homogeneous speech community who knows the language perfectly" (*Aspects of Syntax*, 1965:3). Although fortunately, his contemporaries have chosen to ignore or take this rather lightly, it is clear that SOL could not thrive in either climate and to a large extent owes its existence to these constraints on the purview of linguistics. William Labov, largely working in the Chomskian framework, has suggested more than once that if the social context were adequately taken into consideration by linguists, there would be no need for a sub-discipline such as sociolinguistics.

Meanwhile during all this time, where in the world was dialectology? With few exceptions, it still followed the leisurely *Sprachatlas* approach of the European masters in a day when a leisurely pace allowed ultra-lengthy questionnaires and elicitation,

when interest was focused mainly on regional dialects, and what has been termed the "antiquarian" preoccupation aimed at identifying varying terms for everyday objects and terms applying to agricultural and household equipment. Nevertheless, it would be an injustice to state that Hans Kurath, the "father" of American dialectology, in adapting European methods to the U.S. regional atlases did not improve upon the patterns, cutting out some of the antiquarian concern and stratifying informants into three categories according to socioeconomic status, age and education. It was also at the University of Michigan, where Kurath held sway, that leading investigators of language diversity, like Raven McDavid and Roger Shuy to name a few, acquired their basic training. The uniform procedure used by the principal investigators of the regional atlases was a decided plus in avoiding general chaos. Nevertheless, a goodly number of dialect geographers joined the movement to explore and describe social dialects, while others chose to continue with revisions and modernizations of methods (cf. Allen and Underwood, 1971).

Social history in the United States was, however, ahead of language scholars. As long as the Melting Pot concept had remained a firm foundation of education, immigrants and minorities needed only to become Americanized, getting rid of foreign customs, languages, and, if possible, telltale foreign accents. When, however, social forces turned this around, rejecting the Melting Pot, and proclaiming in the Civil Rights Act of 1964 that our goal was instead "cultural pluralism" within our republic, a host of movements for ethnic and cultural recognition and rejection of enforced assimilation brought new demands and pressures upon our educational and scholarly establishment. These were reinforced by the passage of a great deal of landmark legislation, including various "titles" for the remedy of educational disadvantagedness, and particularly Title VII of the Elementary and Secondary Education Act (ESEA), popularly known as the Bilingual Education Act in 1967. Of utmost importance was the Lau vs. Nichols Supreme Court decision in 1972, mandating bilingual education as a bridge toward the acquisition of English. The burden was then placed on the school system and related entities. Over twenty states had to repeal existing legislation forbidding any language of instruction but English. In all this, the tenure of President Lyndon Baines Johnson may be considered the watershed of contemporary socioeducational legislation aimed at remedying long-existing social inequities.

All these developments brought the study of "social dialects" into the center of attention, also creating an unprecedented demand for more data and insights into the language patterns of the impacted populations of the inner cities, the rural poor, and, above all, the ethnic minorities. There arose urgent needs in ethnolinguistic data on at least 50 of America's more than 100 ethnic groups. Often it was easier to find such information on the languages and dialects of the Somoas, New Guinea, the Dravidians or southern India.

As the dialects of these minorities had largely been brushed under the carpet, few scholars and teachers had dared to devote themselves to such marginal fields. It was the sociolinguists who eagerly stepped into the breach so that the 1960's and 1970's witnessed a rich number of studies on nonstandard varieties of speech, particularly Black English and and varieties of U.S. Spanish. There is little question that the funding by government agencies made possible strides in SOL which the shrinking budgets of academia would not have supported. Individuals of SOL orientation are commonly consultants on the numerous bilingual projects, Head Start, Project Bravo and other intervention efforts supported by federal and state agencies. This section will be devoted to describing briefly the various models for the investigation of linguistic diversity which is the quintessence of SOL.

The study of SOL has been frequently attacked for really having no parameters and trying to be all things to all men. There is truth in some of this, but if SOL had not been developed, something like it would surely have taken its place. This branch or sub-branch of linguistics and of sociology coexists with a number of other fields, particularly anthropology, ethnology, psychology, neurology, speech communication, and others. A certain segment of formal linguists are indeed hostile to it, feeling that sociolinguistics is too watered down to be considered a serious science, while still others consider SOL as one of the recent models of linguistics itself.

Is SOL itself cultivated in a uniform manner? By no means, since its basic requirement is only that both social and linguistic factors be considered. At its best, it tries to establish, when possible, interrelations between both dimensions. A very useful taxonomy has been provided by Walt Wolfram (personal communication). He divides the field into three general patterns, as follows: (a) the anthropological model, (b) the linguistic model, and (c) an intermediate model falling between the two. Regarding the first model, its most distinguished exponents are Dell Hymes and John Gumprez. Hymes argued that if linguists would only broaden their concept of the systems of language to account for the functional varieties of speech, there would be no need for any new name for his proposed study — "linguistics" alone would do. Within five years, in "Models of the Interaction of Language and Social Setting" (1972), he attempted a somewhat more explicit blueprint for his model. "Speech situations," "speech events" and "speech acts" should be differentiated and just as explicitly analyzed and described in terms of rules as is the case with linguistic phenomena. The total "speech economy" should become fair, and indeed, mandatory game for linguistic research. In a clever mnemonic device, SPEAKING, he included some of the primary components of any speech event, proposing that these be incorporated somehow within the framework of generative grammars.

It is, to a large extent, as a catalyst that Hymes exercises a continuing influence upon this new, "more aware," linguistics. Through his prolific writings and his lecturing, he is known throughout the world. Nevertheless, as he himself has admitted (personal communication, 1971), the main shortcoming of his approach — which he is reluctant to term a true model — is its relative lack of use in actual field situations. An updated version of the Models paper also appears in the volume, *Directions in Sociolinguistics* (Gumperz and Hymes, 1972). Hymes edits the journal, *Language in Society*, published at Cambridge University, providing sociolinguists the world over with a much needed and respected forum.

Turning now to the linguistic model, we find a well-known author, William Labov, who entered sociolinguistics through the field of language theory. While a student of the late Uriel Weinreich at Columbia University, he developed an interest in problems of linguistic variation and change, forming the hypothesis that, despite the long-held neogrammarian view, change could as effectively be viewed in progress (synchronically) as in retrospect (diachronically).

Field work on the varieties of English on Martha's Vineyard, an island off Massachusetts (1963), and among the ethnically and socially diverse inhabitants of New York City (1966), enabled him to develop a methodology to identify and measure language change as it was occurring. To many, if not most persons in the field, his "Social Stratification of English in New York City," with its convincing presentation of the concept of the "linguistic variable," stands as the fundamental guide to the contemporary research on linguistic variation and change.

Thoroughly familiar with the principles of modern general linguistics, Labov has become increasingly dissatisfied with a theoretical orientation that assumes excessive linguistic homogeneity in the speech patterns of any given community. Rather than assume an "ideal speaker-hearer" or generalize to the community on the basis of working with a small number of informants, Labov assumes language is inherently variable and that the only viable means to measure its use is to measure it in use. Further, he expects that a great amount of that variability which previously was dismissed as "free variation," in fact, adheres to definite patterns determined by the linguistic environment on the one hand and the social environment on the other.

One of Labov's earliest theoretical contributions was the isolation of the linguistic variable. From the study of Martha's Vineyard on, this has simply been a high-frequency variant (in most of his studies, phonological) which turns up with such frequency in all styles of everyone's speech that, for practical purposes, it cannot effectively be suppressed. It constitutes sociolinguistic data. When correlated with such demographic factors as age, sex, and socioeconomic status and then further correlated with factors relating to style and discourse, interesting patterns of usage emerge for the speech community.

The variable rule first appeared in the exhaustive discussion of the findings of Labov, et al. (1968). It constitutes a device to relate the observed patterns of usage of linguistic variables to the format of writing actual grammars and has far-reaching implications for general linguistic theory. These rules are intended to supplement the bulk of the rules, mostly invariant, in a grammar, specifying the effects of the total environment — linguistic and sociolinguistic — on the output of the grammar.

Labov and his associates found that New Yorkers realize /dh/ and /th/ under certain conditions not as continuants, but as the highly stigmatizing stops, /d/ and /t/, respectively. He would formulate the appropriate variable rule like this:

$$\begin{bmatrix} +\text{cons} \\ -\text{voc} \end{bmatrix} \rightarrow \left(\begin{bmatrix} -\text{cont} \end{bmatrix} \right) / \begin{bmatrix} -\text{strid} \\ -\text{back} \\ +\text{cor} \end{bmatrix}$$

This rule departs from the standard Chomsky-Halle format only in that the additional convention of the parentheses around the feature to the right of the arrow indicates variability. So that they might be applied properly, the grammar must be supplied with an index of variability, $\emptyset = 1 - k_o$, for each rule so marked. Assuming that a rule actually were "categorical," that is, applied universally with no effective constraints, the value of k_o (which indicates the constraints) would be zero. However, this would not be the case with these rules, and degree of variability would be expressed as a function of the constraints of both style and socioeconomic class, as in the following formula:

$$k_o = f(\text{SEC, Style}) = a(\text{SEC}) + b(\text{Style}) + c$$

The precise values inserted into the formula for each linguistic variable would result from meticulous observation of actual speakers from the speech community in question as registered in real situations.

A number of younger and mid-career scholars have chosen to align themselves with the Labovian model. What we might call the "Washington school" is comprised of Roger Shuy, Walt Wolfram, Ralph Fasold, Joan Baratz (now in New York City), Orlando Taylor, David Sankoff and Henrietta Cedergren (1974) (in Canada) and others who have been developing some modification of Labov's model, particularly in the direction of "probabilistic" rather than "frequency" approach, with extensive application of computers and advanced mathematical concepts. Sankoff is a professional mathematician. In Los Angeles, Benji Wald has worked on Chicano Spanish and English in the eastern part of the city, while Shana Poplack has concerned herself

with Puerto Rican Spanish in New York City and John Baugh at the University of Texas-Austin has addressed himself to language varieties of the Southwest, including Black English. In large part, due to the diligence and dedication of these "disciples" of William Labov, Black English has been extremely well-covered sociolinguistically, probably more than any language variety. At the same time, it is safe to assume that most of this group would hardly be flattered to be identified exclusively with only one line of research. Space forbids discussion of other interests. Finally, an able sythesizer of sociolinguistic work along the Labovian model, as well as a researcher in his own right on Brazilian Portuguese syntax, is Anthony Naro (1981) at the Universidade de Campinhas.

Much like Labov, Gumperz and Hymes, Joshua Fishman has become extremely well-known through his copious writings. Of utmost importance, both for the data secured and the research design, is the *Bilingualism in the Barrio* project, performed by Fishman and a number of associates (Fishman, Cooper and Ma, 1972). This seems to be the most comprehensive study made of the bilingualism of any group in the United States, or perhaps the world. Fishman, trained as a social psychologist, makes use of a very broad battery of techniques from both linguistics and the social sciences. The study focused on 431 Spanish-English speaking Puerto Rican subjects from 90 families on three adjacent blocks of Jersey City. They were mostly young unskilled workers with little formal education and from the lower socioeconomic status. For contrast, stratified samples were also taken of intellectuals, high school students and other groups from the Puerto Rican community ranging into New York City. Extensive ethnographic information was gathered on the subjects, including demographic data and details on life-style orientation.

The elicitation battery was extremely broad, including self-report on language fluency, language loyalty and the like. Open-ended interviews supplemented this data, particularly for precise details on the relative use of English vs. Spanish in Fishman's five hypothesized domains of home, neighborhood, church, school and work. Use was made of a small inventory of linguistic variables, along Labovian lines, in both Spanish and English and based partially on Tomas Navarro Tomas's studies of Puerto Rican Spanish dialectology.

The language output elicited from each subject was rated by independent judges along the following dimensions:

1. Accented speech, with a seven-point scale in which the highest scores denoted Spanish dominance, the lowest, English and a medium score assigned to "balanced bilinguals";

2. Reading aloud of selected passages, similarly rated;

3. Writing, based on self-report by one household member, on a three-point scale, subtracting English scores from Spanish, with positive scores indicating Spanish dominance, negative, English;

4. Spanish repertoire range with respondents globally rated on a four-point scale, the top figure indicating control of a number of styles. These styles were defined along Labovian lines and consisted of word list reading, passage reading, careful speech and casual speech;

5. English repertoire range, identical with the above, but on a five-point scale.

Attitudinal testing was modeled on Lambert's techniques, and subjects were asked to make value judgments, along a polarized Osgood semantic differential-type scale, of the personal characteristics of speakers heard on taped samples. One innovation, moreover, was the testing in which similar procedures were employed to evaluate the contextual appropriateness of the English, Spanish, or code-switched variety of speech on taped samples. The significance here is that these researchers displayed a departure from traditional unrealistic tendencies to regard bilingualism as the opposition of two standard languages, rather than as several coexisting varieties within languages.

Elaborate use of such statistical techniques as factor analysis, regressive analysis and others permitted evaluation of a large number of correlations among the numerous socio-educational and linguistic variables represented in the study.

Initially it should be pointed out that Andrew Cohen's Redwood City study is one of the few sociolinguistic studies on a Mexican-American population (1970). Subjects were twenty low-income first and second grade Spanish speakers and ten middle-income English speakers of the pilot bilingual education project in Redwood City, California. Research design represents a combination of techniques adopted from the Fishman Puerto Rican project with guidance from team member Robert L. Cooper of that earlier project and suggestions from Wallace Lambert's well-known attitudinal investigations. The elicitation battery consists of five instruments administered by bilingual interviewers to individual pupils and their parents, respectively. These include demographic items, attitudes toward the languages and cultures, word-association tests for Fishman's five domains, and an ethnic image test based on the Osgood-Tannenbaum model.

Cohen's linguistic observations are mostly impressionistic, but valuable in their insistence that deviant forms may well be socially significant in themselves as expressions of their users' felt cultural identity and should not be automatically assigned to contrastive standard language differences or the "culture of poverty." A further distinctive feature of

the study is the parallel set of questionnaires for parents, instead of relying solely on children as subjects.

Today the field of linguistics has been described as one in which so much reappraisal is going on that it is in a state of utter chaos. It is perhaps because of this ferment that so many look to SOL, whatever be the prototype, in order to save their sanity. Allen Grimshaw, a sociologist, has written of SOL as a hybrid discipline with a short and largely atheoretical history (1971: 135), but that was years ago. Nevertheless, SOL has one redeeming feature and that is its commitment to field study of different populations. It is based not on speculation, but the elicitation and analysis of data.

C. Toward a **Sociolinguistic Orientation**

We have sketched the trends in American linguistics, then moved on to the development of SOL, going from there to research on dialectology, formerly concerned almost exclusively with regional dialects, but which has by now added a new dimension — that of social dialects. At any rate, SOL has shown a very hardy growth, so that it covers not only social dialects but also language policy and planning, reading problems, literacy, and speech communication. At many points, it is sometimes difficult to demarcate where one study begins and the other ends, but I would consider this a plus rather than a minus. Just as national decisions are said to be too important to be left exclusively to the military, SOL problems are far too significant to be entrusted exclusively to narrow-gauge linguists. With over 5,000 languages and countless dialects on Earth, yet only some 250 sovereign nations to house them, our agenda of problems of language and dialect in today's society contains all too many priority items.

It should be emphasized, moreover, that more than being a sanctified body of facts and principles, SOL is at its best an enlightened orientation, that is, a general attitude or approach that attempts to view each language problem in the light of its own sociocultural setting.

Unfortunately, the very speed with which the mushroom-like growth of the bilingual education movement has moved has acted to prevent the application of SOL where it could be helpful, i.e. in general bilingual/bicultural education operations and especially in assessment. The world term "sociolinguistic" or "sociocultural" are bandied about a great deal, but very often that is as far as it goes. SOL has become a "must" when one speaks about bilingual/bicultural schooling, much more than a concept that is applied to any large extent. What passes for SOL in many teacher training programs is merely warmed-over course material on culture and at its worst might be termed the "taco trail" or Mexican hat dance approach, providing pleasant entertainment through dances, songs, and information on exotic customs. Something more preferred is obviously needed, which will walk a tight-rope between engaging in propaganda or partisan advocacy for other cultures and systems and describing variations in them, avoiding dangerous generalizations which do not take into account differences on ethnicity, socioeconomic status, and other factors in a given community, region, or nation.

Addressing assessment for a moment or so, I would say that this is virtually the most difficult operation expected of bilingualists and school personnel, since it consists, by at least conservative estimates, of so very many variables — a couple of dozen at the most up to a hundred or more. Accordingly, an instrument that would address all the variables would become so complex that it would be counterproductive.

At the 1967 International Conference on the Description and Measurement of Bilingualism at Moncton, Canada, several distinguished colleagues addressed themselves to this. E. C. Malherbe, Union of South Africa, declared:

> It is doubtful whether bilingualism per se can be measured apart from the situation in which it is to function, in the social context in which a particular individual operates linguistically. The only practical line of approach to this complicated problem which I can suggest is to assess "bilingualism" *in terms of certain social and occupational demands of a practical nature* in a particular society. Here again the criterion is to be "bilingualism for what." Purpose and function is the only determinant (Malherbe, 1967:50).

Thorny technical obstacles in the way of assessment were also pointed out by Bertha Siertsema, a Dutch linguist. She even went so far as to state:

> As long as the different intonational features can alter the entire meaning of one and the same "phatic" utterance, can even change it into its opposite as in irony or sarcasm, what sense is there in measuring phatic material only (supposing all the time that this would indeed be possible)? Had not we better drop the whole idea of "measuring" in the study of bilingualism? (Siertsema, 1967: 155)

It is time now to present a basic stock of SOL foundations which are pretty generally accepted in the field. It is not felt necessary to indicate origin or attribution of the concepts, since they are well-known in the literature.

Cluster A: Language Is Flexible and Varied

1. Language is not static, but dynamic, ever changing.

2. Variation in language is much more common than homogeneity (cf. Labov).

3. More of the earth's dwellers inhabit regions of language diversity where there is more than one language or dialect, than those that live in fairly homogeneous regions, such as Japan, Uruguay, Iceland.

4. When languages are in contact, they usually influence each other's vocabulary features. In extreme cases, they may become a *mischsprache* or mixed variety which has been common in most immigrant languages spoken in the United States. In most cases, at least, there will be code-switching, often at the subconscious level.

5. All languages, as far as we know, have a number of styles, alternating in their degree of formality vs. informality. One may think of language varieties, a term which we prefer to *dialects* because of the pejorative connotation of the latter term.

Implications: In Cluster A, we encounter a fairly relativistic view of language, rather than an elitist one, which regards language as a set of sacred symbols which must not be defiled and which must be guarded in its pristine form. As in the world of nature, language changes as a result of language contact and other forces, such as deliberate measures by governmental bodies, or even for reasons difficult or impossible to fathom — sounds and features change in certain directions.

Teaching and testing along "puristic" or prescriptive lines tend to aid and abet the already widespread tendency toward ethnocentrism and regarding the variety of speech spoken by some elite as inherently superior to all others.

Cluster B: Language in Society

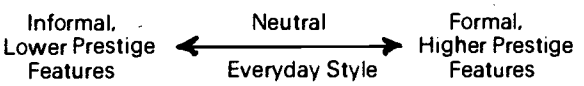
1. Language occurs in a societal context and not in an abstract vacuum.

2. Language is in itself neutral, but values and qualities are assigned to it by people.

3. Language generally follows the stratification of small, specific groups (small groups) of society so that the way we speak and write constitute immediate markers by which people judge who we are and from whence we come, both geographically and socially.

4. Language can be studied, as Fishman suggests, both as "macro-sociolinguistics," or as the role of language among groups and organizations including regions and nations concerned mostly with the speech circumstances of the speech behavior of individuals. This gives added perspectives and a fuller range of views of the different roles that languages play at various levels.

5. A specific language viewed sociolinguistically is more easily seen as a linguistic continuum, with informal, nonstandard features at one extreme and with formal ones clustering more toward the other polarity. Thus, in the middle, we may posit an everyday style, containing both informal and formal elements.



Bilingual communities may have different languages at each point.

Implications: Both teaching and assessment are rendered more complex in view of the continuum. As Labov says, there are presumably no one-style speakers anywhere in the world. Teaching and assessment need to be constantly guarded against the excessive artificiality marking so much instruction and testing. The answer is, of course, not in the direction of "anything goes," but rather in that of exposing the students to variation through their specific samples, which they may in turn learn to "mark," both as a class or in a society-testing heuristic.

Cluster C: Domains of Activity

1. Language may be regarded from many viewpoints, such as in prose and poetry studied as literature, or sociolinguistically from the standpoint of the domains of living in which, let us say, each of a language pair or set are utilized.

2. Sociolinguistics is particularly interested in which domains specific languages or language varieties are utilized. Very commonly, a language of lesser prestige may be used in informal domains, such as the home or the neighborhood, while the dominant or prestigious language may be employed in church, school, and the work place.

3. Social changes and new alignments change from colonial status to independence, which may completely alter the above relationships.

4. A language or language variety may, therefore, be made to serve whatever function for which it may be needed.

5. At the same time, language planning may be necessary before a language used only for informal purposes can be equipped to function in formal domains. Many states often have to take specific measures to provide new and special technical vocabulary in line with needs of modern government and technology.

Implications: The above further extends the picture of language. This broadens the scope of one's knowledge of places where language is not standardized and has been in literate use for many years.

Languages which are only used for informal, everyday spoken purposes are called vernaculars. To a large extent, Spanish in the United States Southwest has a vernacular role, and English serves for formal, technical purposes.

Cluster D: The Variables of Performance

1. Age, sex, language, ethnicity, and socioeconomic class are some of the main variables influencing performance in a given language.

2. Situational factors are another sort of variable. In general, sociolinguists attempt to have this occur in as natural a place as possible.

3. Psychological variables are still not fully understood, but research and observation show that the following are influential: (a) loyalty to a given language or culture; (b) attitude(s) toward a foreign language/culture; (c) attitude(s) toward a foreign language/culture, either negative or positive; (d) motivation for studying the language, be it instrumental or integrative.

4. Coincidental factor(s), such as family crisis, ill health, anxiety neuroses in test-taking, may be very influential.

Implications: As the above constitute a minimal inventory of variables, it is already evident that variables of performance are both numerous and, in many cases, factors or conditions not of the student's making. It would be impossible to design instruction and assessment so that each variable would be explicitly taken into consideration. Accordingly, as far as assessment goes, it is important to give each student sufficient opportunity to perform at his/her optimum.

Cluster E: Desirability of Differing Options

In general, a sociolinguistically trained person would avoid "all or nothing" solutions. Answers to language problems in a multicultural setting need to be developed from an examination of local conditions, if this is at all possible. Fortunately, language is a field in which wide variation exists, and options can vary greatly. To the extent possible, pupils and their parents need to have a voice in the following:

1. Since the home language is already known, all efforts should be expanded in mastery of the dominant language (here, English).

2. The choice can be a "monoliterate" one, in which the home language or dialect is utilized for discussion of cultural materials, but only the mainstream language is used for reading and writing skills.

3. Instruction is given in the minority language, only for developing reading and writing skills.

4. A part of the curriculum might be devoted to the cultural background of the minority group.

5. A maintenance solution, in which a larger part of time is devoted to the dominant language, may be used especially in scientific-technical areas.

Implications: Strategies for dealing with any or all of the above solutions are to be found, although not always in a highly developed manner, in the literature of bilingualism. There is room for a great deal of experimentation, difference of "treatment," between an experimental and control group following a fairly simple research design. Many of the answers needed by educators have simply not been researched. One thing seems certain — the dominant home language usually utilized as a vernacular can and does often serve as a useful bridge toward increasing mastery of formal aspects of the home language. Nevertheless, this does depend upon the "distance" between formal, standard varieties and informal ones. If the distance is too great, it may be best to treat the second language as a foreign language. (Spoken Cantonese vs. Mandarin is a striking example. Even here a bridge is possible through writing and reading, since the written characters communicate to any literate Chinese reader.)

Cluster F: Community vs. Home Standards

In general, there is, in most societies or regions, one language whose speakers form an elite. They are usually fully in the mainstream. There may be any number of "other" ethnic groups, some even more numerous than the elite, with limited access to the mainstream.

1. Situations facing multilingual pupils vary greatly in different countries and regions. The country may have *diglossia*, that is a Low vs. High form of the same language.

2. What is the purpose of bilingual education, assimilation or maintenance of minority language?

3. Does the minority community reject the values of the majority or dominant culture?

4. How different is the variety of the dominant language taught in the school from that spoken by mainstream and minority speakers?

5. Is there a wide separation between the variety of the minority language taught in school and what the children hear in their own home and community?

Implications: A wide spectrum of issues is evoked by this cluster, and they are ones for which there are no pat answers. Sociolinguistically, however, since language varieties (language and dialects) occupy differing communication roles, instruction

and testing can afford to vary pedagogic treatments and approaches to the testing of language proficiency.

Cluster G: Ten SOL Insights for the Classroom

More than any one of the ideological clusters presented, there is room for sociolinguistic consciousness-raising for all involved in the teaching and testing process. Given any reasonable sociolinguistic exposure, an individual will not view language phenomena as a series of dogmatic oppositions, but rather as continua or, in practical terms, as a set of alternatives from which solutions can be chosen. The following items represent a sociolinguistic attempt to arrive at a sociolinguistic type of measure or better measures embodying SOL awareness.

1. Teaching and assessment both should emphasize language performance in a variety of different situations.

2. Although there is a continuing need for numerical assessment, performance-based criteria for specified situations need to be developed far more than is the case at present.

3. An indispensable component of any language training program would be a realistic orientation to the phases of the target culture which must be known in order to function in a given language without avoidable faux pas. Again instruction and testing should be done through the presentation of situations.

4. As part of the cultural component, there is need for training and assessment in kinesics, e.g. the use of space, eye contact, gestures, etc., since this varies from culture to culture.

5. The notion of "only one form can be correct" flies in the face of SOL approaches suggesting artificial contexts for language, which hardly correspond to reality.

6. Knowledge of the regional ethnic culture as well as the ancestral culture and civilization should not be assumed. Unfortunately, there is little agreement on the content of such a component, offering here a challenging field of research. Little research has been done on ways in which children learn and internalize cultural knowledge.

7. In view of the younger ages of most bilingual pupils, ways should be sought to enhance the effectiveness of modalities. The serious drawbacks with paper and pencil and such measures as the draw-a-man are too well-known to need elaboration. Closer to a naturalistic situation is the coordinated use of videotape and slides presenting realistic situations and coordinated with a synchronized tape or cassette recorder.

8. A close tie-in should be sought between entry assessment, subsequent classroom instruction, and follow-up assessments. Types of individualized instruction or small ability-group divisions can be useful here.

9. Sociolinguistic insights from research ought to be combined with those from psycholinguistics (including neurolinguistics, pedagogy, and social anthropology) on a continuing basis. These can be incorporated into both assessment and instruction.

10. SOL research data in "raw" form may be ambiguous or misleading if brought directly to the classroom context. There is need for processing such information, so that interpretations can be made relevant to the experience of both teachers and pupils.

Implications: SOL research data for assessment and related language operations needs careful processing for the classroom context.

D. Applying SOL Notions to Language Assessment and Related Operations

There are perhaps few sub-fields of linguistics and social science which are so amenable to classroom application as SOL. This is simply because the very essence is perceived in the variability of language and speech. Accordingly, it is light-years distant from the prescriptive approaches to language which attempt to promote the concept that there is almost always one, and only one, "best" or "most correct" form in a language and that it should be used in all situations. I have termed this the "unitary fallacy," which is replaced in SOL by the notion of "situational appropriateness" as the determinant of what is "best" to use.

In second language teaching, we deliberately teach the most commonly accepted standard variety, usually of a somewhat formal nature. That makes sense for the purpose. In bilingual/bicultural education, however, more often than not we are dealing with pupils who come to us speaking different dialects of a language and, by definition, somewhat nonstandard varieties. As regards Spanish, which accounts for the majority of bilingual programs in the U.S., there are three major varieties — Chicano, Puerto Rican, and Cuban Spanish, but even a much larger number if we take each national form of Spanish American language spoken by immigrants from Columbia, El Salvador, Dominican Republic, Venezuela, Ecuador, and other nations. Just think of it. There are nineteen countries where Spanish is the official language; no other language can claim such a number. At any rate, it is precisely SOL which prepares us better than any other study to cope with the variation. We ourselves cannot be masters of all varieties, no more than we can be of all varieties of Portuguese — European, Brazilian, Azorean, colonial African, and U.S. The result is that

we program the most versatile of language teaching computers — the human being — to deal with variation so that the students can escape feelings of inferiority because their speech is different and not standard and so that we can put existing language skills to use as a bridge to the standard variety. This goal is overwhelmingly approved by parents and, at the best, is merely giving individuals another option to enrich their linguistic and bilingual repertoires.

Among the linguists who have specifically addressed assessment, as well as a broader spectrum of bilingual education instructional issues, are Rosana Sanchez and Eduardo Hernandez-Chavez. Their ideas are of special relevance for us (Sanchez et al. 1978: 7-63). Professor Sanchez holds that a satisfactory assessment battery must address skills in the areas of linguistics, cognition, and sociolinguistics. The linguistic dimension calls for an examination in the students' language varieties of their ability to communicate in and conceptualize about topics and situations natural to their experiential domains. Once these skills are determined, the task of the school will be to increase student repertoires so that they may express themselves in more than one variety of a language when it is appropriate to do so. This means, frankly, code-switching during interaction with ethnic peers since this is the normal way to converse. Cognitive skills in computation, abstraction, or sythesis must also be assessed as they relate to academic, home, or sociocultural settings. All these skills must be assessed in both languages so that we may be able to determine the students' proficiency and enable them to develop their primary language stage of abstract conceptualization (Sanchez 1978: 51-52).

In order to illustrate Professor Sanchez's blueprint for assessment, we will reproduce here two diagrams detailing her plan.

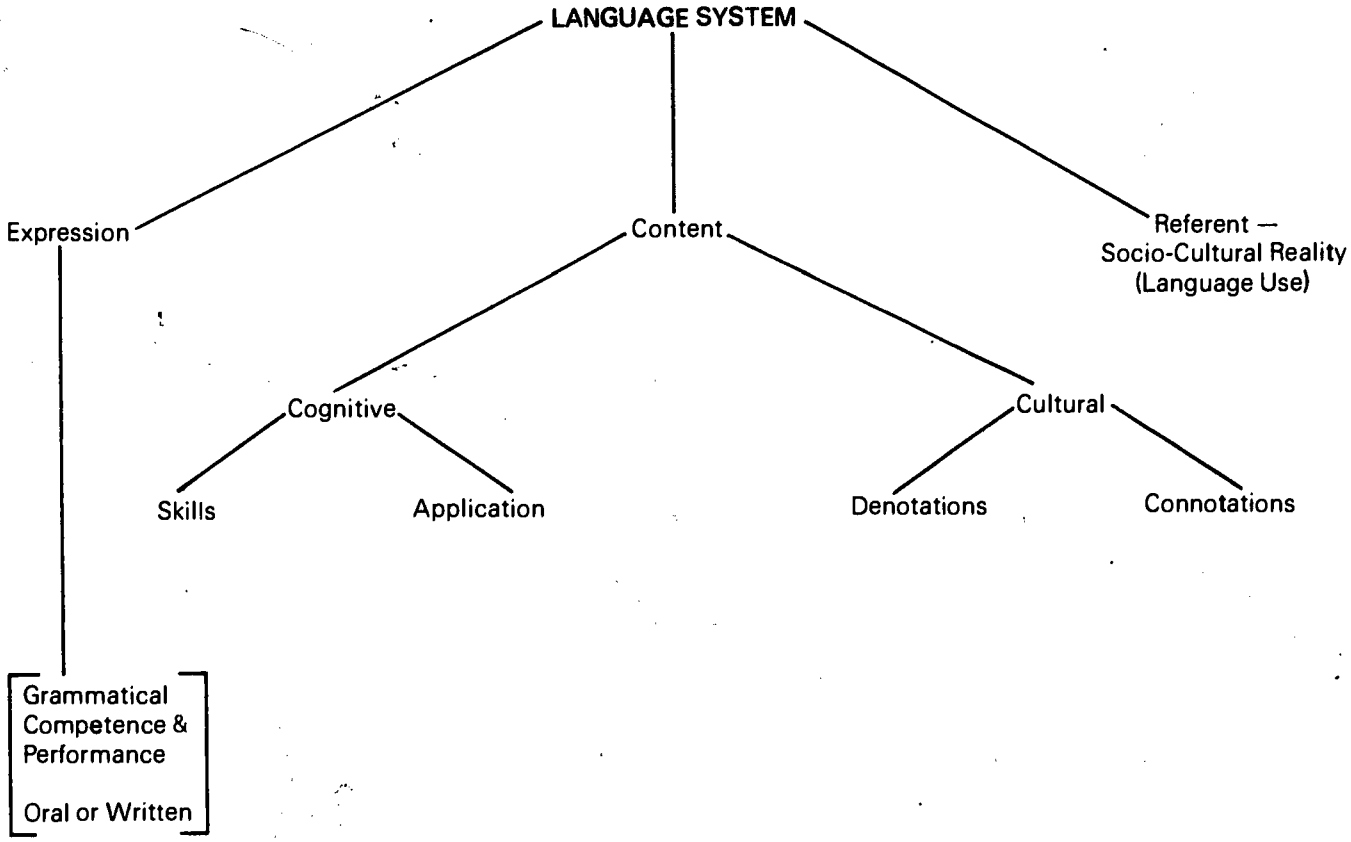
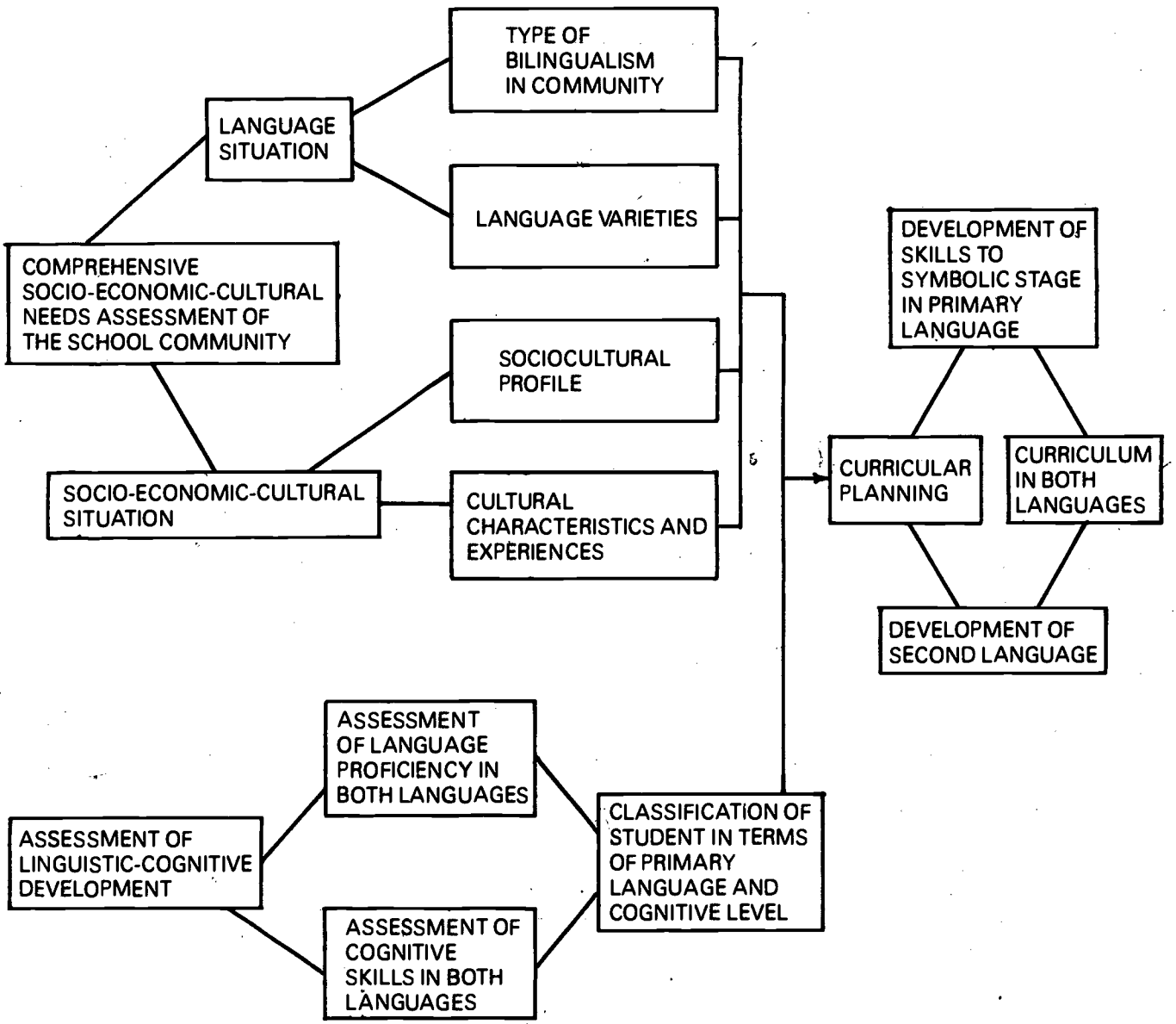


DIAGRAM I

Sanchez, 1978: 17

DIAGRAM II



Sanchez, 1978:15

Eduardo Hernandez-Chavez (1978:47-56) likewise would consider that there is much room for improvement of assessment instruments, although he himself is coauthor of the Burt-Dulay *Bilingual Syntax Measure*. His ideal instrument would seek separate ratings for each of the following dimensions: (1) linguistic, (2) sociolinguistic, and (3) metalinguistic skills. These would be accompanied by specific comments regarding each individual's performance, addressing such items as degree of standardness vs. nonstandardness in the speech output.

As regards the three dimensions mentioned above, the linguistic one involves performance in the reading, speaking, writing, and listening skills of the two languages. The sociolinguistic parameter would, in Hernandez-Chavez's view, seek to determine the pupil's proficiency as well as dominance in the separate skills of each language and in the domains familiar to him or her. Taken into consideration would be the fact that ethnic peers normally code-switch with one another, while with a monoglot Anglo he would resort to such devices as intonational shifts, gestural patterns, special lexical choices, and the like. Quoted aptly, Professor Hernandez-Chavez warns that there is hardly ever an equal commensurability of language use with any two domains in two different languages. This is evidenced by such speech acts as asking for a favor, a problem that is hardly ever solved by translating words directly from say English to Spanish, or vice-versa. It is, in the final analysis, the cultural values of the speech community which account for the shape that such a formula should take.

With respect to the metalinguistic dimension, this refers to overt conscious knowledge about linguistic phenomena, such as grammar and orthographic rules, as well as the meanings and values attached to regional, social, or literary variants of grammatical or vocabulary forms. Obviously, if a student is unaware of such distinctions, he would have real trouble in interacting with other people of the other culture, no matter how impressive the linguistic component may be in his case.

The literature of bilingual assessment is by now considerable, although time and space forbid our doing justice to the entire gamut of experience and opinion. That SOL, although widely recognized as "important," is poorly or not at all integrated into most assessment instruments becomes apparent upon even a cursory examination of commerically available instruments.

An examination of the extremely useful compilation, *Assessment Instruments in Bilingual Education: A Descriptive Catalog of 342 Oral and Written Tests*, prepared by the Northwest Regional Educational Laboratory (NDAC, 1978), helps us gain a perspective on the extent to which SOL has influenced such instruments. For excellent reasons, the instruments reviewed focus on language proficiency sometimes in one, sometimes in two languages, with little overt signs of the presence of SOL. The fourth section is titled: "Sociocultural, English, Spanish, Other Languages." A total of 41 instruments are described, measuring aspects of students' environment and their attitudes toward their environment, particularly the school. Language attitudes are also measured by some tests, while certain ones also offer parental attitude measures. These factors are certainly aspects of SOL, but sociolinguistics goes beyond them. At the same time, the attitudinal dimension is also within the purview of psycholinguistics and even education theory. There is no way to escape the fact that SOL is inextricably interrelated with other cognate disciplines. All in all, this represents positive steps forward by comparison with the state of the art in 1968, ten years before this compilation appeared. A briefer survey can be found in Duncan and De Avila (1979:15-50).

At the same time, a consideration of only commercial and published assessment instruments offers a skewed view of practices in the field. A survey of what is actually employed by bilingual projects for measuring proficiency/assessment, based in part on a sample questionnaire, has been carried out by Stan Seidner. In his paper, "Establishing Entry/Exit Level Criteria for Language Assessment," presented June 5, 1981 at the Ethnoperspective Forums, Eastern Michigan University (proceedings in preparation), Seidner found that in over 66 per cent of project directors responding, the practice was to use several instruments — representing both commercial and self-prepared measures. Additional light will be thrown upon this matter at the Language Assessment Institute, June 17-20, 1981 at the National College of Education, Evanston, Illinois. (This institute was organized by the School of Continuing Studies of that College, with the cooperation of the Illinois State Board of Education, Bilingual Education Service Center, Educational Testing Services and National Clearinghouse for Bilingual Education.)

There is no better way to end this section than to underline the extremely difficult task of embodying SOL in assessment instruments, to a large extent because it is difficult to gain consensus on what shape this should take. In a paper also presented at the Ethnoperspectives in Bilingual Education Research Forum on Bilingual Education Technology, Eastern Michigan University, entitled, "The Assessment of Language Proficiency in Bilingual Chilren: An Analysis of Theories and Instrumentation" by Charles Stansfield, the difficulty of doing justice to a test based on a tripartite theory of language component, communicative skills, and sociolinguistic domains is detailed. As Stansfield works it out, it involves a matrix of 32 different cells, each of whose tasks require separate validation. All in all, a challenge of utmost complexity is involved. In the discussion period referring to Stansfield's paper, he went so far as to claim that instruments which would attempt to specifically address each aspect described above would probably become too unwieldy to administer and analyze (Stansfield 1981, personal communication).

Thus, we are on the horns of a dilemma, and with each effort to expand the scope and depth of assessment instruments, the complexity of such a task is greately compounded. At the same time, many of the thorny issues are illuminated in the other papers of this volume, covering papers presented at the National College of Education Language Assessment Institute (Seidner, ed., forthcoming).

E. SOL as Dynamic Heuristics — Personal and Group Involvement

As already suggested, SOL offers no panacea capable of solving all educational problems. What may be considered its main terminal objective is that of sensitizing the individual to discriminate among the language varieties and to be more aware of the existence of options in the world of language and speech. This, far more than the sum total of facts learned in such training, is what makes it valuable since of all human activities, there is hardly one in which we engage more than speaking and listening. From learning about "language situations" the world over, the student of SOL usually develops the ability to make judgments about the linguistic situations which s/he encounters at home, at work, and at school and elsewhere. One of the most common exercises in SOL courses is the drawing up of a SOL profile of a country, region, or town. Often after students have completed such a profile of their own home town, they will exclaim, "My, I never realized that we had such different kinds of people, and so many relationships among them." This is quite understandable in countries like ours in which the goal of egalitarianism is reflected in such slogans as: "Tall people are equal," or "There's no discrimination in this town." Nevertheless, as George Orwell wrote in his classic 1984, some are *more* equal than others. Even in a democracy, role, power and functional relationships are operative, and these can be identified in any SOL course worth its salt.

Ideally, I would personally present the relationship of SOL to language assessment semiotically in the three following patterns:

SOL and Teaching/Assessment Models

1 Formal/traditional

SOL	Teaching	Assessment
☐	☐	☐

2 Mixed

SOL	Teaching	Assessment
☐	☐	☐

3 Variable SOL Oriented

SOL	Teaching	Assessment
☐	☐	☐

Everything considered, then, the main goal of SOL, aside from the actual data presented, is the development of attitudes which permit the individual to make decisions about language options in a world where some 250 sovereign nations house more than 5,000 distinct languages and an almost infinite number of dialects and sub-dialects. Stated another way, in modern technological terms, the most desirable terminal objective or criterion-referenced goal is to "program" individuals, be they students, teachers, administrators, parents or community activists. This must be done sociolinguistically so that they are liberated from our traditional heavy load of myths, misconceptions, and prejudices about other languages and cultures and can make enlighted and practical language-culture decisions. Having given some notion of the virtual futility of explicitly incorporating SOL into all teaching techniques and materials, the alternative is to program the most versatile of computers — the human being.

Indeed, in her fine review of foreign language testing, Helen Jorstad (1974:223-273) emphasizes the extreme difficulty, if not impossibility, of adequately testing the cultural component of language. The tester is trapped between the Charybdis of selecting an inventory of cultural items about which no two persons usually agree and the Scylla of involving value judgements of a subjective type. What should cultural or sociocultural teaching and assessment materials and instruments contain? It is a wide-open and moot question.

One of the highest priority problem areas that have already received a certain amount of attention is that of attitudes toward other languages and cultures on the part of youngsters at different stages of development. Most of the research on this, as exemplified by a brilliant group of Canadian scholars, has been in the form of the matched-guise technique, mostly applied to young adults. There is a desperate need, particularly in the bilingual/bicultural schooling context, for investigations that will illuminate the ways in which young Americans acquire their abysmally negative perceptions of other languages and life styles. It is too simple to say merely that it is all the fault of the parents. Unfortunately, U.S. social history reveals in our society deeply rooted hostilities toward things "foreign," call it jingoism, chauvinism, know-nothingness or whatever. These are compounded by now outmoded preceptions that foreign language/culture study has little or no integrative value. A child interested in language study is marked as a "lame" by his peers and may drop it hastily. As one who spent most of his life trying to teach and inspire red-blooded American youngsters with my own passion for our subject, I threw up my hands after 25 years or so of the effort and went into research, lecturing, and consulting. Please don't judge me too harshly!

The results of our mercilessly competitive international front has been the retention of a "Let'em Learn English" attitude and the image of the ethnocentric Ugly American. A recent Russian immigrant told me in El Paso that when a Soviet serviceman presented himself at the U.S. Embassy in Kabul, Afghanistan, wishing to defect to the U.S., there was not a single person in our staff there able to speak Russian to him, and a special plane had to be dispatched to another capital, almost a thousand miles away, to bring a trained polyglot. Of the fifty-odd hostages formerly held in Iran, only one or two could speak Farsi with any functional level.

This problem cannot be solved, despite our fondest dreams, by trying to devise still another breakthrough, complete with computer intervention for teaching methodology or testing. A two-pronged attack is needed that will research the anti-language attitudes and their development and, as a second phase, will undertake an aggressive but informed campaign supported by educational, community, and related agencies. This would be in line with the American way of influencing our institutions, effectively applying pressure to the proper governmental and political organizations. Unless this, however, obtains wide grassroots approval, it will fail as have such movements as FLES — Foreign Languages in the Elementary School — which sparkled in the 1950's, only to eclipse in the 1960's.

As this section does not purport to being a comprehensive treatment of research targets in sociolinguistic aspects of assessment, it will suffice to add only that the writer sees two additional broad areas which he would typify as highest priority on our agenda for today and tomorrow.

The first area is in the realm of interdisciplinary r search — precisely at those points where SOL intersects with psychology, sociology, social anthropology, psycholinguistics, neurolinguistics, and speech pathology. The latter field has become unusually open to current trends from linguistics and indeed has begun to share with us some of its important insights, particularly as there are problems of variation in styles of speaking and performance involved. The old days of relegating speakers of ethnic dialects to retarded classes, special education, are unfortunately not completely behind us. There is much more.

The second area may be seen in the still underdeveloped activity of data-gathering and research at the micro level. Language practitioners usually emerge from backgrounds in which they have not been exposed to even the simplest social science research procedures. The awe which then results from the medium and terminology of statistics, with its bell-shaped curves, distinctions between the mean and the average, levels of confidence, chi-squares and regression analysis is well-nigh paralyzing for us. Yet, at the same time, as Cynthia Darche-Park pointed out at the 1981 Ethnoperspective Forum in Ypsilanti, (Lujan & Park) teachers can do important types of data gathering which, if properly processed, can supply insights not available to an outsider with the most sophisticated statistical techniques. This is not to attack the validity of research which employs modern procedures, but rather to suggest that there is room for both types of inputs. Due in part to the great speed with which bilingual education schooling has been organized, many practices today — out of sheer necessity — are performed in an a priori way. There is no time to carry out large longitudinal studies when governmental mandates and requirements demand immediate execution of measures designed to relieve disadvantagedness on ethnolinguistic bases. No aspect, you will agree, has been quite so heavily "under the gun" as the requirement for producing instruments or measures for assessment of language proficiency and dominance.

Encouragement of micro-level research can and should bring us a great deal of data and figures and insights not yet available — all of which are badly needed. I am saying here that although focused research projects are essential, ethnographic or broad-gauge investigations can also, by serendipity, make us more knowledgeable about our communities and the characteristics which distinguish our students, parents, teachers and the community infrastructure. Much of what we believe about our field is simply based on hearsay given authority by dint of repetition. We must open ourselves to learning more about a topic such as bilingualism. As one of the pioneers, Einar Haugen, has said, in it everything and anything is possible. A tiny example will suffice here. In his important study on "The Neurology of Bilingualism," Yvan Lebrun (1976) brings convincing evidence that the long-held belief that when aphasia caused by a trauma occurs, the strongly embedded mother tongue will — if anything — return at least in part to the patient, while the language learned later will recede or disappear. Research on specific cases shows quite the opposite frequently to occur.

F. SOL Research at the Local Level

The remainder of this paper will concern itself with various specific types of research which can be performed with very small or no monetary outlays, at the community level. As Cluster G, Item 10 suggests, the data gleaned can be extremely useful as a continuing data bank for bilingual/bicultural education.

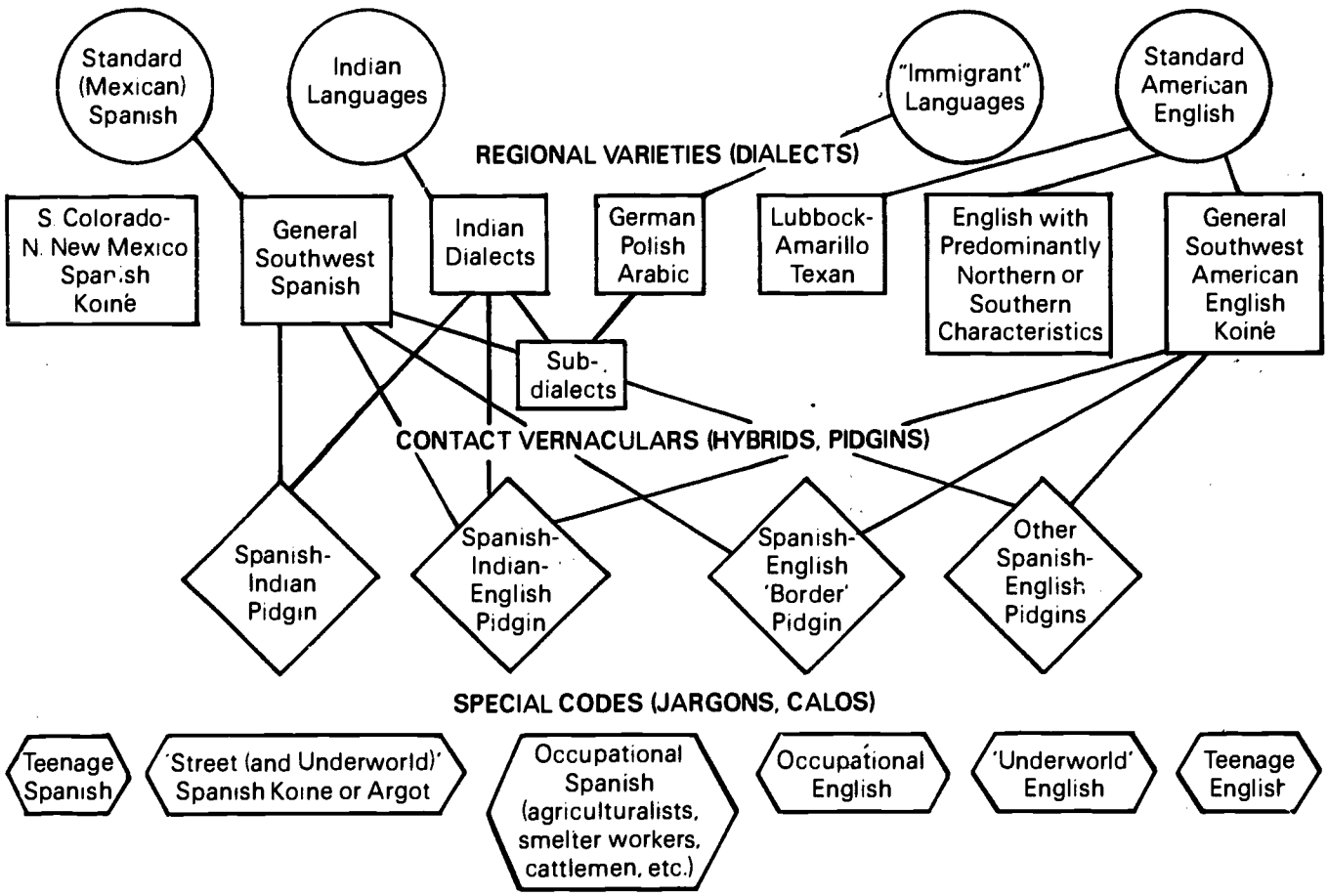
The experience on which the following is based was acquired with a Sociolinguistic Research Team, which was active at the University of Texas — El Paso between 1969 and 1979, mostly on a volunteer basis involving university faculty from different departments. Minimum requirements for such a team are at least one language specialist and one social scientist or educationalist familiar with standard social science research techniques.

In 1969 several of us colleagues at UT-El Paso decided that we needed to know more about our college population to see whether there were significant differences between the Anglo and the Chicano bilingual students, comprising more than a third of the undergraduate enrollment. Our Sociolinguistic Research Team consisted of an educational psychologist, an educational administration specialist, a sociologist and myself as a linguist. Our first task was to devise a Sociolinguistic Background Questionnaire (Brooks, Brooks, Goodman & Ornstein, rev. 1972) which was then administered to the 301 subjects (students), stratified into sixteen groups, including Mexican-American vs. Anglo, age, sex, year of college, college enrolled in, etc. This quota-type sampling amounted to almost 5 percent of the undergraduate enrollment that year. In addition, a similarly stratified sub-sample of 30 bilinguals was taken, who, in addition to the above, were subjected to oral and written linguistic elicitation as Part B of the questionnaire. The basic questionnaire, with its 106 items, mostly of true-false type, covered demography; self-reports of proficiency in Spanish vs. English, as well as domains of usage; and attitudes to the respective languages and cultures, as well as questions on work ethic, life goals, life styles, etc.

Results of the survey have been described in detail through publications by different team members and by appearing in linguistic, social science and education journals. (References to some of these include: Murray, 1972; Dubois, 1979; Goodman and Brooks, 1975; Goodman and Renner, 1971; Brooks and Calkins, 1981; Ornstein 1971, 1973, 1975, 1976, 1976b, 1979; Ornstein and Goodman, 1979; and Bowen and Ornstein, 1976.)

Depending upon local needs, a sociolinguistic research team, in addition to gathering and collating raw data in different categories, decided upon particular targets with the greatest possible relevance to bilingual/bicultural needs. Research monies were subsequently obtained. By way of example, Jaime Lujan and Cynthia Darche-Park have described such efforts in their paper "Collaborative Staff Development for Teachers of Bilingual Students with the Teacher as Researcher" (1981). They remark: "The notion that classroom teachers are incapable of doing important educational research is a bias inherent to research training in many traditional research institutions."

In the UT-El Paso survey we were mostly interested in identifying key sociolinguistic characteristics of Anglo vs. Mexican-American students in our highly bilingual enrollment. Accordingly, our products followed this orientation. One of the most valuable outcomes was a Language Situation diagram of the entire Southwest, largely applicable to the El Paso, Texas-Las Cruces, New Mexico area. (see Diagram 4).



Southwestern Language Situation (Idealized)

Another valuable type of product that can be furnished by a SOL Research team consists of inventories of variants in local-regional speech. Diagram 5, prepared by the El Paso project, shows constantly updated lists of phonological and grammatical variants in the Spanish and English of Southwest Chicanos. For pedagogic purposes, it is best to avoid too much technical jargon.

Spanish	English
1. /ch/ and [sh] alternation	1. /ch/ and /sh/ alternation
2. /r/ with retroflex interference from English	2. /b/ and /v/ alternation
3. /x/ with glottal interference from English	3. [s] as realization of /z/
4. Realization of orthographic "ll" (sia for silla)	4. Consonant cluster reduction and secondary effects
5. Epenthetic /e/ in certain environments (words with final stressed syllables ending in l, n, r, as *Isabele, comere)*	5. Realization of /O/-/ / and /t/-/ /
6. /i/ for /e/ in final position following /ch/ *(nochi)*	6. [dzh] for /y/
7. Paradigmatic leveling of /e/ to i in infinitive and finite forms of *ir* verbs *(vistir, vistimos)*	7. Realization of /iy/
8. Reduplicated plurals. *(cafeses)*	8. Realization of /ae/
9. Overriding English intonation	9. Stress mislocation, particularly in compounds with nouns serving adjectival functions
	10. Overriding Mexican Spanish intonation patterns

Diagram 5: Inventory of Variables of Southwest Spanish and English

In line with the above, this worker has been developing a sociolinguistic marking system, first presented at the 12th International Congress of Linguistics, Bolongna, and subsequently revised. As the particular bilingual classes are a battleground for what forms and usage to follow, some system of classifying standard vs. nonstandard forms in language can be very useful. Along sociolinguistic principles, every attempt is made to present usage as a set of options, depending upon situation and function, rather than attempting to establish one inflexible "unitary" form, which is unrealistic since students will revert to colloquial patterns as soon as they leave the classroom. The SOL marking system devised can be seen in Diagram 6.

Diagram 6
A Tentative Sociolinguistic Marking System

I. Marking System

Symbol (According to Dimension)	Explanation	Markedness
ST	**STANDARD**	−
fo	formal	−
pr	prescriptive	−
no	normative, often of "school variety"	−
sa	sanctioned by recognized authority (academy, etc.)	
of	official	
na	national	
re	religious	
ac	acrolect(al), higher prestige variety[2]	−
sti	stilted	−
ru	rustic, rural	+
jou	journalistic	−
NS	**NON-STANDARD**	±
st	stigmatized	+
ar	argot, calo	+
pa	patois	+
ja	jargon, specialized code	+
ba	basilect(al), lower prestige, stigmatized variety[3]	+
sl	slang	+
ca	cant	+
cr	creole	+
pi	pidgin	+
cd	contactual or bilingual dialect[4], interferential variety	+
if	interferential feature or item[5], usually not assimilated within bilingual community	+
af	assimilated feature[6] or item, integrated within bilingual community	+
lo	loan, feature or item often not part of standard language[7]	− ±
hy	hypercorrect, to point of non-standardness	+
NE	**NEUTRAL (STANDARD OR NON-STANDARD)**	±
la	language	− ±
di	dialect(al)	±
le	lect, isolect (replacing dialect?)[8]	±
va	variety	±
ve	vernacular, "unmonitored" style of speech[9]	±
sc	specialized code, e.g. teenagers, occupational	±
sty	style (write: sty + des.)	±
co	code (write: co + des.)	+

Symbol (According to Dimension)	Explanation	Markedness
re	regional, geographically determined (write first 2 letters of country or region, regist 3 of *la* or *di*, e.g., FR = France; FRE = French)	− ± +
reg	register (write: reg + des.)	+
rep	repertoire = totality of styles or "codes" of language(s) controlled by an individual	− ±
so	social	±
vu	vulgar	±
po	pornographic	±
ob	obscene	±
obs	obsolete, obsolescent	±
arc	archiac	±
hf	high-frequency item or feature	±
lf	low-frequency item or feature	+
wp	word-play usually involving transposed phonotactics, e.g. Pig Latin	− ±
ta	taboo	±
su	status uncertain, lacking further data	±
in	informal	±
un	unstable, fluctuating, neologistic, disappearing, etc.	±
ss	semantic shift, often away from ST	±
ch	changing, shifting; indicate direction with arrow	±
pe	petrified, fixed, often archaic form	±
qu	quasi	+
SI	**SITUATIONAL FACTORS OR DIMENSION**	
SES	socioeconomic status (of speaker)	
ag	age, age group (write: ag + des.)	
se	sex (write: se + des.)	
co	context (write: co + des.)	
au	audience, interlocutors (write: au + des.)	
at	atmosphere (write: at. + des.)	
is	isolating from "mainstream" (write: is + des.)	
et	ethnicity, affiliation to ethnic group (write: et + des.)	
ll	language loyalty (write: ll + des.)	
or	primarily oral usage	−
wr	primarily written usage	+
te	tempo, speed of delivery (not normal; write: te + des.)	+
no	notabilia, special circumstances or details to be noted (write: no + des.)	

Students can ultimately learn to make distinctions themselves regarding the sociolinguistic concepts of "standard" vs. "nonstandard," "formal and informal," "written and spoken" and "informal and calo." In the U.S. Southwest, the Pachuco calo, a street variety of former delinquent zoot-suiters, is often confused with Southwest Spanish and "border jargon" or "Tex-Mex." Sociolinguistic ignorance causes these distinctions to be blurred or confused when the fact is that all these styles of a language, including slang which is quite limited in actual lexical quantity, have their functions in certain contexts. The rich and colorful Pachuco calo has in fact become part of racy, informal Southwest Spanish, used for humorous or often identificational purposes.

Instead of the fruitless hammering away that there is one and only one "correct" form of Spanish, or whatever language, bilingual teachers can and should assist their pupils in perceiving the vast richness and versatility of the great culture-languages, such as Spanish and English. Children are sociolinguistically programmed from the very beginning of their existence and possibly as early as in utero. Unless much of the confusion is sorted out, the variation becomes a burden, rather than a blessing. Thus in concrete terms, for example, Southwest bilinguals can, as a classroom exercise, "sort out" with the teacher such co-occurring forms as the following and mark them sociolinguistically.

Formal	Standard (Formal & Informal)	Calo (street language)
si	si	simón
no	no	nels, chale
automovil	coche, carro (informal)	ranfla
ya partió el senor	Ya se fué el hombre	ha se borró el bato
Donde vive?	Donde vives? or Onde vives? (more inf.)	Onde víboras?* (word play)

*"snake" = SW Spanish

Elsewhere (Ornstein, 1975: 40-42) the writer describes in fuller detail the organization and functioning of regional sociolinguistic research team networks. As noted, shrinking budgets at all levels of the educational establishment make it a question of sheer survival to develop strategies for maximizing funding and individual resources. One of the features of the author's concept, which has been commented upon favorably by Cohen (1975:3-4) is that of the Sociolinguistic Research Kit, to be distributed to collaborating investigators, detailing minimal and similar targets for data collection under the following four rubrics: (1) Linguistic, (2) Socio-Attitudinal, (3) Bilingual Communication and (4) Socioeducational Data/Insights.

(1) Linguistic Data
   a. Regional and nonstandard variants of Spanish (or other language) in the fields of phonology, grammar and lexicon, according to use by different socioeconomic strata
   b. Nonstandard features of the English employed by different socioeconomic groups.
   c. Interference vs. integrated norms.

(2) Socio-Attitudinal Data
   a. Attitudes toward different varieties of regional Spanish (other) and English and degrees of "language loyalty" or commitment.
   b. Attitudes toward comparative life-style and value systems of *Hispano* and *Anglo* populations (or other).

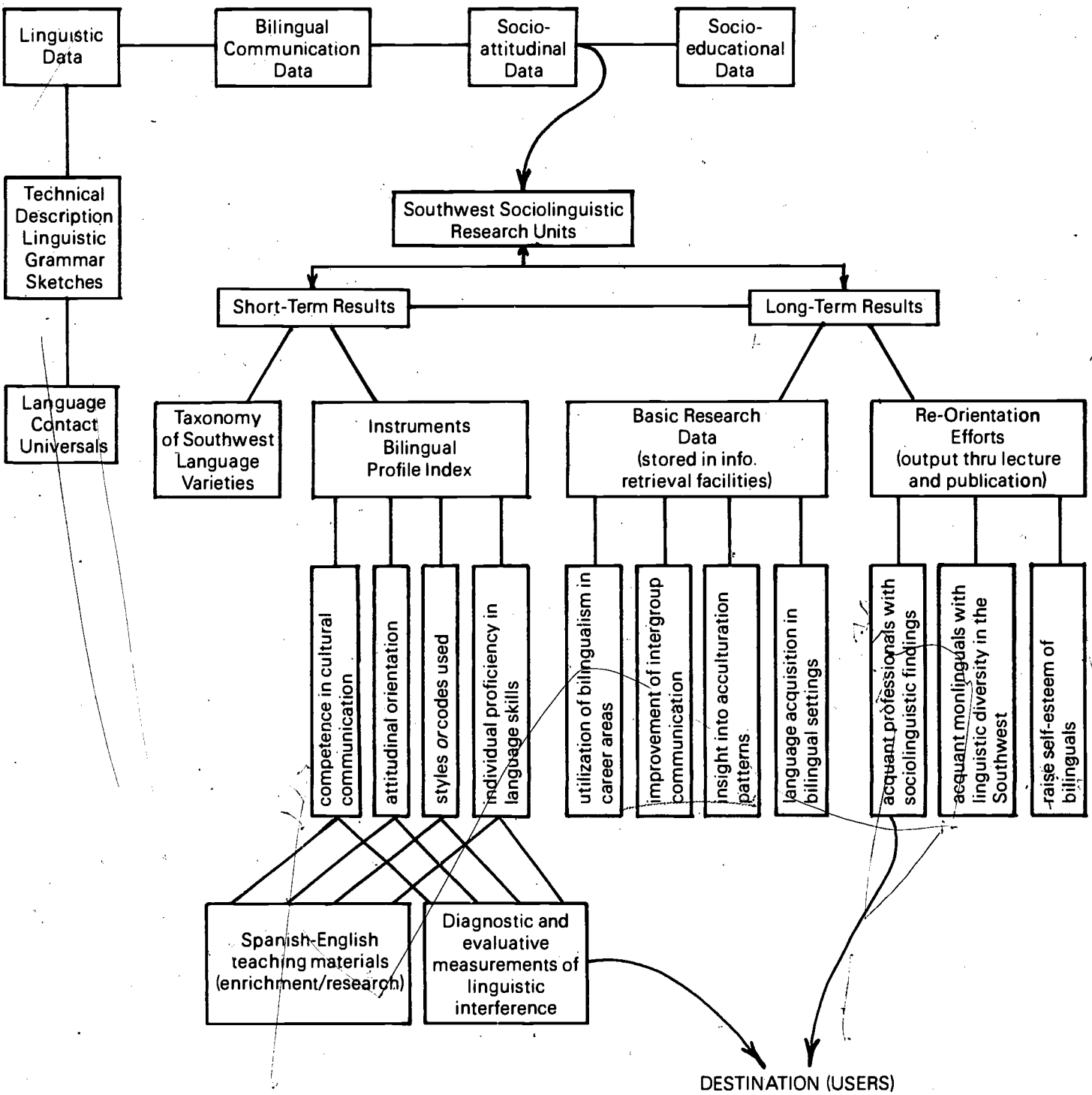
(3) Bilingual Communication Data
   a. Patterns of linguistic dominance.
   b. Distributive role of Spanish vs. English in regional communication networks.
   c. Range of codes, styles, registers employed and by whom (i.e. typical "linguistic repertoires").
   d. Code switching and stylistic shifting.

(4) Socio-Educational Data
   a. Community profiles.
   b. Language/culture attitudes of community, parents, school personnel toward aims and methods of teaching of bilingual language pair (Spanish vs. English).
   c. Testing and assessment procedures, and their special problems as applied to all types of measurements, formal and informal, and aimed at compliance of placement, exit criteria, improvement of skills and fluency improvement (proficiency).

The following diagram (7) then presents the flowchart of the regional or greater urban sociolinguistic research team network.

Diagram 7

Proposed Regional/Metropolitan Sociolinguistic
Research Team Network (U.S. Southwest)



In conclusion, the example of the Southwest U.S., an area marked by vast spaces and often limited population and resources, in pooling expertise in teaching and research may be illustrative. Since 1972, beginning at the University of Texas-El Paso and meeting in a different site each year, the *SOUTHWEST AREAL LINGUISTICS WORKSHOPS* (SWALLOW), have provided a forum for presentation of pedagogic and scholarly papers, published yearly in the *SWALLOW PROCEEDINGS*. The latest may be seen in Elerick, 1981.

No matter how tempting are the ideological disputes in our field we must, without abandoning theory, cooperate to the utmost with classroom practitioners forced to produce or apply instruments for proficiency and other aims upon short notice. As academicians, it is our obligation to provide a solid research base without which tests will result in inevitable disasters.

Summary

In line with the spirit of SOL itself, we have in the preceding pages, largely addressed the spirit of this sub-field, which is larger than the collection of facts and assumptions usually taught as "sociolinguistics." At least two notions override and pervade this study: (1) Language is inherently variable — not rigid and inflexible; and (2) Appropriateness of situation or context is a safer guide than the agonizing search for the unfailingly "correct" form. Applying SOL to assess the preparation and administration of instruments is a mighty challenging task, and utmost caution is recommended for a variety of reasons or else a top-heavy "Frankenstein" type of test whose complexity would be self-defeating will emerge. To the extent possible, instruments should be so contructed that some allowance is made for varying options. Above all, testers, bilingual teachers and administrators, should be programmed sociolinguistically to provide the surest guarantee against excessive elitism and prescriptivism. At the same time, all-out efforts should be made to utilize available grant funds and to mount sociolinguistic research teams, consisting minimally of a social scientist and a language specialist, preferably within regional team networks, addressing similar targets. An interface needs to be developed beyond its present stage, which would "package" SOL research data gleaned and whose products would be particularly aimed at the classroom teacher and tester — our first line of defense and progress in the bilingual/bicultural education movement.

*Appreciation is hereby expressed by the author to the following entities which made possible the research upon which this study was based: The Research Institute of the University of Texas, El Paso; the Hogg Foundation for Mental Health, University of Texas, Austin; the Gulf Education Fund, New York; and finally the Spencer Foundation Study Center (SWESC), which sponsored the Sociolinguistic Studies on Southwest Bilingualism (SSSB) project.

Bibliography

Allen, Harold B. and Underwood, Gary N. eds. *Readings in American Dialectology*. New York: Appleton, Century, Crofts, Inc., 1971.

Bernstein, Basil. *Class, Code and Control*. 2 Vols. London: Routledge, 1971.

Bloomfield, Leonard. *Language*. New York: Holt, Rinehart and Winston, 1933.

Bowen, J. Donald and Orstein, Jacob, eds. *Studies in Southwest Spanish*. Rowley, Mass.: Newbury House, 1976.

Brooks, Bonnie S., Brooks, Gary, Goodman, Paul W., and Ornstein, Jacob. *Sociolinguistic Background Questionnaire: A Measurement Instrument for Studying Bilingualism*. El Paso: Univ. of Texas, Southwest Ethnic Study Center, 1972. (rev. ed.)

Brooks, Bonnie S. and Calkins, Dick S. "Ability and Performance Measures of Anglo and Mexican-American Students at a Southwestern University," in Jacob Orstein-Galicia and Robert N. St. Clair, eds. *Bilingualism and Bilingual Education: New Readings and Insights*. San Antonio: Trinity University, Department of English, 1980. 55-74.

Bright, William, ed. *Sociolinguistics: Proceedings of the UCLA Sociolinguistics Conference, 1964*. The Hague: Mouton, 1966.

Burt, Marina G., Dulay, Heidi C., and Hernandez-Chavez, Eduardo. *Bilingual Syntax Measure*. New York: Psychological Corp. (757 Third Avenue), 1973.

Cedergren, Henrietta J. and Sankoff, David. "Variable Rules: Performances as a Statistical Reflection of Competence" *Language* 50, (1974), 333-55.

Chomsky, Noam. *Syntactic Structures*. The Hague: Mouton, 1957.

_____ *Aspects of a Theory of Syntax*. Cambridge, Mass.: MIT Press, 1965.

Cohen, Andrew D. "A Sociolinguistic Approach to Bilingual Education: The Measurement of Language Use and Attitudes Toward Language in School and Community, with Special Reference to the Mexican-American Community of Redwood City, Calif." M.A. Thesis, Comm. on Linguistics, Stanford University, 1970. ERIC ED 043 007.

_____ *A Sociolinguistic Approach to Bilingual Education*. Rowley, Mass.: Newbury House, 1975.

Cohen, Marcel. *Pour une sociologie de language.* Paris: Albin-Michel, 1956.

Currie, Haver and Currie, Eva. "A Projection of Sociolinguistics: the Relationships of Speech to Social Status." *Southern Speech Journal* 18, (1952), 28-37.

Dubois, Betty Lou. "Societal and Linguistic Correlates in an Investigation of the English Writing of a Selected Group of University Level Chicanos." in Wm. F. Mackey and Jacob Ornstein, eds. *Sociolinguistic Studies in Language Contact: Methods and Cases.* The Hague: Mouton, 1979. 347-62.

Duncan, Sharon A. and DeAvila, Edward A. "Bilingualism and Cognition: Some Recent Findings." *The Journal for the National Association for Bilingual Education.* Fall, 1979, 15-50.

Elerick, Charles, ed. *Proceedings of the Ninth Annual Southwestern Areal Linguistics (SWALLOW) Workshop.* El Paso: University of Texas Department of Linguistics, 1981.

Fishman, Joshua A. *Sociolinguistics: A Brief Introduction.* Rowley, Mass.: Newbury House, 1971.

_____. Cooper, Robert L. and Ma, Roxana *et* al. *Bilingualism in the Barrio.* Indiana University Publications. Language Science Monograph No. 7. The Hague: Mouton, 1972.

_____. *Advances in the Sociology of Language, I and II.* The Hague: Mouton, 1971.

Gilbert, Glenn G. and Ornstein, Jacob, eds. *Problems in Applied Educational Sociolinguistics: Readings in Language and Culture Problems of U.S. Ethnic Groups.* The Hague and New York: Mouton, 1978.

Goodman, Paul W. and Brooks, Bonnie S. "A Comparison of Anglo and Mexican-American Students Attending the Same University," *Kansas Journal of Sociology* 10, (1974), 181-203 (Fall).

Goodman, Paul W. and Denner, Kathryn. "Social Factors and Language," in Glenn G. Gilbert and Jacob Ornstein, eds. *Problems in Applied Educational Sociolinguistics: Readings in Language and Culture: Problems of U.S. Ethnic Groups.* The Hague and New York: Mouton, 1976. 55-62.

Grimshaw, Allen D. "Sociolinguistics," in J. A. Fishman, ed. *Advances in the Sociology of Language I.* The Hague: Mouton, 1971. 92-151.

Gumperz, John J. "On the Linguistic Markers of Bilingual Communication," in John Macnamara, ed. *Bilingualism in the Modern World (Special Issue) Journal of Social Issues,* 23, (1967), 48-57 (2).

_____ and Hymes, Dell, eds. 1964. "Ethnography of Communication." *American Anthropologist* 66, No. 6, (1964), Pt. II.

_____ eds. *Directions in Sociolinguistics: The Ethnography of Communication.* New York: Holt, Rinehart and Winston, 1972.

Haugen, Einar. *Bilingualism in the Americas: A Bibliography and Research Guide.* Alabama: University of Alabama Press, 1956.

_____. *The Norwegian Language in America: A Study in Bilingual Behavior, 2 vols.* Bloomington: Indiana University Press, 1969.

Hernández-Chavez, Eduardo. "Critique of a Critique: Issues in Language Assessment," *NABE (National Association for Bilingual Educators Journal.* 2, (1978), 47-55.

Hertzler, Joyce O. "Toward a Sociology of Language," *Social Forces.* 32, (1953), 109.

_____. *A Sociology of Language.* New York: Random House, 1965.

Hymes, Dell. "Models of the Interaction of Language and Social Setting," *Journal of Social Issues* 23, (1967), 48-57(2).

Jorstad, Helen A. "Testing as Communication," in G. A. Jarvis, ed. *The Challenge of Communication.* ACTFL Review of Foreign Language Education, Vol. 6, Skokie, Illinois: National Textbook Co. 1974. 223-274.

Kelly, L. G., ed. *The Description and Measurement of Bilingualism.* An International Seminar. University of Moncton (N.B.), June 6-14, 1967. Toronto: University of Toronto Press, 1969.

Labov, William A. *The Social Stratification of English in New York City.* Washington, D.C.: Center for Applied Linguistics, 1966.

_____. "The Social Motivation of a Sound Change," *Word.* 19, (1963), 273-309.

_____. "The Study of Language in Its Social Context," *Studium Generale* 23, (1970), 30-87 (Reprinted in Fishman, ed. *Advances in the Sociology of Language I,* The Hague: Mouton, 1971. 152-216).

Lebrun, Yvan. "The Neurology of Bilingualism," in W. von Raffler-Engel, ed. *Child Language 1975.* (Equiv. to *Word,* Vol. 27, 1976)

Lieberson, Stanley, ed. *Explorations in Sociolinguistics.* Bloomington: Indiana University, 1966.

Luján, Jaime and Darché-Park, Cynthia. "Collaborative Staff Development for the Teacher as Researcher," *Ethnoperspectives Forum: Bilingual Education Technology,* ed. Raymond Padilla. Ypsilanti, Michigan: Eastern Michigan University, 1981.

Malherbe, E. C. "Discussion of Paper by R. N. Jones;" in L. G. Kelly, ed. *The Description and Measurement of Bilingualism*. Toronto: University of Toronto Press, 1969. 41-50.

Murphy, R. Paul and Ornstein, Jacob. "A Survey of Research on Language Diversity: A Partial Who's Who in Linguistics." Peter A. Reich, ed. *Second LACUS (Language Association of Canada and U.S. Forum 1975*. Columbia, S.C.: Hornbeam Press, 1976. 423-61.

Murray, Wayne. "Ethnic and Sex Differences as Related to Perceptions of a University Environment" (Dissertation) Las Cruces, New Mexico State University, 1972.

Naro, Anthony J. 1981. "The Social and Structural Dimensions of a Syntactic Change." *Language*. 57, 1, (1981), 63-98 (March).

Northwest Regional Educational Laboratory. *Assessment Instruments in Bilingual Education: A Descriptive Catalog of 342 Oral and Written Tests*. Los Angeles, Ca.: National Dissemination and Assessment Center, 1978.

Ornstein, Jacob. "Sociolinguistic Investigation of Language Diversity in the U.S. Southwest and Its Educational Implications." *Modern Language Journal*. 55, (1971) (April), 224-229.

_____. "The Sociolinguistic Studies on Southwest Bilingualism (SSSB) — A Status Report," in *Southwest Areal Linguistics*. Garland Bills, ed. San Diego, California: Institute for Cultural Pluralism, University of Califorinia, 1973.

_____. "Sociolinguistics and the Study of Spanish and English Language Varieties and Their Use in the U.S. Southwest (with a Proposed Plan of Research)." in *Three Essays on Linguistic Diversity in the Spanish-Speaking World*. Jacob Ornstein, ed. The Hague: Mouton, 1975. 9-46.

_____. *A Sociolinguistic Study of Mexican-American and Anglo Students in a Border University*. (Occas. Paper No. 3, ed. Will Kennedy) San Diego: San Diego State University, Institute of Public and Urban Affairs Press, 1976.

_____. "A Cross-Cultural Sociolinguistic Investigation of Mexican-American Bilinguals/Biculturals at a U.S. Border University." *La Linguistique* (Paris) 12, (1976), 131-45.

_____. "Relational Bilingualism: A Socioeducational Approach to Studying Multilingualism among Mexican-Americans," in *Language and Society*, Wm. C. McCormack and S. A. Wurm, eds. The Hague: Mouton, 1979. 285-305. (another version, "Relational Bilingualism — A New Approach to Linguistic Cultural Diversity and a Mexican-American Case Study," in *Ethnicity* 5, (1978), 148-66.

Sanchez, Rosaura. "Language Theory Bases," in R. Sanchez et al. *Language Proficiency Assessment*. (National Origin Desegregation: LAU Center). San Diego: San Diego State University Institute for Cultural Pluralism, 1978. 1-63.

_____, Romo, Harriett; Rivera, Iris S. and Williams, Byron. *Language Proficiency Assessment*. San Diego: Institute for Cultural Pluralism, 1978.

Seidner, Stan. "Language Assessment at Post Secondary Institutions: Models for Establishing Entry-Exit Level Criteria." *Ethnoperspectives Forum: Bilingual Education Technology*. Raymond Padilla, ed. Ypsilanti, Michigan: Eastern Michigan University, 1981.

Siertsema, Berthe. "Discussion of Paper by Nils Hasselsno." in L. Kelly, ed. *The Description and Measurement of Bilingualism*. Toronto: University of Toronto Press, 1969. 152-55.

Stansfeld, Charles. "The Assessment of Language Proficiency in Bilingual Children: An Analysis of Theories and Instrumentation." *Ethnoperspectives Forum: Bilingual Education Technology*. Raymond Padilla, ed. Ypsilanti, Michigan: Eastern Michigan University, 1981.

Weinreich, Uriel. *Languages in Contact*. New York: Linguistic Circle of New York, 1953.

Williams, Frederick, Hopper, Robert and Natalicio, Diana. *The Sounds of Children*. Englewood Cliffs, N.J.: Prentice-Hall, 1977.

SOCIOLINGUISTIC FOUNDATIONS OF LANGUAGE ASSESSMENT: A REACTION

by
Guillermo Duron
Governors State University

The purpose of this paper is to react to the material presented by Dr. Jack Ornstein-Galicia relative to the sociolinguistic foundations of language assessment. Dr. Ornstein-Galicia's presentation was organized into three areas. The first dealt with a historical perspective of sociolinguistics. This section included cumulative developmental definitions. The second section dealt with the theories underlying sociolinguistics. Included in this section are the ideas of the contemporary theorists from which the discipline of sociolingustics has evolved. The final section of Dr. Ornstein-Galicia's presentation dealt with the contribution of sociolinguistics to language assessment. Specifically addressed were issues related to bilingual bicultural education and ethnic studies.

According to Dr. Ornstein-Galicia, sociolinguistics is a sub-branch of both linguistics and sociology. It has been called the sociology of language. Furthermore, sociolinguistics is viewed as using sociological methods and concepts to analyze the linguistic patterns of groups of individuals. Sociolinguistics assumes that language can only be understood and have meaning when it is examined in light of the social groups utilizing it. Language, therefore, is not fixed in time and space; it is dynamic and ever-changing. The variability of language is the subject of sociolinguistics. Noteworthy is the fact that over 5,000 languages and over 1,000 dialects are spoken.

Historically, sociolinguistics evolved for pragmatic reasons. Sociolinguistics came into focus as a result of the increase in the number of people who travel extensively and the need for communication and understanding for professional reasons. Sociolinguistics tries to negate the idea of language as an elite concept, that is, it avoids the conclusion that there is a "pure" language. In rejecting the elitist point of view which concludes that there is only one correct variation of language, it has introduced the concept that appropriateness of language is relative. Since the purpose of language is communication, it must be situation-specific. An example of this is demonstrated by the following which depicts a situation and an inappropriate linguistic vehicle to accompany it.

Situation	Inappropriate Linguistic Vehicle
• In a bar in the ghetto	• King's English
• In the barrio	• Academia Real Spanish
• Job Interview	• Vernacular Speech

While any of the above linguistic vehicles are appropriate within a given situation, they may be deemed inappropriate based on the specific contextual environment listed above. This illustrates the situation-specificity of language.

Several major sociolinguists were referenced by Dr Ornstein-Galicia as contributing their ideas, philosophies and theories to the emerging field of study. Fishman defined sociolinguistics as the structure that tells us who speaks which language to whom, when, why, and where. This theorist described sociolinguistics as responding to a need for an organized discipline to replace a vacuum that was occuring in the area of linguistics. Due to the influence of philosophy, transformational grammar, etc., linguistics was becoming a science. Other sociolinguists mentioned included "purists" such as Chomsky and Haugen and "interpreters" such as Blumefield and Brite.

Considering the cultural milieu that existed while sociolinguistics was developing, the time was right for its conception. In 1960, Title VII of the Elementary and Secondary Education Act was established; in 1964, the Civil Rights movement was underway; and in 1974, *Lau vs. Nichols* called for equal educational opportunity for students from non-English backgrounds. Legislation occuring at the time was favorable toward linguistic diversity. Sociolinguistics developed as the nod was given to researching "standard" as well as "non-standard" languages in order to interpret their cultural significance. The results of this research were ultimately used to provide insight and information to educators of minority and non-minority students.

Sociolinguistics, according to Dr. Ornstein-Galacia, has contributed significantly to language assessment. As a field of study, sociolinguistics has softened the outrageous demands of some English teachers. These teachers were guilty of imposing elite standards of language regardless of the cultural or contextual milieu. Awareness of language and its various roles have resulted from sociolinguistics. Accordingly, in terms of assessment, lesser ability is not equated with nonstandard speech or language patterns. Sociolinguistics has made teachers and administrators aware of bilingual education as a meaningful, effective, instructional vehicle.

As a result of the influence of sociolinguistics, options are available to producers and consumers of language. There is not one unitary, monolithic language. Sociolinguistics provides a means to look at other disciplines and allows for a point of view that legitimizes those options. Unfortunately, a problem exists when language assessment utilizes a sociolinguistic perspective. Because no two people agree on what should be assessed sociolinguistically, the result is global rather than specific instrumentation. Practicality demands that much of the variation in the population be ignored. As such, tests cannot be profitably made and marketed that take into consideration the experiences of all sub-groups. However, based on the observations of sociolinguists, the results of currently existing instruments should be carefully considered in light of the linguistic and cultural experiences of the examinee.

LANGUAGE PROFICIENCY ASSESSMENT: RESEARCH FINDINGS AND THEIR APPLICATION*

Charlene Rivera
InterAmerica
Research Associates
Rosslyn, Virginia

Carmen Simich
InterAmerica
Research Associates/
Georgetown University
Washington, D. C.

Development of language proficiency assessment instruments has been greatly influenced by the various interpretations of linguistic theories and their application. Assessment measures developed in the 1950's and 1960's were greatly influenced by structural linguistic views of language and psychometric methodology which prompted the testing of discrete aspects of language. These instruments were generally intended for adults learning a foreign language and, as such, assumed literacy in the native language.

During the 1970's, the influence of psycholinguistics was reflected in the attempt to incorporate generative linguistic theory into language proficiency testing. The concept that "the reliable variance in a great variety of educational and psychological tests can be attributed to a single global factor of language proficiency" (Oller, 1979, p. 61) motivated the development of tests which "must invoke the expectancy system or grammar of the examinee" (p. 16). Oller posits that tests which reflect the pragmatic perspective such as dictation, cloze and their variations "probably provide more accurate information concerning language proficiency. . .than the more familiar tests produced on the basis of discrete point theory" (1979, p. 9).

The sociolinguistic/ethnographic perspective entailing "the notion that children's school language should be viewed within a broader framework of culturally acquired communicative competence" (Philips, 1980, pp. 2-3) has recently begun to influence language proficiency assessment practices. Methodologically, this approach implies the focused observation of students' language use in naturalistic settings. In contrast to the traditional approaches to testing, it does not generally rely on the paper and pencil type of tests which can be statistically analyzed.

While traditional psychometric approaches to testing are generally used for purposes of identification and placement, they are recognized to be inadequate. Thus, the controversy remains as to the nature of language and how to best measure it. Issues which have not been adequately addressed by traditional testing procedures include such basic questions as:

- What does it mean to be proficient?

  Does it mean a person's receptive/expressive knowledge of discrete grammatical components of a language?

  Does it mean a person's knowledge of linguistic code(s) and its (their) appropriate use in different social contexts? (Note 1)

- What are the variables which influence language use during communicative interactions?

- How should these variables be incorporated into the development of language proficiency measures?

Current research which has implications for language proficiency assessment practices includes research in adult language proficiency testing, cognitive studies which attempt to conceptualize the construct of language proficiency, development of theoretical models of communicative competence, studies investigating the validity of measures of communicative competence, and ethnographic/sociolinguistic studies of children's functional use of language. Representative research in each area will be reviewed.

Research efforts in adult language proficiency testing are represented by the work of the Foreign Service Institute (FSI) and the Educational Testing Service (ETS). These efforts are discussed because of their potential application to the assessment of language minority students.

The Foreign Service Institute (FSI) has played a significant role in the development of an oral interview testing system to assess foreign language proficiency. The system has evolved over the past thirty years. The significance of this effort lies in the standardization of oral testing procedures, the aspects of performance that are to be observed, and their rating. While the FSI Oral Interview Test represents a positive effort in the measurement of oral language proficiency of adult foreign language learners, the system is limited in its ability to

*This paper was prepared as part of Assessing the Language Proficiency of Bilingual Persons Project (ALPBP). It is funded by the National Institute of Education (NIE 400-79-9942) and administered by InterAmerica Research Associates, Rosslyn, Virginia. The opinions of the contributors are their own and do not reflect those of the National Institute of Education.

measure the effectiveness of the communication process within cultural contexts (Rice, 1959). Currently efforts are being made to emphasize "more realistic uses of language" (Jones, Note 2), with a greater stress being placed on the comprehensive ability of individuals being tested. The goal of the FSI Oral Interview Test is to predict government employees' successful use of the target language in overseas assignments. Although designed to assess adults' foreign language proficiency, it has the potential to be adapted to assess the language proficiency of school age LEP students.

Clark (1980) proposes the development of a "common measure" of speaking proficiency in second and foreign languages for use with high-school and adult learners. He defines "common measure" as "...a uniform testing procedure that can be used with diverse groups of examinees in a variety of language learning situations with testing results reported on a single uniform scale" (p. 15). He argues that there is a need for development of a measure of speaking proficiency that can be validly used in a series of different situations for which development of specific procedures is not realistic. In an update on the development of such a measure, Clark (1981) emphasizes its benefits and potential use:

- to further the development of a more sophisticiated measurement procedure;
- to make available a cost and time effective instrument which evaluates communicative proficiency in different situations and for different second language learners;
- to provide a highly valid and reliable instrument for use in the validation of other tests of speaking proficiency.

Cummins suggests that the present state of the art in language proficiency assessment is confused because of the "failure to develop an adequate theoretical framework for relating language proficiency to academic achievement" (1981). He contends that:

> there has been a failure to adequately conceptualize the construct of language proficiency and its cross-lingual dimensions. In other words, there has been relatively little inquiry into what forms of language proficiency are related to the development of literacy skills in school contexts, and how the development of literate proficiency in L_1 relates to the development of literate proficiency in L_2.

He posits that there are two dimensions to language proficiency: cognitive-academic language proficiency (CALP) and basic interpersonal communication skill (BICS). CALP refers "to the dimension of language proficiency that is strongly related to literacy skills. BICS refers to cognitively undemanding manifestations of language in interpresonal situations" (Cummins, 1980, p. 28). He hypothesizes that these two dimensions of language can be empirically distinguished and that the native (L_1) and second language (L_2) CALP-like skills are manifestations of the same underlying dimension. Based on these hypotheses, which are currently being investigated (Cummins, et al, Note 2), Cummins suggests that "placement of bilingual children in different types of instructional programs should *not* be based only on 'natural communication' (BICS) tasks (but that) developmental levels of L_1 and L_2 CALP should also be taken into account" (1980, p. 54). Thus, he strongly recommends that students' literacy skills be tested in both L_1 and L_2 before placement/exit decisions are made.

Sociolinguistic theory and research in teaching second language learners has brought the concept of communicative competence (Hymes, 1974) into prominence. While many interpretations have been given to this concept, in general, it is used to refer to mastery of communicative skills acquired by second language learners and the appropriate use of these skills during social interactions. Two models based on this concept are presented here because they provide a frame of reference for ongoing developments in language proficiency assessment.

Canale and Swain, (1979) suggest a model which is based on the identification of features considered important for communicating. These communication features are characterized as being interaction-based, unpredictable, creative and purposive. Communication is authentic, rather than contrived, and takes place within sociolinguistic and discourse contexts. Successful communication is judged on the basis of actual outcome. The three components of the communicative competence model are: linguistic, sociolinguistic, and strategic competencies. Linguistic competence deals with mastery of the grammar of a language. Sociolinguistic competence involves mastery of appropriate language use with an emphasis on meaning and the appropriateness of the linguistic forms used to convey meaning. Strategic competence refers to second language learners' ability to compensate, repair and use other strategies in their attempts to communicate with other participants.

Research by Slaughter and Bennett (1981) into the nature of discourse of bilingual children expands on Swain and Canale's discourse component. Bennett (1981) describes discourse as being temporal, reflexive, multivocal and multimodal. It is temporal in that it evolves through time. It is reflexive in that language creates the context for it to be understood. It is multivocal in that different meanings and interpretations are always available to participants, and it is multimodal in that it involves the selection of varied modalities such as choices of grammar, lexicon and nonverbal behaviors.

Briere (1979) developed a model which recognizes sociolinguistic and linguistic competence as components of communicative proficiency. These competencies have two dimensions: one at the abstract level, or linguistic/sociolinguistic competence, and the other at the performance level, or linguistic/sociolinguistic performance. At the performance level, communicative proficiency is associated with the speaker's use of the grammar of the language in appropriate social interactions.

The relevance of both Canale and Swain's and Briere's attempts to develop models of communicative proficiency is that they provide researchers and practitioners with a comprehensive framework. The models suggest that language proficiency tests should consider knowledge of the linguistic code concurrently with appropriateness of its use during social interactions.

Bachman and Palmer (1979,1981) investigated the construct validation of language tests in a two-phase study. They explored the validity of a simple model of language which consisted of two traits — speaking and reading. They found that the model statistically explained test results better than a single factor model (Oller, 1979). Encouraged by the findings, they expanded their research to investigate the construct validity of several other models of language. Using confirmatory factor analysis, they found that the model which showed the highest degree of statistical significance was a model that posited one single factor and three specific trait factors: linguistic, sociolinguistic and pragmatic. The model was developed based on Canale and Swain's (1979) model of communicative competence.

Ethnographic and sociolinguistic studies of children in a New York Puerto Rican community (Poplack, 1981) add to our understanding of the evolution and maintenance of language in stable bilingual speech communities. The role of code-switching of bilingual adults and children investigated from an intergenerational perspective gives insights into the acquisition and use of code-switching and its relationship to language proficiency.

Rodriguez-Brown and Elias-Olivares (1981), in investigating the communicative competence of bilingual children, concentrated on the expressive strategies used to make inquiries. Their findings indicate the importance of focusing on bilingual children's language use in experimental test situations.

Simich and Rivera's work in cooperation with teachers and specialists from Tucson Unified School District, Tucson, Arizona, represents a preliminary effort to identify functional uses of language in elementary bilingual and monolingual classrooms. This work attempts to clarify the reasons why children and other participants communicate and how children's knowledge of acceptable functional uses of language relates to language proficiency.

Sociolinguistic research highlights the difficulty of measuring language via discrete quantifiable means. Shuy points out that the most critical aspects of language "are the ones least susceptible to quantification" (1977, p. 77-78). In general, instruments of language proficiency measure easily quantifiable components of language such as pronunciation, vocabulary, and grammar, rather than those less visible or quantifiable, such as meaning or the functional intent of utterances. Use of discrete point language assessment instruments does not provide valid and accurate information about how effectively students participate in instructional settings. Thus, the sociolinguistic approach promotes evaluation of the appropriate use of language in different communicative situations. Its contribution to the development of language proficiency assessment instruments is based on a broader interpretation of language which includes the use of linguistic code(s) by participants in ways which are acceptable to other members of a speech community.

Studies in language variation have expanded the idea of the linguistic repertoire of speech communities (Gumperz, 1972). Applied to language proficiency assessment, this concept supports the view that minority students' language proficiency should not be measured against the "standard" dialect of a language. Rather, a student's way of speaking should be considered adequate and appropriate in terms of the purposes it serves during communication. For example, in the southwest, where large numbers of Hispanic students live, code-switching incidences are common. This manner of communication should be considered appropriate given its functional use within the community.

If we take the position that students' language should not be measured against a standard dialect, then we need to ask: why measure language proficiency?

The measurement of language proficiency is necessary to provide teachers with an understanding of the language skills students have already acquired in the home environment. A comparison between home language skills and functional language demands in the classroom setting will provide necessary information upon which to place students in appropriate educational programs.

More recent sociolinguistic research relates functional language use to language proficiency. Shuy defines functional language use as "the underlying knowledge that allows people to make utterances in order to accomplish goals and to understand the utterances of 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)" (1977, p. 79). Several researchers in the last decade have investigated functional language use in school settings. The focus of their research has been on specific aspects of why and how children use language in different social contexts (e.g., Cahir, 1978; Cazden, 1979; Simich, 1980; Jacob, 1981). One outcome of the research on functional language use in the classroom has been the development of several sociolinguistically based language assessment instruments.

For example, Shuy, McCreedy and Adger (1979) developed an oral language assessment instrument for use with elementary school children who are speakers of vernacular Black English (VBE). The instrument consists of three components. The first provides for approximate measurement of phonological and morphosyntactic features. The second component evaluates communicative competence according to relative appropriateness and strategies children use for conversational functions such as explaining, describing, etc. The third component evaluates discourse abilities, such as appropriateness of interrupting, use of transitional markers, referencing and style shifting.

Another effort toward the development of sociolinguistic/ethnographic measures is represented in the work of Simich and Rivera (1981) in cooperation with Sunnyside and Tucson Unified School Districts (TUSD), in Arizona. An instrument, entitled the *Teacher Observation System (TOS)*, was developed during a comprehensive two-year teacher training program. The goal of the program was to provide bilingual educators with background in linguistics, sociolinguistics, ethnographic methodologies, measurement, and research methodologies in order to enable them to develop accurate and effective language proficiency assessment strategies. The *TOS* was developed based on the understanding that the range of students' language repertoire can only be determined through systematic and focused observations of interactions in a variety of school settings. A framework to identify the variables that influence students' communicative interactions was developed, based on certain components of speech events (Gumperz & Hymes, 1972). This framework is described in Table I.

TABLE I

Components of Instructional Events to Be Considered in the Development of the Teacher Observation System

Setting	Particpants	Channels of Communication	Language Used	Discourse Characteristics
Instructional (formal) vs. Non-instructional (informal) events	Teacher/Student(s) Student/Student(s)	speaking listening reading writing	English Spanish	coherence complexity adequacy of vocabulary code-switching

The following questions helped participants focus on an ethnographic/sociolinguistic perspective during the development of the *TOS*.

- What kind of communicative skills do bilingual students need to master in order to participate appropriately as members of the sociocultural school environment?
- In which sociocultural situation can these communicative skills be observed?
- What kinds of communicative skills do students bring to school?
- In which language(s) and sociocultural situations do students have the widest contextual range of communicative abilities?

The questions prompted a discussion of factors that influence students' communicative interactions. The components considered were: classroom organization (teacher-centered vs. student-centered), language of instruction, directness or indirectness of "teacher-talk" and students' language use and their sociocultural background; parents' socioeconomic and educational background, number of siblings, and language use at home, school and community.

Through the process of developing the *TOS*, teachers became aware of the sociocultural aspects of language as they acquired more sensitivity and understanding of language use in multicultural/multilingual school settings. This knowledge, they confirmed, assisted them in the classroom situation to make better judgments about their students' communicative proficiency.

Slaughter and Bennett (1981) are attempting to develop a unified framework for the analysis of discourse samples elicited from Spanish/English bilingual students. The framework will be used in the validation of the *Language Proficiency Measure (LPM)*. The instrument developed by the TUSD purports to measure the language proficiency of K-12 English/Spanish bilingual students based on samples of discourse elicited in an experimental setting.

In summary, theoretical and applied research from the psycholinguistic and sociolinguistic perspective is providing basic information that is proving useful in the further exploration of language proficiency assessment practices. The research cited is intended to be representative of both traditional and nontraditional approaches to interpretations of the nature of language, language proficiency and language proficiency assessment.

Reference Notes

1. Social context refers to cultural/physical contexts in which communicative events take place. The social context is an important variable which influences communicative outcomes. In school settings, both academic and nonacademic events are common. Examples of social contexts range from math class to recess.
2. Jones, D. *Personal Communication*. Foreign Service Institute, Rosslyn, VA, June 17, 1981.

References

Bachman, L. F. & Palmer, A. S. "Convergent and Discriminant Validation of Oral Language Proficiency Tests." In R. Silverstein (Ed.), *Occasional Papers on Linguistics Number 6, Proceedings of the Third International Conference on Frontiers in Language Proficiency and Dominance Testing.* Carbondale, IL: Southern Illinois University, 1979.

Bachman, L. & Palmer, A. S. "Some Comments on the Terminology of Language Testing." Paper presented at the Language Proficiency Assessment Symposium. Rosslyn, VA: InterAmerica Research Associates, March, 1981.

Bennett, A. "Reaction to M. Canale, A Communicative Approach to Language Proficiency Assessment in a Minority Setting." Paper presented at the Language Proficiency Assessment Symposium. Rosslyn, VA: InterAmerica Research Associates, March, 1981.

Briere, E. "Testing Communicative Language Proficiency." In R. Silverstein (Ed.), *Occasional Papers on Linguistics Number 6, Proceedings of the Third International Conference on Frontiers in Language Proficiency Testing.* Carbondale, IL: Southern Illinois University, 1979.

Canale, M. & Swain, M. "Communicative Approaches to Second Language Teaching and Testing." *Review & Evaluation Bulletin.* Ontario Institute for Studies in Education, 1979, *1*(5).

Cahir, S. "Activity between and within Activities: Transition." In *Children's Functional Language and Education in Early Years.* A Final Report to the Carnegie Corporation of New York, 1978.

Cazden, D. "Language in Education: Variation in the Teacher-Talk Register." In J. A. Alatis, & G. R. Tucker (Eds.), *Language in Public Life.* Washington, D.C.: Georgetown University Press, 1979.

Clark, J. L. D. "Toward a Common Measure of Speaking Proficiency." In J. R. Fritch (Ed.), *Measuring Spoken Language Proficiency.* Washington, D.C.: Georgetown University Press, 1980.

Clark, J. "Toward a Common Measure: Status Report on the "Common Yardstick" Project." Paper presented at the Pre-conference on Oral Proficiency Assessment. Preceding the Georgetown University Roundtable on Languages and Linguistics, 1981. (In press)

Cummins, J. "The Entry and Exit Fallacy in Bilingual Education." *NABE Journal, 3,* (1980), 25-59.

Cummins, J. "Wanted: A Theoretical Framework for Relating Language Proficiency to Academic Achievement among Bilingual Students." Paper presented at the Language Proficiency Assessment Symposium. Rosslyn, VA: InterAmerica Research Associates, March, 1981.

Gumperz, J. "The Speech Community." In P. P. Giglioli (Ed.), *Language and Social Context.* Great Britain: Cox and Wyman, 1972.

Gumperz, J. J. & Hymes, D. (Eds.), *Directions in Sociolinguistics: The Ethnography of Communication.* New York: Holt, Rinehart & Winston, Inc., 1972.

Hymes, D. *Foundations in Sociolinguistics.* Philadelphia: University of Pennsylvania Press, 1974.

Jacob, E. "Puerto Rican Children's Literacy Learning at Home." Paper presented at the Language Proficiency Assessment Symposium. Rosslyn, VA: InterAmerica Research Associates. March, 1981.

Oller, J. W. *Language Tests at School.* London: Longman Group, Ltd., 1979.

Philips, S. "A Course on Bilingual Language Proficiency Assessment." Report to the National Institute of Education and InterAmerica Research Associates, Rosslyn, VA, 1980. (Unpublished Manuscript)

Poplack, S. "Transcending Testing in Assessing Communicative Competence." Paper presented at the Language Proficiency Assessment Symposium. Rosslyn, VA: InterAmerica Research Associates, March, 1981.

Rice, R. "The Foreign Service Institute Tests Language Proficiency." *Linguistic Reporter,* May, 1959.

Rodriguez-Brown, F., & Elias-Olivares, L. "On Assessing Language Proficiency: Linguistic Repertoires, Communicative Competence and the Hispanic Child." Paper presented at the Language Proficiency Assessment Symposium. Rosslyn, VA: InterAmerica Research Associates, March 1981.

Shuy, R. "Quantitative Language Data: A Case For and Some Warnings Against." *Anthropology in Education Quarterly 8*(2), (1977), pp. 73-82.

Shuy, McGreedy & Adger. "Oral Language Assessment Instrument." Washington, D.C.: Center for Applied Linguistics, 1979. (Unpublished Manuscript)

Simich, C. "Differential Use of Verbal and Nonverbal Strategies of Control Used by Spanish-Dominant and English-Dominant Teachers in a Bilingual Elementary School." Paper presented at the American Anthropological Association Conference, Washington, D.C., 1980.

Simich, C. & Rivera, C. "The Development of a Teacher Observation Instrument: An Ethnographic/Sociolinguistic Approach." Paper presented at the Language Assessment Symposium. Rosslyn, VA: InterAmerica Research Associates, March, 1981.

Slaughter, J. & Bennett, A. "A Sociolinguistic/Discourse Approach to the Description of the Communicative Competence of Linguistic Minority Children." Paper presented at the Language Proficiency Assessment Symposium. Rosslyn, VA: InterAmerica Research Associates, March, 1981.

Part II
Assessment Approaches

THE ROLE OF GRAMMAR IN A COMMUNICATIVE APPROACH TO SECOND LANGUAGE TEACHING AND TESTING*

by
Merrill Swain and Michael Canale

Ontario Institute for Studies in Education
Toronto, Ontario

During the past few years we have been giving some thought to the feasibility of developing assessment instruments which are consistent with a communicative approach to second language instruction. In doing this, it has been essential to come to grips with what is meant by a communicative approach. In general, a communicative approach implies a fundamental shift of emphasis away from language teaching through form to language teaching through meaning. The danger in this shift is that little place may be left for the teaching of form, that is, the grammatical aspects of the language. Morrow (1977) has commented in relation to communicative approaches to second language teaching and testing that "the polarization which can be detected in some quarters between a concentration on *form* and a concentration on *meaning* must be resisted" (p. 11). It is certainly the case that an examination of current literature on communicative competence and performance, of pedagogical materials and teaching techniques which claim to be directed towards the development of communicative skills, and of testing instruments which claim to be measuring the knowledge and use of those skills suggests that there is less than terminological consensus about what is meant by a communicative approach. The importance of reaching a consensus must not be underrated, both for purposes of program design and program evaluation. Without it, for example, competing claims for the effectiveness of different instructional programs in developing communicative abilities cannot be assessed (Canale and Swain, 1979).

Our own consideration of this matter has been significantly influenced by the target population with whom the assessment instruments will be used. This population consists of English-speaking students in Ontario at the grades six and ten levels who will have had a *maximum* number of accumulated hours of French as a second language (FSL) instruction of 540 and 720[1] at the end of their respective school years. Thus our target population consists of school-aged children who are at a relatively early stage of second language learning. It is our opinion that a de-emphasis on the importance of grammar in some of the current discussions about communicative approaches to second language instruction derives from the fact that the target populations under consideration differ from the ones we are concerned with in that they have already achieved advanced levels of grammatical competence in the second language.

In our attempts to outline the essential elements of a communicative approach for general programs of second language instruction, we initially focussed on the literature related to communicative competence. The purpose of this paper is to discuss several different notions of what is meant by communicative competence, arguing ultimately for a framework which includes three components: grammatical competence, sociolinguistic competence, and strategic competence.[2] Several issues raised by this proposed framework for second language teaching and testing will then be discussed. For a full discussion of these and related issues, see Canale and Swain (1980).

It is important at the outset to note that by being concerned with communicative competence, we are not assuming that it is the highest or broadest level of language competence that can be distinguished or that is relevant for second language teaching purposes. The assumption that communication is the essential purpose of language is widespread (see, for example, Campbell and Wales, 1970; Groot, 1975; Munby, 1978) and would seem to imply that communicative competence is the most inclusive language competence. However, as has been pointed out (see, for example, Chomsky, 1975; Fraser, 1974; Halliday, 1978), there is little reason to view *externally* oriented communication as more essential than other purposes of language such as self-expression, problem solving and creative writing. Nonetheless, the communicative purpose would seem to be the most practical concern for a general second language program. For a bilingual education program, however, focussing only on the communicative function of language would appear to be too limited.

Another general point needs to be made at the outset concerning the distinction between competence as knowledge of a system of rules and performance as the actual use of the rules in communicative acts. We have maintained this distinction despite the obvious statement that all observed behaviour is performance. The distinction is significant not only with respect to classroom activities, but to testing as well. For example, the distinction suggests that communicative testing must be devoted not only to what the learners know about the second language and about how to use it, but also to what extent the learner is able to actually demonstrate this knowledge in meaningful communica-

*Invited submission to the Institute. Originally presented as an unpublished paper at the TESOL Summer Institute, U.C.L.A., July, 1979

tive situations. The former suggests the use of competence-oriented tasks with the possibility of discrete-point items for diagnostic purposes. The latter suggests the use of performance-oriented tasks which conform more directly to normal language use in which an integration of these skills is required with little time to reflect on and monitor language input and output (i.e. Clark's 'direct' tests of language proficiency). One would not want to ignore performance tasks completely in a communicative testing program even if more competence-oriented tests that correlated highly with actual performance were developed. As John Clark (1972) states:

> Indirect tests of proficiency do not provide an opportunity for the student to try out his language competence in realistic communicative situations. Although they may correspond in a statistical sense to direct tests of proficiency, paper-and-pencil tests, tape-recorded listening and speaking tests, and similar measures cannot have the same psychological value for the student or the same instructional impact. For this reason alone, administration of a direct test of communicative proficiency at one or more points in the student's language-learning career would be a very worthwhile undertaking (p. 132).

This also corresponds to one of the essential elements of a communicative approach: "that the second language learner must have the opportunity to take part in meaningful communicative interaction with highly competent speakers of the language, i.e. to respond to genuine communicative needs in realistic second language situations" (Canale and Swain, 1980). If the instructional program includes this, then so should the testing program.

Let us turn now to a consideration of the literature on communicative competence. It is common to find the term communicative competence used to refer *exclusively* to knowledge of the rules of language use and the term grammatical (linguistic) competence used to refer to knowledge of the rules of sentence grammar. The terms are used in this manner by, for example, Allen (1978), Jakobovits (1970), Palmer (1978), Paulston (1974) and Widdowson (1971). It is equally common to find the term communicative competence used, as proposed by Campbell and Wales (1970) and Hymes (1972), to include not only grammatical competence, but also contextual or sociolinguistic competence, that is, knowledge of the rules of language use. Cooper (1968), Morrow (1977), Munby (1978) and Savignon (1972), among others, use the term communicative competence in this way.

Munby argues that the view that communicative competence includes grammatical competence is to be preferred to the view that it does not because it avoids the misleading conclusion that grammatical competence is not essential to communicative competence. Hymes (1972) has pointed out that there are rules of grammar that would be useless without rules of language use. Similarly, we have claimed (Canale and Swain, 1980) that there are rules of language use that would be useless without rules of grammar. For example, one may have an adequate level of sociolinguistic competence in Canadian French just from having developed such a competence in Canadian English, but without some minimal level of grammatical competence in French, it is unlikely that one could communicate effectively with a monolingual speaker of Canadian French (cf. also Morrow, 1977 on this point). Thus it seems essential to incorporate into any notion of communicative competence both grammatical and sociolinguistic knowledge.

This position does not, however, include another commonly held view about what constitutes communicative competence — the ability to get one's meaning across regardless of the appropriateness or grammaticalness of the utterance.[3] There would seem to be an underlying assumption in some communicative approaches to second language instruction that more effective second language learning takes place if emphasis is put from the beginning on getting one's meaning across, rather than on the grammaticalness or social appropriateness of one's utterance.

The view that meaning should be emphasized over grammaticalness has three bases, each of which deserves to be examined when the instructional program is intended for adolescent learners. One basis is the observation that children in acquiring a first language seem to focus more on being understood than on speaking grammatically, and, therefore, second language acquisition might be permitted to proceed in this way as well. Secondly, there is the assumption that despite initial ungrammaticalness on the part of the child learner with respect to the first language, the first language is always eventually acquired. In other words, it is assumed that an initial emphasis on getting one's meaning across does not preclude the attainment of full grammatical competence at a later stage. The third basis for suggesting that second language acquisition might be allowed to proceed like first language acquisition is the belief that language learning is more effective when the learner is involved in real communicative acts, rather than rehearsing grammatical forms.

Let us consider each of these points in turn. The suggestion that second language acquisition might be allowed to proceed like first language acquisition derives in part from evidence which suggests that the processes of first language acquisition are similar to those of second language acquisition. The evidence comes mainly from studies which have demonstrated that many errors made by first and second language learners are similar (e.g. Dulay and Burt, 1974; Richards, 1972). However, considerable evidence has also been accumulated to suggest that many errors are different due at least in part to language transfer (e.g. Canale, Mougeon and Beniak, 1978; Harley and Swain, 1978; Schachter and Rutherford, 1979). These errors may lead to grammatical forms that are difficult to understand or that may be misinterpreted by native unilingual speakers.[4] Additionally, neurological (Lenneberg, 1967; Scovel, 1978), cognitive (Rosansky, 1975), affective (Schumann, 1975), and contextual (Stern, 1976) factors have been cited as

grounds for differences between children, learning a first or second language, and adolescent and adult second language learners. For example, it has been argued that the onset of formal operations changes the way in which language data are processed and stored (Rosanky, 1975) and that the presentation of material to adolescent or adult learners should therefore be modified to take account of these learners' dominant processing strategies.

The second assumption — that despite initial ungrammaticalness on the part of the child first language learner, the target language is always eventually acquired — is clearly untenable where adolescent or adult second language learners are concerned. There are many studies which report on the interlanguage of adult second language learners, and the fossilization of errors (see, for example, Canale, Mougeon, Belanger and Ituen, 1977; Richards, 1975; Selinker, 1969). It is also the case that fossilization may occur with child second language learners as well.

Consider the case of children learning a second language in which initial emphasis has been placed on getting across one's meaning. The example to be discussed is that of young children in French immersion programs in Canada. In these programs, English-speaking children who are initially unilingual receive instruction for the first several years of schooling entirely in French. They are young, and it has been thought inappropriate to teach grammar in any formal way. The emphasis is on having the children understand what the teacher has to say through instructional techniques which make considerable use of concrete situations, visual aides, and non-verbal cues. The children initially speak in English to their teacher, who is bilingual and can therefore understand the children and respond appropriately in French. The children are encouraged to use what French they have acquired through their exposure to French as well as through the active instruction of vocabulary. In effect, the children are encouraged to use their French long before they have the grammatical knowledge to express their ideas. They resort to numerous communicative strategies (see Tarone, 1978). For example, they use English words in otherwise French utterances, they use English structures, and they use adverbs of time to express tense. There is initially minimal explicit correction by the teacher partly in order not to discourage the children from using French by constantly interrupting and correcting them, and partly because the curriculum emphasis is on getting one's ideas across. At later grade levels, there is more explicit correction, both of oral and written expression. There is also some teaching of grammar, although we do not know the extent to which this occurs, or the form it takes. It would appear to vary considerably from teacher to teacher, and from school board to school board.[5]

What are the results related to these students' second language learning? The details of the answer depend partly on what grade level the children were in when the data were collected, how the data were collected, and which of the skills — speaking, understanding, reading or writing — were being examined. We do not intend to report here on all the details They are reported elsewhere (see, for example, Andrew, Lapkin and Swain, 1979; Barik and Swain, 1975; 1976; Harley and Swain, 1977; 1978; Lambert and Tucker, 1972; Lapkin ad Swain, 1977; Selinker, Swain and Dumas, 1975; Spilka, 1976; Swain, 1975; 1978; 1979; Szamosi, Swain and Lapkin, in press). Furthermore the research associated with getting one's meaning across has focussed mainly on oral skills, rather than on comprehension, and it is on this former aspect that we want to focus now. It is important to note in the following discussion, however, that the immersion children showed no sign of not understanding what was said to them. Their ability to understand was native-like.

The results indicate that although there is gradual improvement over time, the children's speech contains numerous grammatical errors. For example, in a descriptive study of the verb system of fifth-grade immersion students (Harley and Swain, 1978), it was noted that the French verb system of these children is a simplified one; it does not cover the complexities of, for example, the conditional tense, and morphologically irregular verbs are made regular, most verbs are being treated as though they belonged to the simplest, most regular "er" conjugation (i.e. verbs such as parler "to speak"). A comparison group of francophone children was found to have made very few errors of this kind. The study did find, however, that the range of meanings that the immersion students were able to express in the interviews was comparable to that of the francophone students.

In another study with grade two immersion students, we arranged for several immersion students each to spend a couple of hours each week playing together with a francophone child over a period of two months. Their spontaneous interactions during these play sessions were recorded. The tape recordings tell us two things. First, the immersion children made numerous grammatical errors. Secondly, in spite of this, they clearly had "communicative confidence" — they did not hesitate to engage in communicative activities in French. The children argued, persuaded, agreed, contradicted, played with words, interrupted, changed topics, exchanged information, made jokes, etc. (for an example see Szamosi, Swain and Lapkin, in press).

The conclusion to be drawn from these studies is that early emphasis on getting one's meaning across results in the ability to do so, but it does not necessarily imply that grammatical accuracy is the result.

To be sure, if the immersion children were placed in an environment where only unilingual francophones were present, they might develop this grammatical competence. That is to say, their current communicative needs in interacting with individuals whose own competence includes knowledge of English grammar are satisfied, and further development is not necessary in their particular context. Perhaps the expression of more sophisticated ideas, out of the here-and-now context, with unilingual speakers would lead to the learning of the grammatical means essential to express them. But perhaps not. Perhaps they would continue to rely on communicative strategies that result in getting across their in-

tended meaning, but in a non-native-like way.[6] Knowledge about, and use of, communicative strategies (i.e. strategic competence) is, in our opinion, an important component of communicative competence, but not the sole component.

The third basis for the suggestion that second language acquisition should be allowed to proceed like first language acquisition is the belief that language learning is more effective when the learner is involved in real communicative acts, rather than rehearsing grammatical forms. The previous example concerning immersion children speaks partly to this issue in that they were not involved in rehearsing grammatical forms. If "effective" language learning is to include the concept of grammatical accuracy, then it is necessary to conclude that the program has not been fully effective.

There is no similar study, that we are aware of, related to adolescent or adult learners. However, there are several studies which do address the question of the effect on communicative abilities of second language programs which focus initially on teaching grammar alone or alternatively on teaching grammar and basic communication skills. Let us turn now to these studies to see what insights they can provide concerning the role of teaching grammar in developing communicative abilities.

It is a commonly held view that knowledge of grammar alone does not necessarily lead to effective communication. Tucker (1974), for example, found that students who scored high (95th percentile) in English language proficiency as demonstrated by the Michigan Test of English Language Performance and the Test of English as a Foreign Language, and those who scored low (below the 60th percentile), scored equivalently on three of the four communicative tasks. The communicative tasks involved having the testee describe an object or picture to a listener on the other side of an opaque screen such that the listener could identify the object or picture from among an array of such items before him. It seems that an appropriate conclusion from this and other similar studies (for example, Upshur and Palmer, 1974) is that focus on grammatical competence in the classroom is not a sufficient condition for the development of communicative competence. It would be inappropriate, however, to conclude from these studies that the development of grammatical competence is irrelevant to, or unnecessary for, the development of communicative competence, given that all subjects in the studies did have grammatical training.

Savignon (1972) undertook a study to determine the effect of teaching both grammar and communicative skills. In her study, three groups of students were enrolled in a basic introductory audio-lingual course of French at the university level. All three groups received the same number of hours of instruction in the standard (formal and grammatical) program, but one group had an additional hour per week devoted to communicative tasks in which the emphasis was mainly on getting one's meaning across and on being able to cope in communicative situations. The second group spent an additional hour per week in a "culture lab" program, and the third spent an additional hour per week in a language laboratory program.

Savignon found that although there were no significant differences among groups on tests of grammatical competence, the communicatively trained group scored significantly higher than the other two groups on four communicative tasks she developed. The first task was a discussion in French between a student and a native speaker of French on one of three topics; the second, an information-getting interview in which the student had to find out about the native speaker by asking him questions; the third, a reporting task in which the student had to discuss a given topic first in English, then in French; and the fourth, a descriptive task in which the student had to describe an ongoing activity. The criteria of evaluation for these tests included effort to communicate, amount of communication, semantic accuracy of information, comprehensibility and suitability, and naturalness and poise in keeping a verbal interaction in hand. Savignon claims that "the most significant findings of this study point to the value of training in communicative skills from the very beginning of the program" (1972, p. 9).

Oller and Obrecht (1968) report a similar conclusion. In this study, they found that the effectiveness of pattern drills is significantly increased when the language in the drill is related to communication. Their conclusion is that from the beginning of a second language program, aspects of grammatical competence should be taught in the context of meaningful communication. Thus, in sum, these studies suggest that some combination of emphasis on grammatical accuracy and emphasis on meaningful communication from the start of second language study is effective in developing both grammatical skills and communicative skills.

With these considerations in mind, one of five principles which underlie what we consider to be essential to a communicative approach to second language teaching and testing was formulated. It states that:

> Communicative competence is composed minimally of grammatical competence, sociolinguistic competence, and. . .strategic competence. There is no strong theoretical or empirical motivation for the view that grammatical competence is any more or less crucial to successful communication than is sociolinguistic competence or strategic competence. The primary goal of a communicative approach must be to facilitate the integration of these types of knowledge for the learner, an outcome that is not likely to result from overemphasis on one form of competence over the others throughout a second language program (Canale and Swain, 1980).

A brief outline of the boundaries and contents of the three components of communicative competence is therefore in order. A more complete description is found in Canale and Swain (1980). Grammatical competence is understood to include knowledge of lexical items and of rules of morphology, syntax, sentence-grammar semantics and phonology. Grammatical competence is

an important concern for any communicative approach whose goals include providing learners with the knowledge of how to determine and express accurately the literal meaning of utterances.

Sociolinguistic competence is understood to include knowledge of *sociocultural* rules of use and rules of *discourse*. Knowledge of these rules will be crucial in interpreting utterances for social meaning, particularly when there is a low level of transparency between the literal meaning of an utterance and the speaker's intention. The primary focus of *sociocultural* rules of use is on the extent to which certain propositions and communicative functions are *appropriate* within a given sociocultural context depending on contextual factors such as topic, role of participants, setting, and norms of interaction. The primary focus of *discourse* rules is on the *cohesiveness* of groups of utterances, that is, the grammatical links, and on the *coherence* of groups of utterances, that is, the appropriate combination of ideas and communicative functions (see Halliday and Hasan, 1976, and Widdowson, 1978, for discussion).

Strategic competence is understood to include verbal and nonverbal communication strategies that may be called into action at times when the flow of speech might otherwise be impeded due to performance variables or to insufficient competence (see also Tarone, 1978.) Such strategies are of two main types: those that relate primarily to grammatical competence, for example, how to paraphrase grammatical forms that one has not mastered or cannot recall momentarily; and those that relate more to sociolinguistic competence, such as how to hold one's place in a conversation and how to address strangers when unsure of their social status.

Adoption of this framework of communicative competence raises a number of issues for second language pedagogy and testing. The discussion here is limited to those which are directly relevant to the argument that grammatical competence is to be considered an essential component of communicative competence.

It should be noted that it is possible to have a communicative approach to second language instruction in which principles of organization are either grammatically based or functionally based. In our opinion, even though grammatical competence is considered to be an essential component of communicative competence, a program organized on the basis of communicative functions is to be preferred to a program organized on a grammatical basis. Morrow (1977) and Johnson (1977, 1978) have expressed a concern that a second language syllabus organized on the basis of communicative functions may be disorganized with respect to grammar. Thus Johnson (1977) writes:

> It seems reasonable to expect sentences which form a homogeneous functional grouping to be grammatically unlike. The choice of a functional organization therefore seems to imply a degree of structural 'disorganization,' to the extent that many structurally dissimilar sentences may be presented in the same unit, while what may be taken to be key examples of particular grammatical structures will be scattered throughout the course (p. 669).

Furthermore, it is possible that this disorganization may impede second language learning more at the initial stages of study than at later stages (Canale and Swain, 1980). However, this is an open question; there are no empirical data on the relative effectiveness of grammatically organized syllabuses versus functionally organized syllabuses at any stage of learning, nor will there be any satisfactory answers until tests which incorporate the essential components of communicative competence are developed. Obviously, testing instruments which only measure one component of communicative competence will give a distorted picture of what the second language program is able to achieve. Furthermore, if different evaluations assess programs by focusing on different components, no comparisons across programs can validly be made.

In spite of Morrow and Johnson's concern, it is by no means an established fact that a functionally organized approach cannot achieve a level of grammatical organization that is adequate for second language teaching and learning. As Canale and Swain (1980) suggest, there may be means of introducing an adequate level of grammatical sequencing into a functionally organized approach by:

1. making use of grammatical sequencing criteria — e.g. degree of complexity, generalizability and transparency with respect to functions, and acceptability in terms of perceptual strategies — in selecting the grammatical forms to be introduced in covering a given function;

2. treating such grammatical sequencing criteria as an essential subset of the set of criteria used to determine functional sequencing;

3. making use of repetitions of grammatical forms in different functions throughout the syllabus, assuming that such forms are partially specified on the basis of their generalizability; and

4. devoting a certain proportion of classroom time and textbook coverage to discussion of and/or practice on new or especially difficult grammatical points and on interrelationships among various points as implemented, for example, in the course materials prepared by Johnson and Morrow (1978). Other possibilities will no doubt suggest themselves in the course of research on achieving the optimal balance between functional and grammatical organization at a given stage of study.

Another reason for proposing a communicative approach organized on a functional basis has to do with its face validity. It is our view that a functionally based communicative approach — in particular, one in which units are organized and labelled with reference to communicative functions — is more likely to have positive consequences for learner motivation than is a grammatically based communicative approach — in particular, one in which units are organized and labelled with reference to grammatical forms. This view needs, however, to be empirically tested.

Concerning the development of assessment instruments, it has already been pointed out that the distinction between competence and performance suggests the possibility of discrete-point items to measure the components of communicative competence. It must be noted, however, that in order to be consistent with a communicative approach, the point being tested must be presented in a meaningful context that is relevant to the learner's needs. Morrow (1977, p. 28) has suggested that discrete-point tests may be expected to address the learner's competence (sociolinguistic competence in our framework) in assessing a communicative interaction in the following terms:

1. the settings to which it might be appropriate,
2. the topic which is being presented,
3. the function of the utterance,
4. the modality or attitude adopted by the speaker/writer,
5. the pre-suppositions behind the utterance,
6. the role the speaker/writer is adopting,
7. the status implicit in the utterance,
8. the level of formality on which the speaker/writer is conducting the interaction,
9. the mood of the speaker/writer.

A fleshing out of the components within our framework should certainly lead to additions to this list. An obvious gap in the above list is any direct reference to grammatical competence or strategic competence.

Based on both a theoretical argument as well as on limited empirical evidence, this paper has suggested that communicative competence should be considered to consist minimally of three components — grammatical competence, sociolinguistic competence, and strategic competence. The empirical evidence suggests that grammatical competence is a necessary but not sufficient condition for *effective* communication; that initial emphasis on strategic competence does not necessarily lead to full grammatical competence; and that initial emphasis on grammatical competence and strategic competence does not impede the acquisition of grammatical competence, while it aids the acquisition of strategic competence. Although not discussed in the paper, it would also appear, if the immersion children can be taken as an example, that sociolinguistic competence transfers quite naturally in many instances from first language competence.

In addition, the paper has suggested that although grammatical competence plays a significant role in communicative competence, a communicative approach to second language teaching which is organized on a functional basis, rather than on a grammatical one, is to be preferred. However, considerable research evidence is needed to support this view.

In fact, this paper points to the general paucity of research that has been undertaken with regard to the effectiveness of various programs in teaching and testing communicative competence and communicative performance. It is hoped that the general framework provided here and expanded on in Canale and Swain (1980) will lead to a more comprehensive conception of the nature of communicative competence that will guide program designers and test developers, and against which claims for a communicative-based approach to second language teaching and testing can be evaluated. Clearly there remains an enormous amount of work to be done, from the fleshing out of the framework suggested here to the development and validation of test items, and to research, for example, on how various levels of communicative competence by second language speakers of different ages are perceived by target language speakers, and thus what the relative importance is of the three components of communicative competence at various stages of learning. Ultimately it may be research of this sort which will permit us to determine what the target levels of general second language programs should be, and therefore, what the optimal balance of instructional emphasis on grammatical, sociolinguistic and strategic competence should be.

Footnotes

This work has been funded under contract by the Ontario Ministry of Education. We would like to thank the Ministry for its support. We would also like to thank Birgit Harley and Sharon Lapkin who provided insightful comments after reading an earlier draft of this paper, which was presented as a Forum Lecture at the TESOL Summer Institute, UCLA, July 1979.

1. A student who has had 720 accumulated hours of FSL instruction might have had, for example, 40-minute classes over a period of 6 years.

2. For an update of this theoretical framework and research bearing on it, see Canale (1981).

3. Note that the emphasis here is only on oral productive skills (speaking).

4. Evidence which supports this suggestion comes from a Ph.D. thesis in progress in which it has been found that unilingual francophones judge the errors made by immersion students (age 11) to be less acceptable and more irritating than do bilinguals (French-English) (Lepicq, forthcoming).

5. This is a matter of considerable interest. It would be instructive in the debate of the role of teaching grammar *per se* to compare the spontaneous speech (and writing) of children who have had some form of formal instruction in grammar with those who have not, at various grade levels. One question we have, other than whether teaching grammar *per se* makes a difference at this age level, is how early it can be introduced so as to make a difference. Equally important is to determine whether one form of grammar teaching is better than another at the early elementary level.

6. An important issue in the interpretation of the immersion results is whether certain grammatical aspects have not been acquired/learned because the sociocultural needs are not present, or because grammar has not explicitly been taught. Although

we believe that if the sociocultural needs were present in the *immediate* environment through the presence of native French-speaking peers, the French used by the immersion children would be different; it is the case that the political climate and educational systems in Ontario and Quebec make it difficult to create such an environment in the classroom (see Swain, 1979 and Genesee, 1979 for a discussion of these issues). Given that this is the case, then we must turn to, and experiment with, other techniques such as the explicit teaching of grammar. The extent to which it would make a difference is, as yet, an open question.

References

Allen, J. P. B. "New Developments in Curriculum: The Notional and the Structural Syllabus." Paper read at the TEAL Conference, Vancouver, B.C., March 1978. Mimeo.

Andrew, C. M.; Lapkin, S. and Swain, M. "Report on the 1978 Evaluationof the Ottawa and Carleton French Immersion Programs, Grades 5-7." Toronto: the Ontario Institute for Studies in Education, 1979. Mimeo.

Barik, H. C. and Swain, M. "Three-Year Evaluation of a Large-Scale Early Grade French Immersion Program: The Ottawa Study." *Language Learning*, 25, 1 (1976), 1-30.

Barik, H. C. and Swain, M. "Primary-Grade French Immersion in a Unilingual English-Canadian Setting: The Toronto Study through Grade 2." *Canadian Journal of Education* 1, 1 (1976), 39-58.

Campbell, R. and Wales, R. "The Study of Language Acquistion." In J. Lyons, ed. *New Horizons in Linguistics*. Harmondsworth: Penguin Books, 1970.

Canale, M. "From Communicative Competence to Communicative Language Pedagogy." In J.C. Richards and R. Schmidt, eds. *Language and Communication*. London: Longman, 1981.

Canale, M.; Mougeon, R.; Belanger M.; and Ituen, S. "Aspects de l'usage de la preposition POUR en Francais Ontarien: Interference et/ou surgeneralisation?" *Working Papers on Bilingualism* 12, (1977), 61-78.

Canale, M.; Mougeon R.; and Beniak, E. "Acquisition of Some Grammatical Elements in English and French by Monolingual and Bilingual Canadian Students." *Canadian Modern Language Review* 34, 3, (1978), 505-524.

Canale, M. and Swain, M. "Testing and Communicative Competence" Paper presented at 13th Annual TESOL Convention, Boston, 1979.

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

Dulay, H. C. and Burt, M. K. "You Can't Learn without Goofing." In J. C. Richards, ed. *Error Analysis: Perspectives on Second Language Acquisition*. London Longman, 1974.

Chomsky, N. *Reflections on Language*. New York: Pantheon Books, 1975.

Cooper, R. L. "An Elaborated Language Testing Model." In J.A. Upshur, ed. *Problems in Foreign Language Testing. Language Learning*, Special Issue no. 3, 1968.

Fraser, B. "Review of Searle 1969." *Foundations of Language* 11 (1974), 433-46.

Genesee, F. Response to M. Swain's paper presented at the 13th Annual TESOL Convention, 1979.

Halliday, M. A. K. *Language as Social Semiotic*. London: Edward Arnold, 1978.

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

Harley, B. and Swain, M.. "An Analysis of Verb Form and Function in the Speech of French Immersion Pupils." *Working Papers on Bilingualism* 14 (1977), 31-46.

Harley, B. and Swain, M. "An Analysis of the Verb System Used by Young Learners of French." *Interlanguage Studies Bulletin* 3, (1978), 35-79.

Hymes, D. "On Communicative Competence." In J. B. Pride and J. Holmes, eds. *Sociolinguistics*. Harmondsworth, England: Penguin Books, 1972.

Jakobovits, L. A. *Foreign Language Learning*. Rowley, Mass.: Newbury House, 1970.

Johnson, K. "The Adoption of Functional Syllabuses for General Language Teaching Courses." *Canadian Modern Language Review* 33, 5, (1977), 667-680.

Johnson, K. "The Application of Functional Syllabuses." In K. Johnson and K. Morrow, eds. *Functional Materials and the Classroom Teacher*. Reading: Centre for Applied Language Studies, University of Reading, 1978.

Johnson, K. and Morrow, K. 1978. *Communicate: the English of Social Interaction*. Reading: Centre for Applied Language Studies, University of Reading.

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

Lapkin, S. and Swain, M.. "The Use of English and French Cloze Tests in a Bilingual Education Program Evaluation: Validity and Error Analysis." *Language Learning* 27, 2, (1977), 279-314.

Lenneberg, E. H. *Biological Foundations of Language.* New York: John Wiley and Sons, 1967.

Lepicq, D. "Une etude theorique et empirique de la notion de l'acceptabilite." Ph.D. dissertation. Toronto: The Ontario Institute for Studies in Education, University of Toronto, forthcoming.

Morrow, K. E. *Techniques of Evaluation for a Notional Syllabus.* Reading: Centre for Applied Language Studies, University of Reading, 1977. (Study commissioned by the Royal Society of Arts.)

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

Oller, J. W., Jr. and Obrecht, D. H." Pattern Drill and Communicative Activity: A Psycholinguistic Experiment." *IRAL* 6, (1968), 165-174.

Palmer, A. S. "Measures of Achievement, Communication, Incorporation, and Integration for Two Classes of Formal EFL Learners." Paper read at the 5th AILA Congress, Montreal, August, 1978. Mimeo.

Paulston, C. B. "Linguistic and Communicative Competence." *TESOL Quarterly* 8, 4, (1974), 347-362.

Richards, J. C. "A Noncontrastive Approach to Error Analysis." In J. W. Oller and J. C. Richards, eds. *Focus on the Learner: Programmatic Perspectives for the Language Teacher.* Rowley, Mass.: Newbury House, 1974.

Richards, J. C. "Simplification: A Strategy in the Adult Acquisition of a Foreign Language: An Example from Indonesian/Malay." *Language Learning* 25, 1,(1975), 115-138.

Rosansky, E. "The Critical Period for the Acquisition of Language: Some Cognitive Developmental Considerations." *Working Papers on Bilingualism* 6, (1975), 92-102.

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

Schachter, J. and Rutherford, W. E. "Discourse Function and Language Transfer." *Working Papers on Bilingualism* 19, (1979), 1-12.

Schumann, J. H. "Affective Factors and the Problem of Age in Second Language Acquisition." *Language Learning* 25, 2, (1975), 209-235.

Scovel, T. "The Recognition of Foreign Accents in English and Its Implications for Psycholinguistic Theories of Language Acquisition." Paper presented at the 5th AILA Congress, Montreal, August 1978.

Selinker, L. "Language Transfer." *General Linguistics* 9, 2, (1969), 67-92.

Spilka, I.V. "Assessment of Second-Language Performance in Immersion Programs." *Canadian Modern Language Review* 32, 5, (1976), 543-561.

Stern, H. H. "Optimal Age: Myth or Reality?" *Canadian Modern Language Review* 32, 3, (1976), 283-294.

Swain, M. "Writing Skills of Grade Three French Immersion Pupils." *Working Papers on Bilingualism 7, (1975), 1-38.*

Swain, M. "Bilingual Education for the English-Canadian." In J. E. Alatis, ed. *Georgetown University Round Table on Languages and Linguistics, 1978.* Washington, D.C.: Georgetown University Press, 1978. pp. 141-154.

Swain, M. "Target Language Use in the Wider Environment as a Factor in Its Acquisition." Paper presented at the 13th Annual TESOL Convention, 1979.

Szamosi, M., Swain, M., and Lapkin, S. "Do French Immersion Children Really "Know" French? *ORBIT.* In press.

Tarone, E. "Conscious Communication Strategies in Interlanguage: A Progress Report." In J. C. Richards, ed. *Teaching and Learning.* Rowley, Ma.: Newbury, 1978.

Tucker, G. R. "The Assessment of Bilingual and Bicultural Factors of Communication." In S. T. Carey, ed. *Bilingualism, Biculturalism and Education.* Edmonton: University of Alberta Press, 1974.

Upshur, J. A. and Palmer, A. "Measures of Accuracy, Communicativity, and Social Judgments for Two Classes of Foreign Language Speakers." In A. Verdoodt, ed. *AILA Proceedings, Copenhagen 1972, Volume III: Applied Sociolinguistics.* Heidelberg: Julius Groos Verlag, 1974.

Widdowson, H. G. "The Teaching of Rhetoric to Students of Science and Technology." In *Science and Technology in Second Language.* London: Centre for Information on Language Teaching and Research, 1971.

Widdowson, H. G. 1978. *Teaching Language as Communication.* London: Oxford University Press, 1978.

IMPROVING COGNITION SHOULD BE THE CONCERN OF EVERYONE IN EDUCATION: REACTION TO DE AVILA, COHEN AND INTILI

by
Walter G. Secada

Bilingual Education Service Center
Arlington Heights, IL

September, 1981

The reactor's role is a difficult one. He is called upon to critique work in which others have invested more time and effort than he ever will and to raise issues which otherwise might be overlooked. He must critique thoughtfully. De Avila, et al. make the task easier since their work readily suggests a wealth of issues and ideas. I think that three reactions are appropriate to De Avila, et al.'s contribution to this monograph: delight, an analytic attempt to understand the paper, and projection vis-a-vis future work to build upon their work. I will address each reaction in turn; it is hoped, the manner in which the reactions are framed will fulfill my duties as a thoughtful critic.

DELIGHT:

De Avila should be congratulated for presenting Bilingual Education (BE) beyond the linguistic format by which it is usually portrayed. By focusing upon the teaching and learning of content (science and mathematics) and by stressing the importance of higher order thinking skills, this paper draws attention to two major concerns that BE shares with mainstream education. De Avila's findings show that children in BE can rise to the challenge of higher forms of thought in the classroom. His findings also challenge those in BE to incorporate the more general concerns of content and critical thinking into their own teaching.

The alternative approach — the stress of linguistic differences between populations — has had the effect of typing bilingual education to a conceptual and political straightjacket by casting it as being another form of compensatory education. We find mainstream teachers who do not believe that limited-English proficient students can learn; we must contend with public (and administrative) attitudes and biases against language differences and/or the allocation of resources for compensatory purposes. Thus, the policy implications of De Avila's approach, one which raises issues that lie at the heart of mainstream education, should not be underestimated.

For example, Diennes' pioneering work demonstrated how children as young as fourth and fifth grades can engage in sophisticated forms of mathematical reasoning (algebra, group theory) with the assistance of manipulatives (Diennes, 1960, 1963). Bell, Fuson & Lesh (1976) detail a variety of ways by which to teach arithmetic and algebra through manipulatives and small group activities. For the mainstream, the issue is how to make such skills and knowledge accessible to *all* students. For BE, the issue is how to adapt methods from the mainstream to serve the same ends. Regardless of the name, the goal is the same: to make accessible to children content-based knowledge and to challenge them to use higher forms of thought. De Avila's results show that it can be done.

ANALYTIC ATTEMPT AT UNDERSTANDING:

This study was intended to show that children in BE can learn science and mathematics when the subjects involve higher forms of reasoning than are commonly stressed in the classroom. The first set of analyses demonstrated achievement growth by the MICA students beyond what they would be expected to show. A second set of analyses was used to determine the source of that growth — the regular BE program or the MICA program. De Avila found a similar pattern of overall academic growth by both the MICA and BE (comparison) groups. However, on each of the math concepts and the math applications subtests (each of which measures higher forms of cognition), significant time by condition interactions (Table B) indicates that the MICA group had grown more than expected solely on the basis of pretest scores and participation in the BE program. This finding is heartening since it demonstrates the specific effects of MICA: what was meant to be taught (higher forms of thought) was learned.

As regards general effects, De Avila argues that at least MICA did not have deleterious effects relative to BE. A slightly more optimistic interpretation suggests itself — that student participation in the BE program allowed the MICA program to make its effects felt. Since there is no suitable comparison group to test this interpretation (students who took the MICA program but were not in the BE program), the reasoning to it is indirect. It is unlikely that such a group (MICA, no BE) could have exhibited the range of general and specific growth that the MICA group did. The growth of educational achievement follows the maxim: the rich get richer, the poor get poorer. Those scoring below the norm fall progressively farther and farther behind. Thus, without the intervention of the BE program, the two groups (MICA and comparison) should have shown growth which was considerably less than the growth of the *50th* percentile. But not only were their gain scores comparable to that percentile, they actually surpassed it. It is very unlikely that the MICA's program could, by itself, counter the expected trend, i.e., BE made possible the MICA effects.

The difference in interpretation is subtle, but important; it supports suggestions that good pedagogy can be transferred from the mainstream to the bilingual classroom and that the two need to work hand in hand.

Other findings are worth noting. The teachers' failure to predict the "problem" group should provide some solace to those who worry that similar predictions are self-fulfilling. On the other hand, this result may be due to the teachers' unfamiliarity with the curriculum and its demands. Future work might investigate this possibility.

The need for social interaction in laboratory-based curricula might underly two of the more interesting findings — the number of children who scored high on both English and Spanish LAS increased from pre- to post test; and the existence of a group of over-age children for whom the MICA effects were limited. The first result — increase in the number of bilinguals — is rather straightforward. Laboratory approaches require small groups whose participants constantly interact among themselves as they discuss their findings and help each other understand what is going on. This interaction should lead to improved oral language skills, which are measured by the LAS.

On the other hand, if the over-age children failed to participate in the small group activities, we would expect the results which were found — virtually no growth on the English language measures (LAS, and the standardized tests) nor on the tests measuring understanding of the content (they didn't learn the materials regardless of the language in which they were tested), despite their high scores on the Spanish LAS and on the CCS (they had a language for social forms of communication and were not "slow" students). Thus, the important question becomes how these children might have been isolated. Because they were monolingual Spanish speakers, they might not have been able to participate in the rapid exchanges which characterize small group interactions if these exchanges included language mixing or if the exchanges were primarily in English. Both these patterns of language use would likely occur among the bilinguals in the groups. Exchanges involving the monolingual English children would be predominantly in English. That Spanish was not as important a language of exchange (therefore, this group could not use their Spanish skills) is supported by the fact that the increase in the number of bilinguals was due to monolingual Spanish speakers becoming bilinguals, not due to a change in the monolingual English speakers.

Alternately, other forms of selection (e.g., by teachers or by students themselves) based on the children's age and language seem reasonable. Future research needs to investigate the nature of the interactions, which underly lab approaches to teaching math and science in the bilingual classroom, and the possibility that some groups of students are selected out. Finding such a group of children suggests that there may be limitations to the curriculum and pedagogic strategy employed by De Avila.

The methodology employed by this study reflects the pragmatic compromises which are commonly made by "field based" research. I will leave it to others to sift through the tradeoffs which were made in terms of limited resources and the various validities of this study (Cook & Campbell, 1979). There *is* one weakness, however, which might be remedied by later analyses. The choice of comparison groups was done too globally. De Avila formed the comparison group from students who attended "similar" schools, yet little data was provided to support that the *students* were, in fact, similar. Furthermore, the data were analyzed by a series of two-way (pre/post by group) ANOVAs. I would recommend a more refined strategy, since the analyses were based on individual student scores and not aggregate school scores (though schools were selected). Either new groups should be formed by matching students along pretest scores, or analyses of covariance (using the pretest scores as covariate) should be employed. Such refinements might turn up differences which would otherwise be hidden.

An alternate strategy in overcoming methodological weaknesses would be to conduct modified replications. That is the concern of the next section.

FUTURE WORK:

Research efforts replicating the above study might take any (or all) of three directions: a. defining the connections between the curriculum and the thinking skills it is meant to foster; b. investigating the role of small group processes on the educational outcomes; and, c. studying predictors of observed effects.

A. Thinking Skills

As De Avila notes in the initial sections of his paper, the notion of "higher" forms of thought is found under many names and in a variety of contexts. Future work might attempt to better define the specific forms of thought which the MICA program fosters. This could be accomplished by a longitudinal study of some children in the program. Repeated assessment of the hypothesized forms of thought could include: 1) detailed analyses of student protocols while they attempt to solve problems which use those forms of thought and 2) standard measures of those forms of thought. Both research strategies would require detailed task analyses to clearly show the kinds of thought required and how these forms are used.

B. Group Processes

As noted above, group processes might explain the overall increase in the number of bilingual students and that a group of students were resistant to the program's effects. Research is needed to investigate this possibility, as well as mechanisms by which groups form during laboratory experiences.

C Predictors of Achievement

By including tests of L1 literacy in the testing programs, we could obtain measures of L1 BICS (Spanish LAS), L1 CALP (the L1 literacy tests), L2 BICS (English LAS) and L2 CALP (standardized achievement tests in English) (Cummins, 1980). By employing a pre/post-test design, we could then investigate the interrelationships among the constructs and how they change over time. We could also get an empirical test of Cummins' hypothesis that for minority children, L1 CALP is the most potent predictor of future academic achievement (Cummins, 1980).

Future research should also consider adopting a variety of research methodologies, rather than following the classic field-based designs. Due to tradeoffs among methodological needs, policies which counter such pure research considerations, and limited resources for research and evaluation, we seldom can draw definitive conclusions from large-scale studies (Cook & Campbell, 1979). The alternative is to address the issues from a variety of perspectives and to allocate resources differentially, based on the relative importance of the competing questions and on the return promised by the various methodologies (Cronbach and Associates, 1980). I heartily endorse the alternative approach.

For instance, if all three of the above directions were incorporated into a single study, the pretest battery could be used to classify students according to their BICS/CALP scores in L1 and L2, using either factor analytic methods or multidimensional classifications. Students could be randomly drawn from each of the BICS/CALP groups and then be randomly assigned among the various studies. Numbers of children assigned to each study would depend on the importance of that study as well as methodological considerations for each question. Only those children who were being tested to evaluate the effects of the project (in the classic sense of evaluation) would require a comparison group, which could be formed by carefully matching children attending other schools along some preselected variable (e.g., achievement in English language tests at the pretest). This differential allocation of resources to address a variety of related questions would allow the researcher to study the program in greater detail than would otherwise be possible. The detailed results should also allow him to draw some definitive conclusions.

Finally, from an educational point of view, I would encourage other BE programs to replicate the pedagogic practices espoused by this study. The success of laboratory approaches to the teaching of mathematics and science are many (Bell, Fuson & Lesh, 1976) It is unfortunate that greater detail concerning the MICA curriculum and its teaching methods are not provided by De Avila. However, BE programs interested in attempting a similar project should contact their local Bilingual Education Service Center for assistance, if required.

References

Bell, M. S., Fuson, K. C., & Lesh, R. A. *Algebraic and Arithmetic Structures.* New York: The Free Press, 1976.

Cook, T. D & Campbell, T. D. *Quasi-experimentation.* Chicago: Rand McNally, 1979.

Cronbach, L. J. and Associates. *Toward Reform of Program Evaluation* San Francisco: Jossey-Bass, 1980.

Cummins, J. "The Entry and Exit Fallacy for Bilingual Education." *NABE Journal, IV*3, (1980).

Diennes, Z. P. *Building Up Mathematics.* London: Hutchinson, 1960.

Diennes, Z. P. *An Experimental Study of Mathematics Learning.* London: Hutchinson, 1963.

DILEMMAS IN DIAGNOSIS

by
Robert L. Thorndike
Teachers College
Columbia University

One of the common uses of psychometric devices in the field of reading — as in education generally — is for educational diagnosis. Diagnosis is most often a matter that relates to a specific individual, though we may be from time to time interested in making diagnostic judgments about groups. Diagnostic judgments are often based on the comparison of two measures in order to judge whether the individual shows some genuine discrepancy in the traits or characteristics that the two measures represent. Thus, if a child falls at the 50th percentile on a test of word knowledge, but only at the 25th percentile on a test of comprehension of connected prose, the diagnostician must decide how much confidence to place in the conclusion that this child's ability to read connected prose falls short of his knowledge of word meanings. The whole armamentarium of diagnostic devices in the field of reading has its value in suggesting judgments of the type "Ability A is greater than Ability B."

But differential judgments about individuals are slippery customers. They are peculiarly subject to measurement error. Some 45 years ago, Kelley warned of the need for especially reliable tests if such diagnostic judgments were of be made with confidence. Nothing that has developed since then has given occasion for the psychometrician to change his views on this point. The diagnostician, however, cannot wait for the psychometrician to produce the perfect psychometric instrument in order to deal with the practical problems of his day to day functioning. He must get on with the job. And practical limitations of time and resources for carrying out his assessments mean that he will always have to use tools that fall short of psychometric ideals.

This being so, what help can psychometrics offer the diagnostician to "carry on" while he waits for the perfect diagnostic battery? Perhaps some guidance on the level of confidence that he should place in diagnostic judgments might be useful to tide him over.

We must always remember that any test, or any other type of behavior observation, represents only a limited sample from some domain of behavior. It represents the domain only imperfectly, and the score that it produces is only an approximation to the score that the individual would get for the whole domain — or more realistically, that he would get on other samples drawn from that domain. We get evidence on this variability from sample to sample of behavior through the various procedures for obtaining a reliability coefficient, and we express it most usefully for our present purposes as a standard error of measurement. The standard error of measurement may be thought of as the standard deviation of a series of equivalent measures of the same individual, displaying the extent to which the measures scatter away from his "true score."

Suppose, now, we have two measures, X and Y. For concreteness let us say that X is a measure of word knowledge and Y, a measure of paragraph comprehension. Suppose that results from the two measures are expressed in a common equal-unit score scale, such as T-scores or stanines for a common sample of sixth grade pupils. Suppose that Peter differs on the two tests by an amount D, and for concreteness let us say that this difference is 10 points on the T-score or 2 points on the stanine scale, i.e., a difference of exactly one standard deviation. How much confidence should we have that this difference represents something real, and didn't just happen because of errors of measurement in the two tests? How confidently can we expect a difference in the same direction, though obviously not of identically the same amount, if Peter is retested with equivalent forms of each of the two tests?

In setting our level of confidence, we need to take account of three things, two of which have already been mentioned. In the first place, we need to take account of the size of the standard errors of measurement for the two variables. The larger the errors — that is, the lower the reliability — the lower the confidence. The appropriate degree of confidence depends secondly upon the size of the observed difference between the two scores. The larger the difference, the greater the level of confidence. It depends finally, and quite critically, on the correlation between the measures, X and Y, of the two attributes that we are studying. The higher that correlation, the other two factors remaining the same, the less confidence one can have in the meaningfulness of the difference.

Let us look at the rationale for these relationships with specific figures for a definite example. Suppose that the word knowledge test (X) and the paragraph reading test (Y) are each known to have reliability coefficients of .90 for a sixth grade sample and that for the same sample that correlation between the two tests is 0.80. Consider Peter, who scored one standard deviation lower (relative to the standardization group) on the paragraph test than on the word test.

For a single test with reliability of 0.90, the standard error of measurement, expressed in standard deviation units, is:

$$\sqrt{1 - r_{11}} = \sqrt{1 - 0.90} \qquad (1)$$

Reprinted with permission from Walter H. MacGinitie, ed., *Assessment Problems in Reading*. Newark, Del.: I.R.A., 1973.

For the difference between two tests, both expressed in standard deviation units, the standard deviation of differences arising purely from measurement errors, which we might call the standard error of measurement of difference is:

$$\sqrt{2 - r_{xx'} - r_{yy'}} = \sqrt{2 - .90 - .90} = \sqrt{0.20} = 0.45. \quad (2)$$

Thus, a difference between scores of one standard deviation is equal to

$$\frac{1.00}{0.45} = 2.22$$

standard errors of measurement of the differences. Turning to tables of the normal curve, we find that a difference this large or larger could be expected to occur in 13 cases out of 1,000.

A parallel formula gives the standard deviation of differences between two tests when one knows what the correlation between the two tests is. When, as before, each test's scores are expressed in standard deviation units, the formula for standard deviation of differences is:

$$\sqrt{2 - 2r_{xy}} = \sqrt{2 - 2(0.80)} = \sqrt{.40} = 0.63 \quad (3)$$

Thus, a difference between scores of one standard deviation is equal to

$$\frac{1.00}{0.63} = 1.59$$

standard deviations of the differences between these two quite highly correlated variables. Turning once again to our table of the normal curve, we find that, given a correlation of this size, differences this large will occur in 56 of 1,000 cases. Of these 56, on the basis of our earlier calculation, we should expect that 13 were the result of nothing more than measurement error. This leaves 43 that represent presumably "real" differences. Thus, we may say that the odds are 43 to 13 or about 3 to 1 that the difference is a genuine one. The betting odds of 3 to 1 represent one way of expressing the confidence that we should feel in the diagnostic judgment that Peter is better at word knowledge than at paragraph reading.

Following the same rationale that we have used in our illustration, it is possible to prepare tables showing the "betting odds" for representative combinations of reliability, intercorrelation, and size of difference. An illustrative set of such tables is presented in Table 1.

Table 1

Confidence Tables for Diagnostic Judgments: Odds That an Observed Difference between Two Variables Is a Real Difference

Section I: Average Reliability = 0.98

Correlation between Variables

Difference in SD Units	.95	.90	.85	.80	.75	.70	.60	.50	.40	.00
0.25	1:1	2:1	2:1	5:2	5:2	5:2	3:1	3:1	3:1	3:1
0.50	9:1	20:1								
0.75										
1.00										
1.25										
1.50										
1.75										
2.00										

All others greater than 20 to 1

Section II: Average Reliability = 0.95

Correlation between Variables

Difference in SD Units	.90	.85	.80	.75	.70	.65	.60	.50	.40	.00
0.25	1:3	1:2	3:5	2:3	3:4	3:4	4:5	5:6	7:8	1:1
0.50	5:4	2:1	5:2	3:1	7:2	7:2	4:1	4:1	9:2	5:1
0.75	7:2	8:1	11:1	14:1	16:1	18:1	20:1			
1.00	13:1									
1.25										
1.50										
1.75										
2.00										

All others greater than 20 to 1

Section III: Average Reliability = 0.90

Correlation between Variables

Difference in SD Units	.85	.80	.75	.70	.65	.60	.55	.50	.40	.00
0.25	1:7	1:5	1:4	2:7	1:3	1:3	2:5	2:5	2:5	1:2
0.50	1:3	1:2	3:4	1:1	1:1	7:6	5:4	4:3	3:2	7:4
0.75	4:5	3:2	2:1	5:2	3:1	3:1	7:2	4:1	4:1	5:1
1.00	3:2	3:1	5:1	7:1	8:1	9:1	10:1	11:1	13:1	17:1
1.25	3:1	7:1	12:1	18:1						
1.50	7:1	20:1								
1.75										
2.00										

All others greater than 20 to 1

Section IV: Average Reliability = 0.85

Correlation between Variables

Difference in SD Units	.80	.75	.70	.65	.60	.55	.50	.45	.40	.00
0.25	1.18	1:9	1:7	1:6	1:5	2:9	2:9	1:4	1:4	1:3
0.50	1:6	1:3	2:5	1:2	1:2	2:3	2:3	3:4	4:5	1:1
0.75	1:3	2:3	1:1	8:7	4:3	3:2	5:3	7:4	13:7	5:2
1.00	2:3	4:3	2:1	5:2	3:1	10:3	11:3	4:1	4:1	6:1
1.25	1:1	5:2	4:1	5:1	6:1	7:1	8:1	9:1	10:1	15:1
1.50	5:3	4:1	8:1	10:1	14:1	17:1	20:1			
1.75	9:2	13:1								
2.00	7:1									

All others greater than 20 to 1

Section V: Average Reliability = 0.80

Correlation between Variables

Difference in SD Units	.75	.70	.65	.60	.55	.50	.45	.40	.00
0 25	1 19	1 11	1 9	1 8	1 7	1 6	1 6	1 5	1 4
0 50	1 8	1 5	1 4	1 3	2 5	2 5	1 2	1 2	2 3
0 75	1 4	1 2	2 3	3 4	4 5	1 1	1 1	9 8	3 2
1 00	2 5	4 5	1 1	7 5	8 5	9 5	2 1	11 5	3 1
1 25	5 8	5 4	2 1	5 2	3 1	7 2	4 1	9 2	7 1
1 50	1 1	2 1	3 1	9 2	11 2	7 1	8 1	9 1	13 1
1 75	3 2	7 2	6 1	8 1	11 1	14 1	17 1	19 1	39 1
2 00	2 1	6 1	11 1	16 1					

All others greater than 20 to 1

Section VI: Average Reliability = 0.75

Correlation between Variables

Difference in SD Units	.70	.65	.60	.55	.50	.45	.40	.00
0 25	1 33	1 18	1 13	1 11	1 10	1 9	1 8	1 5
0 50	1 12	1 8	1 5	1 4	2 7	2 7	1 3	1 2
0 75	1 6	2 7	2 5	1 2	3 5	2 3	2 3	1 1
1 00	1 4	1 2	2 3	6 7	1 1	6 5	4 3	2 1
1 25	2 5	3 4	1 1	7 5	5 3	2 1	9 4	4 1
1 50	1 2	1 1	2 1	7 3	3 1	7 2	4 1	7 1
1 75	4 5	8 5	5 2	7 2	9 2	6 1	7 1	14 1
2 00	7 6	5 2	4 1	6 1	8 1	11 1	13 1	30 1

Section VII: Average Reliability = 0.70

Correlation between Variables

Difference in SD Units	.65	.60	.55	.50	.45	.40	.00
0 25	1 47	1 23	1 16	1 14	1 12	1 11	1 6
0 50	1 20	1 10	1 7	1 5	1 5	1 4	2 5
0 75	1 10	1 6	1 4	1 3	2 5	1 2	3 4
1 00	1 6	1 3	1 2	3 5	7 10	4 5	7 5
1 25	1 4	1 2	5 7	1 1	1 1	4 3	5 2
1 50	1 3	5 7	1 1	3 2	9 5	2 1	4 1
1 75	1 2	1 1	8 5	2 1	3 1	7 2	8 1
2 00	2 3	3 2	5 2	7 2	9 2	5 1	14 1

Section VIII: Average Reliability = 0.60

Correlation between Variables

Difference in SD Units	55	50	45	40	00
0 25	1 56	1 35	1 24	1 20	1 9
0 50	1 30	1 14	1 10	1 8	1 4
0 75	1 14	1 7	1 5	1 4	1 2
1 00	1 9	1 5	2 7	2 5	4 5
1 25	1 5	2 7	2 5	5 9	13 10
1 50	1 5	2 5	3 5	4 5	2 1
1 75	2 7	3 5	4 5	6 5	3 1
2 00	3 4	3 4	6 5	5 3	5 1

Consider first Section III of Table 1 — the section for an average reliability of 0.90 — since 0.90 is a fairly representative reliability for good quality ability tests. Note first that no column is shown for an intercorrelation of .90 or higher between the two tests. Whenever the intercorrelation of two tests is as high as their respective reliabilities, they are effective measures of identically the same trait. Differences between the two are then equivalent to (and equal in number to) differences arising solely from measurement error; there is no basis for a diagnostic judgment, and any diagnostic statement should be made with exactly zero confidence.

Note next that when the difference is small, the betting odds are low that this is a real difference, no matter what the correlation. In the row corresponding to a difference of a quarter of a standard deviation, the odds that the difference is a "real" one range from 1 real difference to 7 chance differences when the correlation is 0.85, to 1 real difference to 2 chance differences when the correlation is zero. Most small differences are readily attributed to measurement errors, and our confidence that there is any "real" difference must be correspondingly low.

Finally, in this table we can see the role that the correlation between two test scores plays in our confidence in the reality of any observed difference. This is seen perhaps as clearly as anywhere in the row corresponding to one full standard deviation of difference — a difference that would correspond roughly to falling at the 70th percentile of a group on one measure and the 30th on the other. For a difference of this size, our betting odds would be 3 to 2 in favor of a "real" difference if the correlation between the two test scores were 0.85, 3 to 1 if the correlation were 0.80, 9 to 1 if the correlation were 0.60, and 17 to 1 if the correlation were zero. The confidence we should have in a diagnostic judgment rises sharply as the correlation between the two measures on which the judgment is based decreases.

To view the effect of test reliability on the confidence appropriate for our judgments, it helps to arrange the tables in a somewhat different way. Table 2 shows the "betting odds" when the size of the difference between X and Y is fixed at one standard deviation, but the values of the average reliability and the intercorrelation are allowed to vary. This table makes it emphatically clear how crucially one's confidence depends upon the reliability of the measuring instruments. If the average of the two reliabilities is 0.98 (one should live to see the day when such measures are available!), even the smallest differences, i.e., those of a quarter of a standard deviation, can be accepted with great confidence as real and not the result of measurement error. With a reliability as low as 0.60, a full standard deviation of difference justifies betting odds of less than even money, even when the correlation between the two measures is zero. For intermediate reliabilities, considerable confidence is justified if the correlation between the two measures is low, relatively little confidence is justified if the intercorrelation approaches anywhere near the reliability

Table 2

Odds That an Observed Difference of One Standard Deviation between Two Variables Is a Real Difference

Average Reliability	Correlation between Variables										
	.95	.90	.85	.80	.75	.70	.65	.60	.50	.40	.00
.98	All greater than 20 to 1										
.95	13:1 Remainder greater than 20 to 1										
.90		3:2	3:1	5:1	7:1	8:1	9:1	11:1	13:1	17:1	
.85			2:3	4:3	2:1	5:2	3:1	11:3	4:1	6:1	
.80				2:5	4:5	1:1	7:5	9:5	11:5	3:1	
.75					1:4	1:2	2:3	1:1	4:3	2:1	
.70						1:6	1:3	3:5	4:5	7:5	
.60								1:5	2:5	4:5	

What do the tables that we have looked at imply when this type of thinking is carried over to some samples of actual tests with the reliabilities and intercorrelations that characterize them?

Davis[1] has carried out some of the most meticulous research on the differentiability of different types of reading skills. Among the abilities that he studied, two that were most readily distinguishable were word knowledge and drawing inferences. His tests had to be quite short, since he was measuring some eight different aspects of reading, so the reliabilities of these two tests were only .58 and .59. The correlation between them had an average value of .45 in several sets of data. Given these values, the betting odds are only 1 to 4 that a difference of one standard deviation between scores on the two tests is "real"; for a difference of two standard deviations, the betting odds are 9 to 8. As they stand, the tests hardly justify diagnostic inferences even when the differences are *very* large. But these tests were short — only 12 items each. If they were lengthened to 48 items, which might be a reasonable length for a test in practical use, one estimates that the reliabilities would be increased to .85 and .86, and the intercorrelation to .66. Then the betting odds are respectively 5 to 2 for a difference of one standard deviation and 80 to 1 for a difference of two standard deviations. Thus, we see how very critically diagnostic inferences depend upon the reliabilities of the constituent measures.

Two of Davis' tests that measure more similar functions are the test of inference and a test that calls for identification of the author's tone, mood, and purpose. Here the reliabilities are .59 and .63, and the intercorrelation is .53. Given those values, the betting odds for the existing test are only 2 to 11 for a difference of one standard deviation and 3 to 5 for a difference of two standard deviations. Lengthened to 48 items, reliabilities become .84 and .88 and the intercorrelation 0.75. For this lengthened test, the betting odds are 5 to 3 that an observed difference of one standard deviation is "real" and 23 to 1 for a difference of two standard deviations.

Let us turn our attention now to the Stanford Diagnostic Reading Tests, the distinctive value of which is presumed to lie in their diagnostic effectiveness. Here, unfortunately, the manual provides only single-testing estimates of reliability, and these are certainly somewhat inflated. We cannot know how much. If we take the figures at face value, the average of the subtest reliabilities is 0.90 and the average of the subtest intercorrelations is 0.65. A more realistic estimate of alternate-form reliabilities might be 0.85. If we assume that figure, and turn to Section IV of table 1 for reliability 0.85, we find figures in the column for intercorrelations of 0.65 as follows:

0.25 S.D. (which would occur for 38% of children)	1 to 6
0.50 S.D. (which would occur for 27% of children)	1 to 2
0.75 S.D. (which would occur for 19% of children)	8 to 7
1.00 S.D. (which would occur for 12% of children)	5 to 2
1.50 S.D. (which would occur for 4% of children)	10 to 1
2.00 S.D. (which would occur for 1% of children)	over 20 to 1

Thus, if we limit our diagnostic inferences to the one percent with the most extreme differences, our judgments will almost always have a real basis. If we set a lower threshold, and undertake diagnostic statements for as many as 10 percent of children, there will be a basis in reality for something like three-fourths of our judgments. If we set a still more liberal standard and venture diagnostic statements based on observed differences for as many as 20 percent of the group, the statements will correspond to real differences only about half the time.

Finally consider a set of data for the reading test of the Stanford Achievement Battery given once in the sixth and once in the eighth grade. For one suburban New York school system, the correlation between the two testings was .747. An estimate of reliability drawn from the test manual is .93. How much would a child have to change his position in his group from the first to the second testing for us to have an even-money bet that there was a real change? The answer comes out to be 0.40 standard deviations. If a child were to improve his position in his group by four-tenths of a standard deviation (for example, from the 50th to the 65th percentile), it is a fifty-fifty proposition that this represents some degree of real change and not just the effect of measurement errors.

The tables and illustrations that we have examined illustrate the impact of reliability, intercorrelation, and score difference upon the confidence that one can logically place in an observed difference between two scores. They illustrate that over the realistic range of test reliabilities, and using the kinds of pairs of measures that we are likely to want to use in diagnostic studies, the confidence is often distressingly low. But children with reading disabilities are there, and they won't just go away until that happy day when we have diagnostic tools of reliability high enough to permit us to make judgments of score difference at a high level of confidence. Therein lies our dilemma. Wherein do we find our salvation?

If salvation exists, it lies in the fact that most of the actions following from diagnostic judgments are reversible, and if they are unfounded, they are likely to result in wasted time or effort rather any more crucial loss. In this respect, instructional decisions differ from selection and classification decisions, since these are typically permanent. The young person who is denied access to a

particular educational institution or job is not likely to be given a second chance. But if the special instruction in word-analysis skills that seems to be called for by a diagnostic reading profile is not effective, it is always possible to hold up, take stock, get new or additional evidence, and follow up some alternative hypothesis. Our tables of betting odds suggest how tentative our hypotheses should often be. Fortunately, they often can be tentative. It is important that we keep them so.

References

Davis, Frederick B. "Research in Comprehension in Reading." *Reading Research Quarterly*, 3 (Summer 1968), 499-545.

Kelley, Truman L. *Interpretation of Educational Measurements*. New York: World Book, 1927.

LANGUAGE PROFICIENCY ASSESSMENT: ISSUES AND DEFINITIONS

by
Richard E. Baecher
Associate Professor
Fordham University at Lincoln Center

Confusion, misinterpretations, and conflicting theoretical beliefs characterize the language assessment landscape as currently understood and practiced by college and university personnel. A somewhat similar, pessimistic picture describes the efforts of state and local public school personnel in classifying and measuring their "limited English proficient" populations and bilingual teachers, respectively.

Confusion is apparent in the plethora of terms that are used to identify some aspects of the langauge assessment process over a specific period of time. For instance, there are the terms, "language proficiency," "language dominance," "bilingual proficiency"; another set includes "language aptitude," "language ability," "language attainment," "linguistic academic achievement," and "global language proficiency." With emphasis upon the bilingual individual, such classifications as "balanced bilingual," "equilingual," "comparably limited," and finally, "semilingual vs. alingual" have occurred in the literature produced by "academic scholars" and "professionals," respectively (Note 1). A careful analysis of these terms that have in common the notion of language assessment results more often in confusion than clarity.

In addition, misinterpretations are another outcome of the incomplete definitions of these terms and their operations. For example, "limited-English proficient" pupils whose functional language is English are programmed to receive instruction in their primary language to better understand such content-area concepts as "sets," "electricity," and "democracy"; or there are first grade children whose native language is not English and who are taught in English only. In the academic setting, misinterpretations are evident in admission committees rejecting promising bilingual teachers and administrators solely on the basis of standardized test scores that presumably yield valid and reliable information about an individual's English verbal ability.

Besides these misinformed plans of action resulting from the process of language assessment, opposing theoretical frameworks continue to be debated, particularly in the area of the meaning of the term, "language proficiency." While this debate attests to the vitality and dynamism of the fundamental disciplines (Note 1), especially, psycholinguistics and sociolinguistics, academic scholars in these disciplines actively and learnedly continue to refine their knowledge of the forms and functions of language; however, analysis of the types of definitions accorded the notion of language proficiency, to be discussed in a later section, reveals the ambiguity and vagueness in which this important concept is expressed.

The purpose of this article, then, is twofold. Based upon selected aspects of the theory of definition as proposed by Leonard (1967), issues resulting from the different conceptions of language proficiency will be highlighted. Properties of definitions, such as ambiguity and vagueness, and basic rules for defining terms will be applied to judge the completeness of current definitions of language proficiency. The second aim of this article is to recommend certain guidelines in formulating definitions that can be useful to professionals and practitioners in their practical endeavors to determine this attribute of individuals.

Selected Features of the Theory of Definition

According to Leonard (1967), one of man's great devices for communicating and clarifying the meanings of words or phrases is the definition. A definition usually means *the act of stating* the signification of a word or phrase, and *a statement* of the signification of a word or phrase.

Insofar as definitions are *acts of stating*, they are important pieces of productive discourse. The definer may have different purposes, usually accompanying the statements of the meanings of terms, i.e., cognitive or pragmatic. A cognitive purpose is when the definer wants to affect or change someone's belief or knowledge about the meaning of a term; a pragmatic purpose is evident if the definer's aim is to influence other people to use certain words or phrases in a manner different from that which they have employed in the past. Leonard (1967) has divided the types of definitions in the following original list of exhaustive pairs:

I. Complete — Incomplete
II. Nominal — Real
III. Informative — Hortatory
IV. Linguistic — Conceptual
V. Extensional — Intensional

These types of definitions will be clarified as the need arises in the next section, but they are presented here to indicate the variety of purposes for which individuals can use definitions. (See Note 2 for their meanings.)

Definitions are especialy helpful in eliminating ambiguous meanings. Since most terms or expressions allow for more than one distinct meaning, e.g., "run: the clock runs, the stocking runs, etc.," there are circumstances in which one cannot tell which of *two or more* customary meanings of the expression was intended by the user of

*Gratitude is expressed to Mr Adel El Banna who compiled many of the definitions cited in the article

that expression. This situation can be cleared up by defining each customary use of the term.

Vagueness, on the other hand, presents a difficulty involving only *one* customary meaning for a word or phrase. The custom in question has never fixed the exact limits of what is included within the meaning of the term. In some circumstances, it is used in a more inclusive sense, in others, in a less inclusive sense, e.g., "greater Chicago," "greater New York." Through the device of definitions, meanings of terms are clarified, thereby reducing vagueness. Figure 1 contrasts the two distinct properties of ambiguity and vagueness.

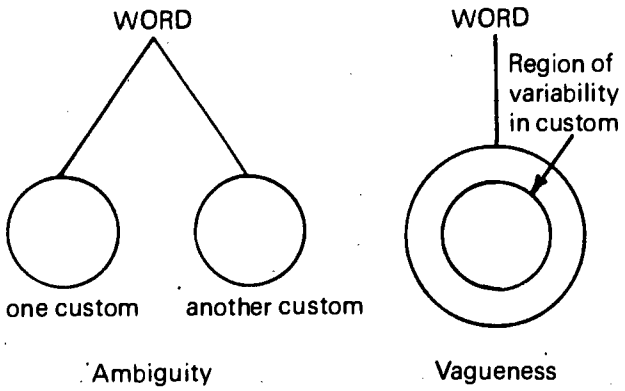


Figure 1. (From Leonard, 1967, p. 27)

Two technical terms employed in the theory of definition are *definiendum* and *definiens*. To illustrate these terms, consider the following expressions:

1. Set $=_{Df}$ a carefully defined collection of elements.

Here the term to be defined is given first (definiendum), then an equal sign, and that, in turn, followed by the expression which states the meaning of "set," (definiens). However, the normal conventions of English would permit any of these expressions:

2. By a *set* is meant a carefully defined collection...
3. A set is a carefully defined collection...
4. A carefully defined collection of elements is called a *set*.
5. Let us understand by the term "set" a carefully defined collection...

What these five definitions have in common, despite their different linguistic contexts, is the expression — definiendum ("set") — to be defined and the expression — definiens ("a carefully defined collection of elements") — selected by the producer as the means of communication. Example one will be the standard form in which definitions will be presented in the rest of this article.

Regardless of the purposes, types and forms of definitions, certain basic rules for defining terms are useful because they propose certain criteria in terms of which one may criticize definitions that one reads or hears. Leonard (1967) divides these rules into three groups: rules stating *literary* requirements, those stating *factual* requirements, and those stating *formal* requirements. These rules, with brief explanations, follow.

Rule 1. A definition should be as clear as possible.

Applicable to all discourse, this Rule requires the producer to accompany the key statement of the definition, or its core, with additional remarks that will clarify the kind of definition being proposed, i.e., real, nominal, linguistic, conceptual, etc. (See Note 2.)

Unless this is done, or the context indicates the concern of the definer, the definition will not be as clear as possible. In addition, the producer must express the definiens in language that the receiver can be expected to understand. Finally, Rule 1 recommends the giving of complete, intensional definitions, unless the circumstances dictate another procedure. (A complete, intensional definition is one intended to explain completely what characteristics are in the total strict intension of the definiendum, e.g., "a circle $=_{Df}$ a closed plane curve all points of which are equidistant from a given point called the center"; contrast this with "a circle $=_{Df}$ the minimum perimeter area of a given size.")

Rule 2. A definition should avoid figurative and metaphorical language.

Since definitions are considered as ingredients of technical discourse, as such, they should not contain anything tending toward vagueness, imprecision, or obscurity. Metaphorical or figurative expressions, e.g., "the devil" as "the prince of darkness," are vague and imprecise and do not have a place in any technical discourse. Therefore, the definiens must not employ vague or figurative language.

Rules 1 and 2 state the *literary* requirements of definitions.

Rule 3. The definiendum and its definiens should be coextensive.

This Rule requires that the extensions of the definiendum and definiens, respectively, be identical; nothing must be in the extension of either term which is not in the extension of the other. As illustrations, these violations of Rule 3 are presented:

a. **Example:** A Mazda $=_{Df}$ an automobile

The definiens is too broad: its extension includes objects that are not in the extension of the definiendum.)

b. **Example:** A Mazda $=_{Df}$ a red car used as a means of transportation.

(The definiens is too narrow: its extension excludes objects that are in the extension of the definiendum.)

c. **Example:** A Mazda $=_{Df}$ a red car.

(The definiens is both too broad and too narrow.)

Rule 4: An intensional definition must give the essential characteristics of the term being defined.

In general, Rule 4 affirms that the total strict intensions of the definiendum and the definiens must be identical. A test of this Rule would find out whether or not: (1) the characteristics "given" in the definiens are *necessary* members of the total contingent intension of the definiendum, and (2) that set of characteristics is *necessarily* jointly peculiar to the extension of the definiendum. An example in which Rule 4 is not violated is the following: "Thesaurus $=_{Df}$ a dictionary of synonyms and antonyms." A violation of this rule is evident in this instance: "Giraffe $=_{Df}$ a mammal with extremely long front legs and an extremely long neck;" although some essential characteristics are listed in the definiens, the definition is incomplete, omitting back legs and a black-blotched fawn, also essential characteristics of a giraffe.

Rules 3 and 4 capture the *factual* requirements of definitions.

Rule 5. The simple definiendum of a definition must not appear in the definiens.

Examples of violations of Rule 5 are: "a house $=_{Df}$ a house"; and "snake $=_{Df}$ the offspring of a snake." Definitions violating this Rule are said to be circular.

Rule 6. The definiens of a definition should avoid the use of simple synonyms of the simple definiendum.

Rule 6 is applicable to conceptual definitions, but not linguistic ones, as in dictionary definitions. A violation of this rule involves the fallacy of word substitution, e.g., "a wagon $=_{Df}$ a cart"; contrast this definition with, "a wagon $=_{Df}$ a wheeled vehicle designed to be drawn by an independent source of power."

Rules 5 and 6 address some of the formal requirements of definitions.

These basic rules for definition, in addition to the process of definition, can be employed in understanding, clarifying, and constructing a definition of the frequently used term, language proficiency.

Current Definitions of Language Proficiency

This section of the paper will summarize current definitions of the term, language proficiency, and apply the concepts and rules for definitions delineated previously.

In an unpublished manuscript, Farhady (n.d.) argues that although current theories of language proficiency testing have generated numerous hypotheses, many have to be questioned because the term is inadequately defined. "Langage proficiency is one of the poorly defined concepts in the field of language testing" (Farhady, n.d., p. 3). Illustrations of this conclusion will follow citing representative academic scholars and their definitions. The standard form for definitions will be used.

Briere (1972, p. 332), acknowledging the complexities of the concept of language proficiency, states his definition:

Proficiency $=_{Df}$ the degree of competence or the capability in a given language demonstrated by an individual at a given point in time independent of a specific textbook, chapter in the book, or pedagogical method.

This definition is complicated and includes words that are vague and unspecified, e.g., "competence" could refer to linguistic, sociocultural, or other types of competencies. Vagueness is apparent in the use of other terms such as "demonstrated" (how? orally, written modes) and "at a given point in time." Ambiguity is evident in viewing language proficiency as either a competence or capability. This definition, then, appears to violate Rule 1 (clarity) and Rule 4 (requirements of an intensional definition). The concepts of "competence, capability, demonstration, given point in time" must be clarified, and the essential characteristics of the term must be identified.

Addressing the area of proficiency testing, Clark (1975, p. 10) adds another feature to the term proficiency:

Proficiency test $=_{Df}$ any measurement procedure aimed at determining the examinee's ability to receive or transmit in the test language for some pragmatically useful purpose within a real-life setting.

In this definition, language proficiency includes another element, the use of language for real-life purposes. This definition, then, includes all the complexities of previous definitions in addition to another concept, "real-life setting." What consititutes a "useful purpose within a real-life setting" remains unclear and vague. In another article focusing upon the differences between direct and semi-direct tests of speaking ability, Clark (1979, p. 37) employs the term "global proficiency":

Global proficiency $=$ the examinee's ability to carry out various language-use tasks appropriately and effectively in realistic communication settings.

This definition remains incomplete and vague because "carry out" and "language-use tasks" are not specified; furthermore, Rule 3 is violated since the definiens is too narrow: it excludes academic settings that require individuals to be proficient in classroom communication.

Upshur (1979), formulating a functional proficiency theory for language tests, distinguishes two kinds of language proficiency:

1. Language proficiency: $_{Df}$ a relation between an individual and a situation requiring the use of language.

2. Language proficiency: Df a psychological capacity of an individual which together with other capacities enables him to function in a situation requiring the use of language.

In Upshur's view, tests developed with this latter definition in mind, essentially language tests for research purposes, seek to answer the question, "Does somebody have proficiency?"

Although this conceptual distinction between two types of language proficiency merits closer attention because of its implications for the validity of such tests, greater clarification of both definitions is necessary. For example, the use of the term "relation" in the first definition remains vague, even in the vivid examples provided by Upshur: what kind of relation is meant? The second definition violates Rule 3 despite the following claim: "We find that the construct of *proficiency* has become virtually coextensive with *human psychology*" (Upshur, 1979, p. 83). The definiens of the second definition has become too broad according to this statement; moreover, the definition does not meet the requirements of Rule 4 whereby the essential characteristics of the term are identified.

These definitions of academic scholars, in particular the areas of linguistics, psychology, and psychometrics, have their origins in the testing of foreign language proficiency. The next set of definitions addresses the educational development of limited-English proficient individuals, and varying theoretical viewpoints toward language proficiency are evident.

In an article that delineates some guidelines for the assessment of oral language proficiency, Burt and Dulay (1978, p. 178) present the following definition:

Language proficiency: Df the degree to which an individual exhibits control over the use of the rules of a language for one, some, or all of its numerous and diverse aspects.

These aspects include "the phonological, syntactic, lexical and semantic systems, and discourse and stylistic rules for oral and written communication for different varieties of a given language in various domains and social circumstances." This definition, in contrast to the previous ones that emphasize the individual's ability to use language for real-life contexts, focuses attention upon the multidimensional nature of language proficiency, in particular, the subsystems of a language. What remains unclear and ambiguous in their definition is the phrase, "exhibits control," which later in the same article is distinguished by means of two types of oral language elicitation tasks, natural communication vs. linguistic manipulation tasks, respectively. The former task, while demanding the communication of something to someone, yields the speaker's unconscious use of language rules, the latter in which the focus of the individual is on performing a conscious linguistic manipulation of the language, demonstrates the individual's meta-linguistic awareness. Two different meanings are implied by the phrase, "exhibits control"; one is conscious control and the other is subconscious. Moreover, the nature and types of language rules are never clarified. Rule 4 is violated in this definition because the definiens does not list the necessary characteristics of the definiendum: proficiency is never clearly defined, whereas language is comprehensively covered.

DeAvila and Duncan (1980, p. 111) view the term as follows:

Language proficiency = Df the student's language skills in English which are learned in both school and natural settings. . . It is not necessarily dependent upon specific instruction or content. . .language achievement is more likely to be dependent upon proficiency than vice-versa.

Examination of the technical manual accompanying their test, *Language Assessment Scales* (LAS) (DeAvila and Duncan, 1975), indicates that a student's oral language proficiency is viewed as performance across four linguistic subsystems, instead of one, single aspect of language. These are: 1) the phonemic, 2) the referential, 3) the syntactical, and 4) the pragmatic subsystems. Their definition represents a different theory of language proficiency from that of Burt and Dulay; in addition, it confuses achievement and proficiency and never specifies language skills that are learned in school and those acquired in natural settings. Rules 1 and 4 are violated in this definition.

Cummins (1978; 1979) divides language proficiency into "cognitive academic linguistic proficiency" (CALP) and "basic interpersonal communicative skills" (BICS).

CALP = Df the ability to make effective use of the cognitive functions of the language, i.e., to use language effectively as an instrument of thought and represent cognitive operation by means of language.

An illustration of CALP would be the individual who is in tune with the semantic complexity of a language, both denotative and connotative, and is capable of carrying out cognitive operations in the language. CALP is in contrast with BICS, or the "surface" linguistic features such as pronunciation, vocabulary, fluency, sociocultural competence, accent. CALP is definitely an important concept because of the nature and requirements of classroom learning with its emphasis upon literacy skills and deriving meaning from printed materials. However, CALP as a definition of language proficiency suffers from vagueness in the use of the phrase "cognitive functions"; are analogies, synonyms, antonyms meant by these functions? Moreover, how does one's knowl-

edge of the technical aspects of language, e.g., morphology and syntax, relate to CALP? Because the essential characteristics of CALP are not specified, Rule 4 is violated, thereby making this definition incomplete.

This analysis of current definitions of the term, language proficiency, based upon selected features of the theory of definition, has demonstrated the different theoretical viewpoints attached to its meaning. Issues relating to different assumptions of language and proficiency, to purposes in testing language proficiency, and to the meanings of this term were identified. Dieterich and Freeman have appropriately summarized the state of the art (1979, p. 2): "English proficiency — what it is to *know* English — is given different operational definitions in each theoretical, historical, and legislative context." In addition, this section has shown the need for greater clarity and precision in the use of this term on the part of those who will define it. What procedures, then, can be employed in the construction of a definition of the complex term, language proficiency?

Some Guidelines for Defining Language Proficiency

The following guidelines, although incomplete, are suggested as aids in assisting professionals and practitioners in their practical endeavors to assess the language proficiency of their students. Before these guidelines are presented, the term, assessment, must be clarified in relation to measurement and evaluation. They are adapted from Ryan and Cruz (1974, pp. 4-5) and defined within a view of education that is contextualistic, i.e., educators — administrators, teachers, parents — are called upon each working day to make decisions about the curriculum and its relevance to the student. To enable these educators to make the *best* decision (and not necessarily the *ideal* one), various forms of information are gathered and interpreted. Measurement, evaluation, and assessment are terms that are applicable to the types of decisions made by educators in the context of helping students understand their world.

Measurement: the application of a standard to a set of data.

Example: (1) This pencil is 6 inches long.

(2) José, a fourth grade pupil, is reading English at the 2.1 grade level (according to test manual norms).

Measurement: information + a standard

Evaluation: the consideration of a set of measurement data in terms of specified priorities for change.

Example: (1) This 6-inch pencil is not long enough to reach the floor.

(2) Since José is in the first month of grade 4, he's labeled as "limited-English proficient," or a "slow reader."

Evaluation: Information + a standard + priorities

Assessment: a process or program of inventorying an individual's strengths and weaknesses, skills, and attitudes that are useful in relating to the roles and symbolic conditions required by various educational tasks. (Note 3)

Example: (1) Although this 6-inch pencil is not long enough to reach the floor, this new pointer that folds into itself helps me do the job more effectively and easily.

(2) Although this standardized test placed José at the 2nd grade level in reading, empirical observations of José in class indicate he can read most materials presented to him and interact in a positive manner with other students in the class.

Assessment: information + standard + priorities + context

While the terms, measurement, evaluation, and assessment are interrelated, assessment is to be preferred as a basis for decision making in educational contexts. This conclusion is based upon the notion that the context or situation does make a difference.

With these distinctions in mind, the following guidelines are presented in constructing a definition of language proficiency. Analysis of the definitions in the previous section had one common feature: individuals differ among and within themselves with regard to language proficiency. The use of such words and phrases as "degree of competence," "capacity," "control over the rules of language," and "cognitive operations" is an indicator of these individual differences; furthermore, classifications such as linguistic subsystems or types of relative language proficiency — "proficient bilinguals," "partial bilinguals" — point out the following rule: individuals differ among themselves in language proficiency or ability.

One guideline, then, is to specify or define the property with which one is concerned. This definition, in turn, will yield a series of operations that will allow the description of individuals in terms of that property. According to Ghiselli (1964, p. 16), "a good definition of a variable is precisely formulated." Specificity identifies the essential characteristics of the property and facilitates the development of a series of operations that enable one to observe similarities and differences among individuals. Adherence to the rules for defining terms, awareness of the definer's aims, and avoidance of vague and ambigious expressions will help in the statement of precisely formulated definitions.

Another guideline is to distinguish between what Ghiselli (1964) calls "trait names" and "trait definitions." Trait names are employed to identify, label, and reasonably represent the definition of a property. As illustrated in the previous section, the label, "language proficiency," was used to represent several different definitions, and since the name comes from the definition, it would be mistaken to claim that one or the other is a better or more valid definition of "language proficiency." Care must be exercised in comparing the results of different definers or the findings with different language proficiency tests in which the same property is nominally involved.

Another guideline of particular importance to bilingual educators is the kinds of individuals with whom one is trying to determine the property. The kinds of individuals of concern to educators have an important influence upon the way in which one may define the property. Ghiselli (1964, p. 18) captures this guideline: "variables defined in certain ways are not appropriate for certain kinds of individuals. A consideration of the nature of the individuals may require us to redefine the variable. Furthermore, the nature of the individuals may dictate the type of variable we conceive ours to be." Justification for this attention upon the nature of the individual is evident in the use of such terms as language minority student, relative language proficiency, language dominance, and bilingual discourse; in addition, the articles by Farhady (n.d.) and Garcia (1980) address this relationship of learner attributes and language proficiency.

These few guidelines pertaining to the definition of language proficiency were presented as suggestions to professionals and practitioners as they endeavor to assess the cognitive and linguistic capacities of different individuals. Once a clear and precise definition is formulated, one is in a position to develop operations that permit the observation of individal differences in the property (Note 4). This summary aptly describes the dynamic process of defining any important variable, especially language proficiency:

> the way in which we define a variable is a function of the concepts, the theories, and the knowledge we have about the property and the individuals for whom it is a variable. With one particular set of concepts, theories, and knowledge we might define a variable one way; whereas with different concepts, theories, and knowledge we might define it quite differently. As conceptualizations and theoretical formulations change and become more refined, and as our knowledge increases, our definitions of variables change so that what we once defined as a simple variable, we now see is a complex variable, and what we had taken as five different variables we now see is one variable. Variables are not static, unchanging, universal truisms. They are modified, given up, and created as our concepts, theories, and knowledge grow and develop. (Ghiselli, 1964, p. 19)

This quote accurately portrays the language proficiency assessment landscape of today.

Reference Notes

1. "Academic scholars" refers to those individuals of an academic community who generate a ***fundamental discipline*** (e.g., linguistics, psychology, mathematics) that consists of a body of knowledge made up of pure and distinctive forms of information pertaining to so. ? phenomenon; "academic," in this use of the term, means not constrained by practical consideration, or learned and scholarly, but not necessarily practical. "Professionals" refers to those individuals responsible for establishing ***applied or derivative fields of knowledge*** (e.g., medicine, law, engineering) that include those bodies of information composed of concepts and terms from the fundamental disciplines and cognate "fields" to deal with practical problems and phenomena found in those aspects of the human condition to which these specializations pertain. See J. E. Hill (1981) for a more detailed discussion of this important distinction.

2. These ten types of definition provide a jointly exhaustive pair, i.e. every definition is either complete or incomplete, nominal or real, etc. These types are defined as follows (Leonard, 1967, pps. 608-616):

   I. **Complete definition**: an act of definition intending completely to give the signification of its definiendum.
   Incomplete definition: one intended only partially to explain the signification of its definiendum.

   II. **Nominal**: one intended to explain what the definer means when he uses the definiendum.
   Real: one intended to explain the signification of a word or phrase as that word is used by authors other than the definer.

   III. **Informative**: one aimed at informing its receivers as to the meaning intended by authors who use the defined expression.
   Hortatory: one that recommends to its receivers that they adopt in their productive discourse the indicated meaning for the definiendum.

IV. **Linguistic:** one intended merely to explain that a certain word or phrase has such and such a presumably already familiar signification.
Conceptual: one that at least, in part, analyzes (instead of merely identifying) the meaning of its definiendum.

V. **Extensional:** one intended to explain (completely or incompletely) what objects are in the extension of the definiendum.
Intensional: one aimed at explaining what characteristics are in the total strict intension of the definiendum.

3. The framework of educational cognitive styles and the concept of matching styles with educational tasks are pertinent here. For more information, see Baecher (1976; 19B1a; 19B1b) and Hill (19B1) in the reference list.

4. For more information on the nature of operations and those of classification, ranking and measurement, see Ghiselli (1964).

Reference List

Baecher, R. "Bilingual Children and Educational Cognitive Style Analysis." In Simoes, A. (ed.), *The Bilingual Child: Research and Analysis of Existing Educational Themes*. N.Y.: Academic Press, 1976.

_____, "The Educational Cognitive Styles of Hispanic American Students: A Report." In Keller, G. and Fishman, J. (eds.), *Bilingual Education for Hispanic Students in the United States*. N.Y.: Teacher College Press, 19B1 (a).

_____, "Matching the Educational Cognitive Styles of Bilingual Students and Modes of Understanding Required by Bilingual Education." In Padilla, R. (ed.), *Ethnoperspectives in Bilingual Education Research: Bilingual Education Technology*. Ypsilanti, MI.: Eastern Michigan University, book in press, 19B1 (b).

Briere, E. J. "Are We Really Measuring Proficiency with Our Foreign Language Tests?" In Allen, H. B. and Campbell, R. N. (eds.), *Teaching English as a Second Language: A Book of Readings*. New York: McGraw-Hill Book Co., 1972.

Burt, M. and Dulay, H. "Some Guidelines for the Assessment of Oral Language Proficiency and Dominance." *TESOL Quarterly* Vol. 12, no. 2 (197B): 177-192.

Clark, J. L. D. "Theoretical and Technical Considerations in Oral Proficiency Testing." In Jones, R. L. and Spolsky, B. (eds.), *Testing Language Proficiency*. Virginia: Center for Applied Linguistics, 1975.

_____, "Direct vs. Semi-direct Tests of Speaking Ability." In Briere, E. J. and Hinofotis, F. (eds.), *New Concepts in Language Testing: Some Recent Studies*. Washington, D. C.: TESOL, 1979.

Cummins, J. "Educational Implications of Mother Tongue Maintenance in Minority-Language Groups." *Canadian Modern Language Review*, 34 (197B): 395-416.

_____, "Linguistic Interdependence and the Educational Development of Bilingual Children." *Review of Educational Research*, 49 (1979): 222-251.

DeAvila, E. and Duncan, S. "The Language Minority Child: A Psychological, Linguistic, and Social Analysis." In Alatis, J. E. (ed.), *Current Issues in Bilingual Education*. Georgetown University Round Table on Languages and Linguistics. Washington, D.C.: Georgetown University Press, 19B0.

Dieterich, T. and Freeman, C. *A Linguistic Guide to English Proficiency Testing in Schools. Language in Education: Theory and Practice Series #23*. Virginia: Center for Applied Linguistics, 1979.

Farhady, Hossein. "Measures of Language Proficiency from the Learner's Perspective." Unpublished manuscript, U. of California, Los Angeles, n.d.

Garcia, M. "Linguistic Proficiency: How Bilingual Discourse Can Show That a Child Has It." In Padilla, R. (ed.), *Ethnoperspectives in Bilingual Education Research, Vol. II: Theory in Bilingual Education*. Ypsilanti, MI: Eastern Michigan University, 19B0.

Ghiselli, E. *Theory of Psychological Measurement*. N.Y.: McGraw-Hill Book Co., 1964.

Hill, Joseph E. *The Educational Sciences: A Conceptual Framework*. Hill Education Sciences Research Foundation, P.O. Box 5053, West Bloomfield, MI, 4B033, 19B1.

Leonard, Henry S. *Principles of Reasoning: An Introduction to Logic, Methodology and the Theory of Signs*. N.Y.: Dover Publications, Inc. 1967.

Ryan, T. and Cruz, J. *Assessment: Humanizing Decision Making*. Washington, D. C.: American Association of Elementary-Kindergarten-Nursery Educators, 1974.

Upshur, J. "Functional Proficiency Theory and a Research Role for Language Tests." In Briere, E. J. and Hinofotis, F. (eds.), *New Concepts in Language Testing: Some Recent Studies*. Washington, D.C.: TESOL, 1979.

FOREIGN LANGUAGE AND BILINGUAL ASSESSMENT: ISSUES APPROACHES*

by
Protase Woodford
Educational Testing Service
Princeton, New Jersey

The National Defense Education Act (NDEA) was passed in 1958. The supposed threat of Soviet superiority in the sciences exemplified by the first successful orbiting of a man-made satellite in 1957 proved the catalyst for wide-ranging reforms or at least alterations in traditional academic programs — particularly in the secondary schools — in mathematics, the sciences and modern foreign languages.

Among the programs and areas receiving support through NDEA were:

1. Vocational Education Programs
2. Research and Experimentation in More Effective Utilization of Television, Radio, Motion Pictures and Related Media for Educational Purposes
3. Improvement of Statistical Services of State Education Agencies
4. Guidance, Counselling, and Testing: Identification and Encouragement of Able Students
5. Loans to Students in Institutions of Higher Education
6. National Defense Fellowships
7. Financial Assistance for Strengthening Science, Mathematics and Modern Foreign Language Instruction

By the end of the fourth year of operation, the National Defense Language Institute's program had provided training for nearly one-fifth of the elementary and secondary school foreign language teachers in the U.S. Over ten thousand teachers had received instruction in new methods of language instruction and had improved their foreign language fluency at 218 language institutes in the U.S. and abroad. There were foreign language institutes for Spanish, French, German, Russian, Italian, Chinese, Modern Hebrew and Japanese.

By June, 1962, 65,000 elementary school pupils were studying a foreign language as compared with 5,000 in 1959. In one Eastern state, foreign language enrollments in secondary schools were up 188%.

New methods were developed. The "new key" or "audio-lingual" method was "in"; "grammar translation" was "out"; and Glastonbury, Connecticut became a mecca for foreign language teachers. Millions were earmarked for language laboratories and new materials, both audio and visual.

There was, indeed, a revolution in foreign language teaching. New goals required new methods. The new methods required new resources — hardware and software. The revolution required new leaders — classroom teachers who could really understand and speak the languages they taught.

The goals of foreign language instruction before the 1960's had been limited, in most instances, to developing in the student some modest reading skills, mastery of grammatical rules and some translation ability. The objectives of the Spanish language classroom of 1955 were not dissimilar to those of the Greek or Latin classroom of 1800. Generations of students who could barely read a sentence in French, and who could speak not at all could, with ease, identify the negative imperative, the regular, irregular and perfectly healthy past participles, as well as other grammatical esoterica.

Foreign language teachers taught *about* language. In very rare cases did they actually *teach* language. This is not a criticism of previous generations of foreign language teachers. They knew what their goals were. They were disciplining the minds of their students. Communication in another language was *not* a goal. If someone really wanted to learn to speak French or German, she could jolly well go to Paris or Berlin, and not clutter up the foreign language class. Besides, who would really want to talk to foreigners anyway?

Testing has usually been a mirror of instruction. The goals and objectives of the instructional program should be reflected in the tests. Thirty years ago most foreign language tests were reading tests. When students wrote, they usually wrote in English. They wrote the English translation for a foreign language word or phrase.

A typical test consisted of a number of vocabulary questions like this. The student sees a word in Spanish (or French or German) and five English words or expressions. The student selects the best translation of the Spanish word or expression. Sometimes they were tricky. They reversed the process. The student looked at one English word or phrase and five words or phrases in the foreign language.

Then there were curious hybrid sentences. The student would see half a sentence in French and the other half in English. Then she would see five possible translations for the English half.

* r Woodford's paper was delivered by Ms Ines Bosworth

Students were also given passages to read in the foreign language. Parts of sentences were underlined and the students had to translate them. And always there were exercises where the student had to provide a label. Name the tense, the conjugation, part of speech or declension. These exercises may or may not have indicated whether a student could communicate in another language.

The good people who sponsored and carried out the NDEA foreign language program realized that new ways to test would have to be found if the new foreign language goals and objectives were to be properly assessed.

The Modern Language Association of America requested of the U.S. Office of Education funding for the development of tests in all four language skills: listening, speaking, reading and writing in French, German, Italian, Russian and Spanish at two levels for secondary schools and a similar battery of tests for teachers and advanced students. The tests for teachers and advanced students included measures of civilization and culture and applied linguistics.

The classroom tests were tried out with thousands of secondary school and lower division college students throughout the nation. The advanced tests were administered to foreign language majors and to participants at NDEA Language Institutes in the United State and abroad. For the first time, measures of oral skills formed an integral part of a foreign language test battery in a national testing program. For the first time, large numbers of high school and college foreign language students were *expected* to demonstrate their ability to understand and to speak the language they were studying.

The MLA tests were an important contribution to foreign language teaching per se, as reliable and valid measures of language performance. But more important, perhaps, was the catalytic effect the MLA tests had on all foreign language testing and, indirectly, on foreign language instruction.

The MLA testing program proved that it was feasible, if not always easy, to assess oral skills — listening and speaking — in a reliable manner, in classrooms from Hollywood to Harrisburg.

The College Entrance Examination Board introduced optional listening comprehension tests in French, German, Italian, Russian and Spanish. Shortly thereafter, composite listening/reading tests were incorporated in the admissions testing program.

The Advanced Placement Program of the College Board considered the development of "Language" Examinations in addition to the Foreign Language Literature Examinations. The late Paul Pimsleur, working with a leading publisher, developed batteries of tests in French, German and Spanish that provided measures in the four skills. Pimsleur also developed at that time a foreign language aptitude test battery, so, too, did John arroll and Stanley Sapon — The Modern Language ptitude Test (MLAT).

Foreign language textbooks that had been fixtures in school and college programs were thrown out and replaced by new texts accompanied by audio publishers' tapes, filmstrips and motion pictures. As part of the "packages," publishers offered tests. The tests, though often of questionable psychometric quality, *did* provide for testing listening and speaking skills in addition to reading and writing.

Listening comprehension tests appeared in the examination programs sponsored by the American Association of Teachers of French, of German, and of Spanish and of Portuguese.

The days of promise of the early 60's were followed by precipitous decline in foreign language enrollment of the late sixties and early seventies, the elimination of foreign language requirements, the virtual disappearance of language instruction in the elementary schools and apparent public disinterest in foreign languages.

Foreign language testing could not be unaffected by the general condition of foreign language education. The composite tests in the College Board program and the optional listening comprehension tests soon disappeared. The Modern Language Association classroom tests were never revised, nor were the Pimsleur Tests. From the mid-sixties to the present, there has been almost no new development in foreign language testing for the masses of secondary school and college students. Some significant work was done in the Advanced Placement Program of the College Board, particularly in tests of speaking and writing. However, the impact of the Advanced Placement Program, though great qualitatively, is less so quantitatively since it reaches fewer than 12,000 students annually. It is interesting to note, nevertheless, that the increase in volume for the Advanced Placement French and Spanish "language" examinations is significantly greater than that of the Advanced Placement Program in toto.

It is ironic that contemporaneous with the decline in volume and impact of academic foreign language programs, there has been a dramatic increase in the number and quality of foreign language programs outside the traditional school and college setting. The Peace Corps provided instruction in over 196 languages to over 27,000 volunteers between 1967 and 1981. The Foreign Service Institute, the Defense Language Institute, and other federal government agencies have taught common and uncommon languages to tens of thousands of their personnel. Proprietary language schools are thriving. Berlitz alone teaches 50 languages to an annual learner group numbering in excess of ten thousand.

While foreign language testing was nearly stagnant on campus, off-campus developments were taking place that could well have a profound effect on foreign language education in the future.

The Peace Corps, the military, the foreign service and the proprietary language schools were all under obligation, explicitly or implicitly, to verify the level of ability achieved by their clients. New ways to assess foreign

language performance were explored. Promising techniques were tried out. Many were abandoned because they were invalid, unreliable or simply too impractical for operational use

In assessment and measurement, there is regular reference made to "formative" and "summative" evaluation. In foreign language education, "formative" evaluation is usually the testing or evaluation of progress through a program of studies. "Formative" evaluation in the classroom takes the form of chapter quizzes and unit tests that are based on the content of the lesson plans, the course outline and the basic text. "Summative" evaluation should provide evidence of how well the learner has achieved the major goals or objectives of the program. In our field, summative evaluation should answer the basic questions:

> How well does the student understand and speak the language studied?
> How well does the student read and write the language studied?

In many foreign language programs, "summative" evaluation is simply formative evaluation on a grand scale. The final exam is simply a compilation of elements from the progress tests given throughout the course.

The danger inherent in evaluation of this sort is that successful performance on measures bound to, and reflective of, the course of study may not necessarily reflect real ability to communicate in the language.

That a student can memorize and recite some lines of dialogue and correctly replace elements in a pattern drill does not mean that she can meet basic needs through the medium of the foreign language in a real-life context.

One cause of the disillusion felt by foreign language learners and their consequent abandonment of academic language programs may well have been the lack of correspondence between academic grades and real ability to perform in the language. Whereas, in the school or college, the foreign language teacher is confined to a credit hour or semester system and is in direct competition with other disciplines for the student's time, in programs such as those of the Defense Department or the Foreign Service, the student's total effort is applied to the acquisition of language skills during a concentrated period of time — six hours a day, for example, for forty-four weeks. At the end of training, the learners must demonstrate ability to function in the foreign language in situations such as those they will encounter on the job. An "A" or a "C" is meaningless to an ambassador, a chief of mission or an artillery officer. Faced with the need to evaluate the actual performance ability of their graduates and the concomitant need to provide labels that would be readily understood by non-specialists, the linguists at the Language School of the Foreign Service Institute developed a rating scale to describe speaking ability and an interview-based evaluation procedure for assigning ratings. Although the scale and testing procedure were designed for use with their own graduates, the Foreign Service Institute linguists were called upon to administer their test and rate the performance of personnel from other government agencies such as AID and the Peace Corps. What is important is that the measure and the scale were equally valid when used with those who went through the FSI course and with those who learned the language at home, in school or overseas. The interview and rating constituted a summative evaluation procedure independent of any program of studies, a procedure that established as the criterion the student's performance compared to the speaking or oral interaction ability to be expected of an educated native speaker.

Performance-based evaluation of foreign language skills, as exemplified by the language proficiency interview, has been adopted by agencies within and without the federal government. State and municipal education agencies have chosen the interview and scale for purposes of bilingual teacher certification. The Province of New Brunswick (Canada) uses the interview with secondary school students for evaluation of the provincial second language (French/English) program. Missionary groups, international student exchange programs and major industrial organizations require the interview for assignment overseas.

Unfortunately, evaluation procedures such as the Language Proficiency Interview are still little used in traditional academic foreign language programs.

In bilingual education, the need for evaluation of language performance in a real-life context is of paramount importance.

In assessing the English language skills of children with limited proficiency in the language, the need is to determine how well each child can function through the medium of English in the classroom.

While many of the test exercises used in the foreign language classroom have application in the bilingual classroom, there are indeed some major differences.

The most obvious difference is in the expected use of language skills. In most foreign language classrooms, the time lag between instruction and real life application of skills may range from months to never. In the bilingual context, there is usually immediate and almost simultaneous application of newly acquired skills. Concomitant with the enhancement of motivation that comes with immediate gratification, there is the need for appropriate criteria. In the traditional foreign language classroom, the criteria against which student performance was measured often had little to do with communicative ability. In the bilingual and ESL classroom, the criteria are real-life communicative criteria exemplified by what is going on in the English-medium classroom down the hall. The ultimate test is whether the youngster can function in *that* classroom.

The major measurement needs in bilingual education remain the same as they have been for over a decade:

1. A means to determine whether a child should be placed in a bilingual program or not,

2. Procedures for determining mastery of subject matter in the bilingual classroom. (The whole issue of the optimum format for subject-matter tests in the bilingual classroom has yet to be resolved.)
3. Reliable determiners of *when* a youngster is ready to profit from instruction given in the English language.

Over a decade ago, Theodore Andersson and Mildred Boyer predicted that within ten years bilingual education would have to prove itself to its critics.

Irrefutable evidence of the efficacy of bilingual programs would have to be presented to the uncommitted. The ten years have passed. We have had the AIR report that proponents of bilingual education all reject. Yet, we have little or no evidence to present in support of our programs.

Reliable evaluation and valid tests are an absolute necessity for bilingual education. We cannot use a questionable test that shows our children to be reading dramatically below grade level as justification for funding and then deny the validity of the same test when it indicates that our "treatment" of the children has had no effect or even a deleterious effect. We can't have it both ways.

Assessment or evaluation is necessary to recognize deficiencies in a program and to discover strengths and weaknesses in students or in teachers. The information gained from such assessment should be used to improve instruction and meet the needs of children better. The need for evaluation and assessment is clear, but the procedures, the means, and the instruments, present problems. To evaluate a program, the goals and objectives of the program must not only be stated, but must be stated in such a way that their attainment is clear and measurable. Affective goals are particularly difficult to define in a measurable way. If there is no way to determine whether goals have been achieved, the goals themselves tend to be no more than words. "Improved self-image" and "better concept of self" are laudable objectives, but are of little value to a program unless behavioral manifestations of the improved self-concept and image can be perceived and evaluated. To evaluate a program, a variety of assessment tasks must be carried out — aptitudes and abilities of children, mastery of subject-matter content, language proficiency (in both English and the home language and in all four skills, if appropriate), and attitudes. The measurement of aptitudes and mental abilities or IQ of children who are not middle-class members of the dominant culture has long been under attack. The practice of placing children into classes for the educable mentally retarded on the basis of an IQ test administered in English, when their home language is other than English, has received widespread publicity and general opprobrium.

If English-language IQ tests are inappropriate measures for non-English-speaking children so, too, can be tests administered in the child's home language. Such tests may simply be translations of English-language instruments with cultural irrelevancies intact, or they may be instruments designed for children who, though sharing essentially the same language, come from a very different cultural background. Before any foreign tests are considered for use with non-English-speaking children, they should first be scrutinized to determine the extent of their linguistic and cultural "fit" for the population to be tested.

Children in a bilingual program must be tested routinely in various content areas to assess their progress. The language medium for testing should be the same as the medium for instruction. If children are taught social studies in French, the social studies test should be in French as well. Although this is recommended for early grades, there are indications that once linguistic competence is attained in the second language, transfer of information across languages can readily be accomplished.

Subject-matter tests are normally in one language and limited to two skills, reading and writing. If bilinguals have an unequally developed set of skills in two languages, it might be desirable to create measures in skill modalities that would conform to the linguistic strengths of the examinee. For example, if a student has a fully developed oral ability in Navajo, but no reading or writing ability in that language, and has moderate oral skills in English and moderate reading and writing skills in English, the optimum test modalities for him might be spoken Navajo and printed English. If bilingual students require evaluation of their attainment in subject-matter areas, they also require continuous evaluation of their linguistic development, both in English and in the home language. This evaluation is essential for decision making at various times and for various purposes. An initial assessment of language dominance is needed to determine whether there is a need for a program of instruction in the home language. This is done in some programs by means of a questionnaire to parents. Reliance on indirect indicators of language dominance such as surnames, ethnicity, and parents' language dominance is questionable.

The most valid measure of language dominance is obtained through an assessment of performance in the language. Such assessments, despite their high validity, are open to questions of reliability because they depend on a rater's judgment.

Evaluation of children's language abilities in English and in the home language should be carried out routinely in all language skills. Of particular importance is the determination of when the bilingual child is ready linguistically to enter the regular English-medium stream in the transitional bilingual program. In the bilingual education legislation of one state appears the statement that the child of limited English-speaking ability must be enrolled in a bilingual education program for three years "or until such time as he/she achieves a level of English language skills which will enable him/her to perform successfully in classes in which instruction is given only in English, whichever shall first occur." Legislation in a number of states requires the establishment of a bilingual program only if there are 20 or more children of limited English-speaking ability be-

longing to a specific home language group. Problems with finances and staffing can lead administrators to attempt to mainstream bilingual children as soon as possible. Because legislation does not specify the procedure used to determine that a child has achieved "a level of English language skills that will enable him/her to perform successfully in classes in which instruction is given only in English," great care must be taken to ensure that the child is indeed ready. What must be measured is the bilingual student's acquisition of the subject-matter content prerequisite for success in the English-medium class and ability in English in the four skills necessary for functioning in the classroom. The student's English language readiness must be assessed against a real-life criterion. This criterion might be those English language abilities in understanding, speaking, reading, and writing that are possessed by monolingual English-speaking students of the same age and grade who are at least minimally competent in the English-medium class. What is needed is an analysis of minimum requirements in English listening comprehension, speaking, reading, and writing for success in each of the grades from kindergarten on. Instruments for assessing reading ability and writing for monolingual English speakers at different grade levels exist and may be useful with bilingual students. Some measures for listening comprehension and speaking ability have been developed. Much work still needs to be done to determine the validity and reliability of such measures.

With any standardized tests, great care must be taken to determine their appropriateness with regard to the specific population with which they will be used. Some of the questions that should be asked are: Is the content sampled in the test representative of the classroom content of the program? What are the characteristics of the reference population (the children in the norming sample)? Does the test require a skill that the children have not developed? (For example, must they be able to read in the home language to answer questions when they might not have been taught to read yet?) Is the variety of the home language used on the test familiar to the children? Will the information to be gotten from the test be useful?

Virtually all assessment of language skills in foreign language and bilingual education is being done by those involved in the training process. They use instruments of their own design, or they use tests designed by the producers of the teaching materials — the textbook publishers. In a Mexican university the other day, the head of the Engineering Department said to me that education is one of the few fields in which the manufacturing and quality control are done by the same person. Is there anything wrong with that? There can be. There is no problem with that kind of measurement if the content of the course or the textbook really teaches language skills and the tests measure real language. There is a problem, however, if the course of textbook content does *not* lead to development of real language skills. The test, then, may accurately measure mastery of the course, and *not* tell us a thing about how well the examinee can communicate. The student who does well in the French course gets an "A." The student who does poorly in the course gets a "D." The assumption is that the "A" student speaks and understands French better than the "D" student. Maybe, if the final examination is a reading test only, then it is quite possible that a student with an "A" might not speak a word of French. A native Frenchman who was illiterate would fail. One of the most damaging practices in language teaching has been the "internal accountability" system. If I teach Chinese and I alone evaluate my students, then I essentially define what is the Chinese language.

For years American students studied foreign languages. They spent two, perhaps three, years in a foreign language class. They memorized their word lists. They learned their verb conjugations. They received their A's and B's. They fulfilled a foreign language requirement for a college or university. They *thought* they had learned another language. Heywood Broun, the humorist, once said that he had received an "A" in Beginning French. But when he got to Paris, he found that nobody there spoke Beginning French.

These remarks are not a criticism of foreign language teachers. They did not develop accomplished speakers of other languages because they simply could not. Government agencies that teach foreign languages have determined that it takes from 250 to 400 hours of intensive instruction to reach a survival level. 250 to 400 hours to be able to order a simple meal or to ask directions. It takes almost four years of high school language study to accumulate 400 hours of foreign language instruction. In four years the high school student is expected to learn to understand, speak, write and read the classics in the language and know the history, the culture, the mountains, rivers and lakes of the country whose language he is learning. Impossible!

We are now enjoying a reawakening of interest in foreign languages on the part of government, commerce and the general public. There is the promise of federal funds. In the 1960's, competence in foreign languages was considered essential were we to compete with the Soviet Union in outer space and in the battle to win over the hearts and minds of the uncommitted around the world. Today we are told that competence in other languages is essential if we are to compete with the Empire of Japan in the market place. Whatever the motivation for it, we are all pleased with the attention given us and our mission, and we hope for more tangible evidence of interest such as the NDEA programs of the 1960's.

This time, however, we will need to be more cautious when we accept the gifts. Because, even though it may not be obvious, there will be a string attached. The string is an expectation that this time we really will teach people to communicate effectively in languages other than English.

In order to understand what our needs are, in order to monitor student progress in the acquisition of foreign language skills, in order to assess the level of mastery of individual learners and in order to determine the kind of language performance requisite for different tasks and

occupations, we will need reliable, valid and practical measurement tools.

There are three critical needs that must be addressed:

1. Improvement of the test development and test interpretation skills of classroom teachers;
2. Development of tests of receptive skills — listening comprehension and reading for national administration;
3. Development, adoption and dissemination of common descriptors of language performance.

Most language teachers have had little or no training in test development and test interpretation beyond an introductory statistics course. The number of foreign language testing courses offered in the United States is insignificant. Yet, major decisions regarding students and programs are made on the basis of tests that are often unreliable and inappropriate.

In the training or retraining of bilingual and foreign language teachers, provision must be made for the development of testing expertise. The focus should be on how to design tests to reflect the objectives of the curriculum and how to create the appropriate measures for testing each of the language skills. In addition, teachers should be taught how to review external tests to determine their appropriateness and how to interpret standardized test results. Teacher training institutions should be urged to create language testing courses and should be given the necessary resources to initiate such courses. Testing should be included as a major area in language teachers' in-service training programs.

The foreign language test batteries developed in the 1960's are in desperate need of revision or replacement. Skills of reception — listening comprehension and reading — are most amenable to large-scale objective testing procedures. It is recommended that listening and reading tests covering a wide range of abilities be developed in the most commonly taught languages. If the range of abilities tested is broad enough, then the same tests can be used in secondary school, college and university from beginning through advanced levels of language training. Separate norms for specific sub-groups can be developed to provide for meaningful interpretation of scores. Multiple forms or versions of the tests could be developed periodically. Older versions could be made available to individuals for self-testing. Longitudinal studies of language development would be facilitated by having the same kinds of data available year after year. Significant economies might be effected by making unnecessary a proliferation of language tests for different programs.

Too often we don't know what we mean when we make statements regarding language ability. We have no way of describing our "product." There is a real need for all of us involved in foreign languages as teachers, as students, as administrators and evaluators, to have available a common measure, a common, accepted set of descriptors of foreign or second language ability, so that everyone will know what we mean when we put a label on someone's language performance. The "A"s, the "B"s and "C"s, the Superiors, Goods, Averages and Poors, the 95's and 80's and 60's obviously have all lost more than something in translation. We, the language specialists, whose raison d'etre is communication, have failed abysmally in communicating with one another.

The idea of common standards or common, universally accepted criteria may be abhorrent to some. It may smack of regimentation or an abrogation of academic freedom. On the other hand, the adoption of criteria that are comprehensible to all will allow us to be realistic in our aspirations. Here is a sample level description from the scale used by the Peace Corps, the Foreign Service and others. This is the criterion for Level 1: "Able to satisfy routine travel needs and minimum courtesy requirements."

The Level 1 criterion is further elaborated as follows: "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.

If a person cannot ask for shelter or lodging, cannot order a simple meal or make a purchase, that person has *not* fulfilled the requirement for Level 1.

The Foreign Service estimated that it would take some 360 hours of intensive training in Russian for a learner of average aptitude to attain Level 1. This estimate was made on the basis of long experience with large groups of learners all evaluated by the same procedure. It is very difficult to argue that they should be able to develop Level 2 ability speakers in 200, 300 or 500 hours when solid evidence to the contrary exists.

In the academic world, we have the freedom to give grades or credits based on criteria that we ourselves determine. Our clients, however, have different expectations. They have been led to believe that if one receives high marks for two or three years of language study that somehow she will be able to speak French or German or Portuguese. In fact, were we to compare contact hours in school or college programs with intensive government programs, it would be evident that it would take about three years of high school study to achieve a Level 1 in Spanish, French or German and about four years in Hebrew or Russian.

Adoption of a common scale will allow us to plan efficiently and to set forth clear, attainable goals. Is it not obvious that in the time required in an intensive language program with small groups of highly motivated students under optimum conditions for learners to

achieve a survival level of performance, we *cannot*, under far less than optimum conditions, have our students achieve even higher levels of communicative ability in addition to an acquaintance with and appreciation of the literature and culture of the people whose language is being studied?

School administrators, students, parents and the public at large must know that it is nonsense to expect us to give students even basic survival skills in another language in two short years. We must be able to show people what can be expected, realistically, in two years or three or four; in six or eight semesters. In order to do that, however, we need to be able to describe language performance in terms that everyone can understand.

A great deal of work in the development and refinement of language performance scales has been carried out by various federal government agencies. Recently, a group of linguists and foreign language teachers from Great Britain, the United States, Japan and Germany, under sponsorship of the English Speaking Union, the British Council and Educational Testing Service, have met in Princeton and London to review existing scales and to suggest further refinements.

Common descriptors of language performance facilitate the determination of minimum standards of language performance for specific tasks.

The State of New Jersey wanted to determine what level of ability in English and in the home languages of children in bilingual programs should be expected of bilingual teacher candidates. In order to answer the critical question of "How good is good enough?", the following study was carried out.

Five-minute segments from twenty recorded and rated language proficiency interviews were selected. Only performances rated between 1+ and 4+ were included. It was assumed that no performance below 1+ could possibly be acceptable and level 5 performances would all be acceptable. The segments were ordered as follows: 2, 4, 4+, 3, 1+, 2+, 3+, 3, 4+, 3+, 2+, 2, 4, 1+, 2+, 4, 2, 3+, 1+, 3. Eleven judges for English bilingual and eleven for Spanish bilingual received instructions emphasizing that their task was to judge whether the speaking proficiency of the person being interviewed in each segment was "at least minimally sufficient for this person to function adequately" in a bilingual classroom.

The judges were, themselves, bilingual teachers, supervisors of bilingual programs and administrators in bilingual schools. The judges, of course, were not informed of the ratings and were instructed to judge each sample independently and without discussion.

The results of the study indicated that there was a critical level of ability below which most judges indicated that the performance was inadequate.

Similar studies could be carried out for any number of job categories — travel agents, flight attendants, tour guides.

We appear to be on the threshold of another promising era for language education. We must not repeat the errors of the past. We must know where we are and where we are going. We must have our goals set forth clearly and comprehensibly. Our objectives must be attainable.

Effective measurement and evaluation will play a critical role in the process.

INTEGRATING LANGUAGE ASSESSMENT WITH TEACHING PERFORMANCE IN SUBJECT AREAS

by
Gary D. Keller
Graduate School
Eastern Michigan University

Introduction

The title of this article is somewhat ambiguous and requires further definition. Specifically, what will be treated here is the following question: How do we test bilingual education teachers in order to be sure that they can teach bilingually in subject areas such as mathematics, science, social studies, and so on? In order to illustrate the topic, I will use examples taken mostly from Spanish/English bilingual education. The same issues are pertinent, nevertheless, to most, if not all, of the other languages which are encompassed within the bilingual education movement (both through Title VII and in the parochial school sector) in the United States. An additional observation must be made at this time: to the knowledge of this researcher (who has conducted a rather extensive, but not necessarily exhaustive, search of the literature), the specific topic at hand has not been investigated.

This is not to say that the general field has not been well addressed by researchers including linguists, educators, psychologists, sociologists, and other social scientists. It has. The bibliography refers to a number of important works which impact generally on the topic at hand, such as Oller and Perkings (1980), Jones and Spolsky (1975), and Clark (1978), all of which treat in detail the theoretical and practical aspects of the language assessment of teachers. Additionally, Oller and Perkins (1978) review the questions of how to assess the assessment instruments themselves in terms of their validity and effectiveness; a paper by Gonzalez (1980) treats the use of English as a Second Language materials in mathematics education (but only focusing on curriculum materials for the student and not an assessment of the teacher's competency); Zamora (1981) treats the issue of language instruction for bilinguals at the college level; and Keller (see bibliography) has treated in some detail the question of choice of register in the bilingual classroom. Nevertheless the observation stands that to the best knowledge of this writer the topic of assessing language competencies in bilingual teachers of subject areas such as mathematics and science, is addressed for the first time here. Consequently, as with any paper that attempts to chart new territory, I am constrained to a rather high level of theory, speculation and issue identification, some of which it is hoped will provide heuristic value to the specified studies which are sure to follow in the coming years.

This paper will contain four parts: a discussion of the issue of the appropriate language variety or register for the bilingual education content area teacher, which at first blush would appear to be an easily resolved matter; a description of some issues of pertinence in the selection of the types of tests which are generally available for language assessment; an introduction to the ideal qualities that an assessment instrument for evaluating bilingual education content area teachers ought to have; and finally, a return to the critical problem of a lack of a clear or well-defined language variety or register in subject-area bilingual education.

What Language Variety or Register Should Be Assessed in the Bilingual Subject-Area Teacher?

A problem which has existed in bilingual education generally is which variety of a language (in this case, for illustrative purposes, Spanish) should be used in the classroom?

The answer has been often made in the form of one of two extremes. There are those who exalt the ethnic form of their locality and denigrate what the American Association of Teachers of Spanish and Portuguese (AATSP) has called "world standard Spanish." Conversely, there are those who exalt "world standard Spanish" and denigrate the ethnic or folk form. The first group is often found in ethnic studies departments on the college campus or in alternative colleges, of which there are 15 to 20 mainly in the Southwest, and among the ethnic communities themselves, particularly among radical spokespersons for the Chicano and Boricua communities.

The second group has been well described by Rolf Kjolseth. It includes the majority of Spanish teachers, both nonethnic and ethnic. Often the ethnic Spanish teacher fears the ethnic variety of the language as the stigma from which he or she has only recently escaped.

Of course, I am overgeneralizing the dichotomy. For example, the AATSP, the professional group of most importance in the United States, has made very sensible statements with respect to the potential domains of ethnic varieties of Spanish versus world standard Spanish. This is so because the AATSP has turned the question over to its professional linguists for public comment. However, the general membership of the AATSP tilts toward negative attitudes about United States vernaculars of Spanish.

Let us describe very briefly the course that can be taken by those who exalt the vernacular at the expense of the standard form. Lozano (1974) claims that "the regional varieties of Spanish in Mexico and the (United States) Southwest which share vitually the same morphosyntactic characteristics should be considered a binational macrodialect." (p. 147). Lozano adduces some linguistic reasons for subsuming Southwestern United States Spanish and Mexican Spanish, but in the sociopolitical arena, his conclusions are nil.

The Chicano is engaged in combat not only with the "Anglo establishment," but with the disapproving "Mexican establishment" of which even such a distinguished Mexican linguist as Antonio Alatorre (1955) may be taken as a representative voice. Alatorre compares the Chicanos to the *mozarabes* of medieval Spain, intimating that the former, like the latter, have served to introduce many foreignisms into Spanish. He defines the Chicano (except that he uses the term *pocho* which is pejorative in Mexico) as a Mexican who permits himself to be seduced by the American way of life and for whom Mexican ways are contemptible and American ways unsurpassable. As for the language, it is the product of a border society "that has created a type of dialect or creole in which elements of English and Spanish are fused." (pp. 11-15)

Naturally, the Chicano, when confronted with these sorts of stark expressions of prejudice on the part of Mexicans, is compelled to systematically minimize the Mexican element in the language and systematically single out that which is autochthonous.

Let us look at the same phenomenon at the micro, rather than the macro, level. Troike (1968) points out that "there are, in fact, several native dialects of Spanish spoken in Texas alone—even in a single city such as San Antonio or El Paso—and most of these are simply local varieties of the much larger regional dialect of North Mexican Spanish." Troike goes on to observe the classroom implications of these differences, but once again, the chances of implementing the fact of different subdialects in the Southwest into either a coherent corpus plan or a classroom pedagogy are obstructed by the overriding ideological exigency that Chicano Spanish is the vehicle of Chicano self-identify. Thus, Fishman's observation (1968) that while often linguistically salient differences go unnoted in society while others, perhaps of a more minor nature, are socially pressed because divisiveness is an ideologized position, must be recognized as a sociolinguistic fact, a fundamental consideration in analyzing the development of in-group attitudes toward the vernacular. Moreover, these attitudes have had a clear impact on the educational process. For example, Gaarder (1977) attests to the fact that numerous Chicano Studies programs actively denigrate what he calls "world standard Spanish" and insist that for their purposes the only languages needed are English and "Barrio Spanish"

Let us turn now to the second group. Those who exalt the standard and denigrate the vernacular are a particularly vexatious lot. I shall be brief here because of the widespread familiarity that linguists have with the types of arguments, pedogogically, or politically, that are made in this regard. Kjolseth (1972) and also Steiner (1969, 212-213) have characterized this purist approach which posits a single variety of language, or of culture for that matter, as "correct" as one which involves "de-education," that is, the belief that the lower class Chicano, Boricua, or Cuban-American child has to be de-educated before he or she can be re-educated. Kjolseth's graphic conclusion is strong, but valid: these sorts of people liken themselves to priests of education busily civilizing the savages. In fact, they are engaged in a type of self-arrogating educational colonialism.

In addition, I must note that there is in the United States a special intense hostility directed toward non-English vernaculars which linguistically display the evidence of daily English language contact. Specifically, I am referring to what in this country is popularly termed as "Spanglish"—often used as a buzz word by "vernacular denigrators," analogical in its extreme negative connotations to terms such as barrio Spanish, *calo*, *pocho*, Nuyorican, or Rican, which are often proffered as terms descriptive of a new, vibrant language emerging from the United States Hispanic ethnic communities by "vernacular exalters." Those who use the descriptive term Spanglish for politico-pedagogical purposes often take it to mean a pidgin or a hybrid, the illegitimate fruit of English-Spanish contact.

It is necessary, however, that we recognize the profoundly psychopolitical motivation behind the manifest antagonism to United States vernacular Spanish. Alfonso Reyes, the Mexican thinker, once observed about his people, "Pity us who are so removed from God and so near the United States." It is especially painful for many Spanish-speaking persons to see the English language affect Spanish because of the obvious analogues with imperialistic exploitation. Those feelings of frustration are linguistically spurious, but are psychologically intense. Yet, while we can understand some of the motivation as attributable to a feeling of anger and frustration directed toward what is seen as another instance of United States domination—what we cannot accept is the result: stigmatizing both United States vernacular Spanish and its legitimate speakers.

Between the two extremes, there are those who chart a middle course. These tend to be sociolinguists and include such persons as Gumperz, Dillard, Kjolseth, G. Valdes, and Keller. For the sociolinguists, the solution tends to be the fostering of bidialectalism. The world standard variety is to be added to the ethnic variety that the child already brings to the classroom.

This tug and pull with respect to differing language varieties as the medium of classroom instruction within bilingual education has had a profound effect on the actual creation of curriculum materials. In 1974, the National Institute of Education funded a project evaluating approximately 1,000 curriculum titles in Spanish bilingual education. I was the linguist for that project, the results of which were later published by the Educational Products Information Exchange Institute (popularly known as EPIE). The eight types of Spanish that we found to be in actual existence in 1974 in Spanish bilingual education programs were as follows.

Types of Spanish Used in Instructional Programs for Bilingual Education in the United States

1. Programs which use "world standard Spanish." The language is free of regionalisms. Some of the language may not be understood by United States Spanish speakers who use a regional or ethnic designation instead of the standard one (e.g., program uses *autobús*, but not *camión*-SW; nor *guagua*-NE).

2. Programs which use language specific to particular regions or social groups of the Hispanic world outside of the United States, such as Spain, Bolivia, or Chile. For example, these programs may use *micro* (Chile) or *autocar* (Spain), but not *autobús, camión* or *guagua*.

3. Programs which use language characteristic of all the regions and ethnic varieties of United States Spanish. (e.g., program uses *guagua* and *camión*, but not *autobús*).

4. Programs which use language characteristic of the eastern United States and the Caribbean (e.g., program uses *guagua*, but not *autobús* nor *camión*).

5. Programs which use language characteristic of the western United States and Mexico (e.g., program uses *camión*, but not *guagua* nor *autobús*).

6. Programs which use non-standard, non-Spanish (as in bad translations).

7. Programs which use both the regional or ethnic varieties of language as well as the "world standard Spanish" variety. (e.g., program uses *camión* and *guagua* in addition to *autobús*).

8. Programs which use controlled "world standard Spanish," using only language in the standard for which there are no alternate regionalism or ethnic varieties (e.g., program eliminates *camión, guagua* and *autobús* from instructional materials).

Clearly there are types of Spanish now in use which are totally inappropriate with respect to United States bilingual education. These include type 2 and of course, type 6, which tends to be a bad translation of an English-language program. Yet even once we discard these two types, there surely remain too many corpora. In addition, the viable corpora that do exist are often found to overlap. They require the appropriate "compartmentalization," to use a term that Fishman has advanced. Type 1, which is a rather common language variety in which bilingual education materials are published, in my mind, is most compellingly used for the type of bilingual instruction we are discussing here: the content areas such as mathematics, the natural sciences, health, and so on.

Types 4 and 5, for the relevant regions, recommend themselves for employment in transitional bidialectal education. When students enter the school system with only a knowledge of their ethnic or regional variety, it is logical to build upon their knowledge, at least for the first year or two, by teaching them what they don't know on the basis of what they are competent in. This is particularly true with respect to the pedagogy of beginning language arts, with its extensive use of sound-symbol and picture-symbol matching techniques, all of which are short-circulated when a child uses the ethnic term instead of the standard one, when the latter is expected by the pedagogy: for example, when the book expects the child to say *puerco* and the childs says *chancho* instead

On the other hand, as a result of my participation in this massive evaluation, I hypothesized an eighth type, which at that time did not exist. Subsequently, a number of programs have been written in type 8, including one of my own. I believe that they successfully deal with the miscues that crop up otherwise and therefore, are able to teach decoding, encoding, word-attack and word analysis skills in world standard Spanish without interference effects from the ethnic or regional variety.

Finally, I find it hard to rationalize the use of either program types 3 or 7. Under the guise of completeness or fairness, they offer mind-boggling numbers of synonyms for the same meaning. In this sort of program, to give one simple example, northeastern children are bombarded with southwestern Spanish in semantic areas such as the desert, agricultural communities, the mountains, the mines, which are simply irrelevant to them; the converse is true as well. To give an example, southwestern children learn the words *plátano, guineo, plátano dedo*, and other plants from the banana family, when for their language and culture merely one term suffices. Unfortunately, this trend is being exacerbated.

Attempting to be all things to all groups, publishers have tended to supplant synonymy or dialectal equivalence instead of pedagogical logic. Of course, from the publisher's point of view, it is wise to print a national edition, one that will sell everywhere. Perhaps the culmination of this trend can be found in the Santillana program, *Aprendiendo en dos idiomas*, a slick reading program which is marred, for me, by the fact that the reading lessons feature a teacher pointing out the lexical varieties for different meanings in different parts of the United States and abroad as well. I can't fathom to what positive purpose we should teach first and second graders eight synonyms for ball point pen and ten for bus when what we should really be engaged in is the expansion of their vocabulary to meanings that are totally unknown to them. To the extent that this synonymy proliferates in the classroom, we sacrifice language development.

I have stated earlier that Type 1 bilingual instructional materials, those which use world standard Spanish, would appear to be those most appropriate for use in such bilingual education content areas as mathematics, the natural or exact sciences, health, and so on. This would appear to be, at first blush, a fairly obvious point. On the one hand, the use of world standard Spanish is justified on the grounds that the applicability of the content areas transcends any regional, sectarian or other circumscribed linguistic, cultural or political entity. One does not think of Chicano mathematics, *boricua* biology, etc., and therefore, the use of world standard Spanish as the medium of instruction for mathematics or biology is justified by the universality, the transcendence from ethnicity of these content areas. On the other hand, and this is part and parcel of the same observation, there are very few, if any, ethnically marked lexical or other linguistic items which relate to the content areas. Linguists judge the domains of the vernacular varieties to be centered in such areas as family, peer and kinship relations, church, popular art (*dichos, corridos*, etc.). Science and mathematics have most emphatically been

the domains of the standard, and thus there are no competing language varieties in the vernacular for the lexicon of these fields. Since there is nothing for the standard to compete with in terms of vernacular varieties, the question of which language variety to use would appear to be moot. (It isn't as we shall see later.) The linguisitic base for instruction in the content areas has to be the world standard language variety, for there is nothing else.

So far I have limited myself to the language variety in which instruction should proceed, and I have arrived at the conclusion that it should be the standard. However, the topic of this paper is the language variety in which bilingual content area teachers should be assessed. Here the same arguments that I have just been making are even further heightened. Indeed, while I don't believe that it is either practical or profitable, one could make the argument, in theory, that the standard lexicon of the content areas, having no competing lexicon in the vernacular could be, for the purposes of instruction, in say, a transitional bidialectal education program at the early grade levels, introduced in common vernacular trappings. Thus, for first and second graders, early science and mathematics vocabulary from standard Spanish could be presented in the language (e.g. vernacular) with which the students are acquainted.

This is all a rather speculative point that I am discussing, however, because it presupposes a vernacular grammar in which standard content lexicon would be embedded. For Spanish at least, the forte of the vernacular is lexicon. There is very little difference in grammar between standard Spanish and the many vernacular varieties.

Nevertheless, to move forward, whatever attraction that a speculative instructional format using vernacular grammar and standard content area lexicon has for the young child, even this totally breaks down with respect to the issue of assessing the bilingual education teacher. Clearly the content area teacher would have to know the standard variety of the language, both to have been trained and to be able to communicate with peers in the content (e.g. mathematics, chemistry, etc.) field. Ergo, the obvious conclusion is that the assessment instrument that would evaluate the linguistic proficiency of bilingual content teachers would, of necessity, be cast in the standard variety of the language. Although the bilingual education teacher must know much more than merely the standard in order to understand and/or teach the language arts and culture of the bilingual child, for the mere purpose of assessing that teacher's competency in subject areas such as mathematics or science, an assessment instrument that is expressed in the standard is required.

While we have arrived at a rather straightforward conclusion that the assessment of bilingual education content teachers should be in the standard, we have, unfortunately not exhausted the topic. We shall return to this question of language variety later in this paper. For while it is true that Spanish at least does not have significant lexical competition between vernacular and standard with respect to the content areas of the bilingual education curriculum, the proliferation of scientific lexicon (some less charitable souls might call it jargon) translated from English into standard Spanish, and for which in Spanish no equivalence existed, has caused difficulty for the bilingual educator and will cause difficulty for those whose responsibility it is to assess the linguistic competency of the bilingual educator.

A Review of Issues Dealing with Test Formats for Assessing the Bilingual Educator of Content Areas

Having briefly outlined a critical language issue to be considered in the construction of a proficiency test, let us turn our attention to test formats. One of the first distinctions deals with the issue of "discrete point" versus "integrative" tests. The discrete point test assumes that all of the tasks and topics involved in language learning can be disentangled, listed, tested separately, discretely as it were. Phonemes, morphemes (free or bound), syntagms, lexicon, idioms, etc., are tested separately. The advantage of such a format is that individual learning problems can be separated out and assessed without the confounding of intervening, "fudge" variables. A discrete point test is diagnostic in nature. It permits us to ascertain if the testee has acquired each unit of the language. Discrete point testing finds some of its basis in native language interference research (Lado, 1961) and in contrastive analysis.

Integrative test formats are the natural opposite of the discrete format. Integrative test items combine numerous units of language into a single assessment, and thus, their claim is to greater validity. Performance on an integrative item is more closely tied to contextual restraints, as in the cloze test in which, for example, a paragraph is systematically mutilated by deleting every seventh word. (The examinee has to fill in the blanks with the correct word or an acceptable substitute.) Contrast the integrative cloze test with the tin/thin sound discrimination task, a discrete point item appearing on the *Language Assessment Scales* (LAS). It is readily apparent that the integrative test item involves a great many units of language, not merely one. This is both an advantage and a disadvantage. The advantage lies in validity, in the fact that in real conversation successful performance involves the use of numerous units of language in a single utterance. On the other hand, the discrete point test is especially useful to the evaluator in that one can identify each unit of language which has not been learned by the testee. While the integrative test may be a better indicator of language learning, the source of error or lack of competence can not be easily identified.

Another element surrounding the subject of test formats relates to the issue of natural versus unnatural language. Natural language involves the use of structured and nonstructured communication tasks. Unnatural language involves the use of various types of linguistic manipulation tasks. Nonstructured communication involves a conversation between the student and the examiner, as is required on the *Basic Inventory of Natural Language* (BINL). Structured communication involves student responses to specific questions asked by

an examiner in a pragmatic context. An example of this format is the *Bilingual Syntax Measure* (BSM).

Current bias runs heavily in favor of natural unstructured communication since this is the type of communication the speaker participates in daily. There are certain disadvantages to nonstructured natural communication tasks, however. A great deal of speech must be elicited in order for a sufficient range of language structures to be obtained. Also, it is not possible to make judgments as to the testee's command of any structures not obtained since the situation discussed may not have called for them. Performance on such tasks is often dependent on personality since some testees are much more willing to talk expansively than others.

Structured communication tasks seem to solve most of these problems. One can diagnose the testee's ability to use the conditional or conditional perfect, for example, by posing a question the answer to which will require such tenses. Thus, on the *Bilingual Syntax Measure*, we find questions like "What would have happened if the dog hadn't eaten the food?" The most common response is "The king would have eaten it." Through the careful development of appropriate pictorial situations, it is actually possible to construct a diagnostic profile of the child's internalization of numerous structures and to use the profile as a checklist. A disadvantage of structured communication is that with this format it is not readily possible to ask questions which will elicit all structures. For instance, how would one elicit a question, a very common structure in conversation?

Another concern surrounding structured communication is that it produces an inflated indication of the test taker's proficiency. There may be discrepancy between the test taker's correct usage of a language form on a structured task and his/her usage of that same form in nonstructured speech. This relates to the overriding issue of validity and explains why nonstructured communication tasks are generally preferred.

The question of the generalizability of performance on structured communication tasks to performance in nonstructured communication is in desperate need of research. Nonstructured communication tasks generally require a larger corpus of language and a more complex scoring system, proficiency testing might be simplified considerably if there were research data indicating a strong correlation between the two types of tests. Unfortunately, such research has not been conducted to this date.

Previously we mentioned that some test formats employ unnatural language. Such formats involve either mimicry or manipulation. Mimicry is used on the LAS, the *MAT-SEA-CAL*, the *Del Rio Language Screening Test*, the *Northwestern Syntax Screening Test*, and numerous others. It is often used to see if the test taker can pronounce certain words or sounds. In the form of sentence repetition, mimicry is considered to be an indicator of syntax acquisition (Natalicio, 1979). Linguistic manipulation tasks are often based on a foreign language teaching technique known as pattern practice. Thus, a student may be given a sentence in the present and told to change it to the past. Since children and even many teachers, especially if they are in content areas, normally do not know grammatical terms, on tests that employ linguistic manipulation tasks, the examiner usually initiates the response for the student as in the following examples from the *Spanish/American Oral Proficiency Test:* (Politzer and Ramírez, 1975)

> Today it's clear. Yesterday...
> He studies in the library. They... .
> Hace frio hoy. Ayer...
> Él estudia en la biblioteca. Ellos...

It has been widely questioned whether linguistic manipulation tasks are valid measures of language proficiency. (Sánchez, 1976; Burt and Dulay, 1978). Indeed Politzer and Ramírez (1975) found that they do not work with many students. It appears that performance is related to a kind of "metalinguistic awareness" which can be defined as the conscious knowledge of the forms of a language. Some students seem to possess this awareness, while others do not, in spite of their proficiency in the language. Thus, there is good reason to believe that linguistic manipulation tasks are less desirable measures of proficiency than structured communication tasks.

An hypothesis that is relevant to the debate over natural communication versus linguistic manipulation tasks is that recently put forth by Cummins (1980). Cummins distinguishes between cognitive/academic language proficiency (CALP) and basic interpersonal communicative skills (BICS). CALP is defined as those aspects of language proficiency which are closely related to the acquisition of literacy skills in L1 and L2. While the aspects are not defined, there is at least an implication that it refers to such things as a greater vocabulary and the internalization of more complex linguistic structures. BICS, which includes oral fluency and sociolinguistic competency, may exist independently of CALP. According to Cummins, CALP is more likely to be assessed by linguistic manipulation tasks than by means of natural communication tasks.

A Review of the Relevant Characteristics for an Assessment Instrument of Bilingual Education Content Teachers

Now that we have reviewed the question of language variety and register and discussed some pertinent aspects of proficiency tests, we are in a position to make at least some preliminary recommendations concerning the characteristics and parameters of an instrument to assess bilingual content area teachers.

1. A preferable mechanism for assessment of the bilingual content area teacher is to have the evaluation of proficiency for teaching subjects such as science and mathematics to be part of a general assessment of language proficiency, rather than a separate test. It would be more appropriate to create a science/mathematics/other content area subcomponent of a general test, for these reasons, among others:

a. The language which this test assesses will be the world standard variety. However, because it is important that all bilingual education teachers, including instructors of content areas, know well the vernaculars of the students, this ethnically marked variety of language must also be tested. Convenience and practicality would dictate that the assessment of proficiency for content areas be part of a battery which would test for knowledge of the culture and vernacular of the student population as well.

b. Below, the suggestion is made that discrete-point testing would be appropriate for this type of assessment. However, discrete-point testing is best used when combined with integrative test items. Thus, another reason for having the content area assessment merged into a more general assessment (which would use integrative test items) is highlighted.

c. Reasons a. and b. are merely subparts of the more general proposition that while there ought to be an assessment of competence for teaching subject areas bilingually, the nature of this assessment is restricted and can not encompass all that should be assessed in the bilingual educator generally. Since it would behoove the tester to certify the general proficiency of bilingual educators, the content area evaluation component would naturally be an optional assessment unit to be given only to those educators for whom it would be relevant.

2. Since, for diverse reasons discussed earlier, it is the formal code, the world standard variety, which we are assessing, the assessment procedure should be appropriately constructed to restrict so-called "natural language." Margaret Zamora makes the same kind of observation for college-level language instruction programs. A relevant portion of her paper is worth quoting here:

> A language program for bilinguals is essentially different from Anglo-oriented programs in that it is designed to teach students who have a working knowledge of syntax, vocabulary, and pronunciation and to make maximum use of their limited competence. This competence, however, is adequate within a very reduced context—the home environment. The bilingual's command of the target language is usually oral, familiar, and colloquial, hence, not suitable for formal communication. When the student is placed in an environment which necessitates the use of a vehicle for communication other than familiar conversation (for example, the written word), functional use of the language deteriorates, often to the point of incoherence. (Zamora, 1981)

3. It is fairly self-evident that the sort of assessment instrument that we are discussing here needs to focus on cognitive/academic language proficiency, rather than on basic interpersonal communicative skills.

4. The assessment of proficiency for the bilingual instruction of content areas such as science and mathematics would appear to be focusing mostly on two elements: the lexicon of the content area, and the knowledge of the correspondences or equivalencies in L1 and L2 of the lexicon of the content area. Discrete-point test formats would seem the most appropriate for the evaluation of such competencies for a number of reasons:

   a. Discrete-point testing was itself based on contrastive analysis which in this instance closely approximates a skill which the instrument itself will attempt to assess.

   b. The linguistic element to be assessed—lexicon—lends itself to discrete-point test formats.

   c. As I have observed earlier, if this assessment were to be a subcomponent of a larger instrument, then other parts of the battery (for evaluating language arts, culture, etc.) could use integrative test formats.

The Problem of a Lack of a Clear or Well-Defined Register in Subject Area Bilingual Education

I have observed, immediately above, that the assessment of language proficiency in the subject areas will redound mostly on the area of lexicon. We have further come to the ad hoc conclusion that the level of that vocabulary which should be evaluated is the world standard. We have now arrived at a final, critical problem area in the assessment of bilingual content area teachers, at least with respect to the Spanish language. The problem is that much of the vocabulary (or jargon) of the content areas is borrowed into Spanish from English. This has been occurring in a haphazard fashion so that in many cases there is no clear lexical item in Spanish which can be said to be the standard. The problem is *not* that for a specific lexical item in English there are, in Spanish, competitors between the standard variety and one more vernacular varieties. The problem is that an English concept (lexical item) is translated in a variety of ways by different speakers (more appropriately, writers) of Spanish in accordance with subjective preferences, and accordingly, there has been no standardization of lexicon.

Other lexical items have not yet even begun the process of being translated into Spanish. For example, in the area of sports, we have basketball which has been translated into Spanish (with Royal Academy approval) as baloncesto, but which is more commonly used, in standard Spanish as basquetbol. In linguistics this is a

very common problem: there is quite a variety of Spanish equivalents or translations for the key words of sociolinguistics, such as pidgin, standard, or code-switching. In social work, words such as foster parents have no one standard translation, but a variety of equivalences. Thus, once again there is no easy way to determine within standard Spanish itself what the lexical item is in the language because of competitors (or perhaps, more appropriately, embryonic or pre-competitors) within the standard, or because of the actual nonexistence of equivalent terms in Spanish for scientific words in English.

We return to the problem that we have too many corpuses in Spanish and not enough standardization. There is a weak status base upon which United States Spanish rises. One fact is that authorized corpus planners of United States Spanish derive their authority from rather unique quarters, if they have any authority at all. A second fact, which returns us to the vernacular denigration/exaltation part of this paper is that corpus language planning is not socially innocent. Technical expertise alone never seems to be sufficient—there are always habits, attitudes, values, loyalties, preferences and ideologies among the planners themselves.

In many instances, Spanish lexicon appropriate for teaching science and mathematics concepts is and will remain for a long time either in flux or will not even be present. Full resolution of this problem awaits a long term solution based on the emergence of Spanish-speaking scientifico-linguistic academies, standardization mechanisms such as dictionaries, nomenclatures, grammars and research studies, and the emergence, above all, of a cadre of authoritative speakers who operate either formally as in France and Spain or along informal lines as in the United States, as a sort of national standardization model.

Conclusion

I have attempted to review the test issues involved in assessing the proficiency of teachers of bilingual subject areas. I have tried to apply the fruit of that review to a series of test construction recommendations. Finally, I have shown how the question of *what* to test, which would seem to be a fairly simple and forthright matter, in some instances will defy resolution.

References

Alatorre, Antonio "El idioma de los Mexicanos." *Revista de la Universidad Nacional Autónoma de México*, Vol. 10, (1955), 11-15.

Burt, Marina and Dulay, Heidi "Some Guidelines for the Assessment of Oral Language Proficiency and Dominance." *TESOL Quarterly*, 12:2, (1978), 177-192.

Clark, John L.D. (ed.) *Direct Testing of Speaking Proficiency. Theory and Application*. Princeton, N.J.: Educational Testing Service, 1978.

Cummins, Jim. "The Cross Lingual Dimensions of Language Proficiency: Implications for Bilingual Education and the Optimal Age Issue." *TESOL Quarterly*, 14, 2, (1980), 175-187.

Fishman, Joshua "Nationality-Nationalism and Nation-Nationalism," in *Language Problems of Developing Nations*, ed. by Joshua A. Fishman, Charles A. Ferguson and Jyatirindra Das Gupta. N.Y.: John Wiley and Sons, 1968.

Gaarder, Bruce A. "Language Maintenance or Language Shift: The Prospect for Spanish in the United States." *Bilingualism in Early Childhood*, ed. by Theodore Andersson and W.F. Mackey. Rowley, Mass.: Newbury House, 1977.

González, Phillip "English as a Second Language in Math Education." *Journal of the National Association for Bilingual Education*, 5:1, (1980), 93-108.

Jones, Randall L. and Spolsky, Bernard. *Testing Language Proficiency* Arlington, Virginia: Center for Applied Linguistics, 1975

Keller, Gary D. "Psychological Stress among Speakers of United States Vernacular Spanish," Luis Ortega, ed., *Introduction to Bilingual Education*. New York: Las Americas-Anaya, 1975.

_____ "Constructing Valid Goals for Bilingual and Bidialectical Education: The Role of the Applied Linguist and the Bilingual Educator," in Gary D. Keller, Richard V. Teschner, and Silvia Viera, *Bilingualism in the Bicentennial and Beyond*. New York: Bilingual Press, 1976.

Kjolseth, Rolf "Bilingual Education Programs in the United States: For Assimilation or Pluralism?" Bernard Spolsky, ed. *The Language Education of Minority Children*. Rowley, Mass.: Newbury House, 1972.

Lado, R. *Language Testing: The Construction and Use of Foreign Language Tests*. London: Longman, 1961.

Lozano, Anthony. "Grammatical Notes on Chicano Spanish." *The Bilingual Review*, 1:2, (1974).

Natalacio, Diane. "Repetition and Dictation as Language Testing Techniques." *Modern Language Journal*, 63:4, (1970), 163-176.

Oller, John W. Jr., and Perkins, Kyle. *Language in Education: Testing the Tests*. Rowley, Mass: Newbury House, 1978.

_____. *Research in Language Testing*. Rowley, Mass: Newbury House, 1980.

Politzer, Robert L. and Ramirez, Arnulfo G. "The Acquisition of English and the Maintenance of Spanish in a Bilingual Education Program." *TESOL Quarterly*, 9:2. (1975), 113-124.

Sanchez, Rosaura. "Critique of Oral Language Assessment Instruments." *NABE Journal*, 2. (1976), 120-127.

Steiner, Stan. *La Raza: The Mexican Americans*. N.Y.: Harper and Row, 1969.

Troike, Rudolph C. "Social Dialects and Language Learning: Implications for TESOL." *TESOL Quarterly*, 2. (1968), 176-180.

Zamora, Margarita. "Language Instruction for Bilinguals at the College Level: Strategies for Self-Correction," *The Bilingual Review*, 8:1, (1981).

ON ASSESSING THE ORAL LANGUAGE ABILITY OF LIMITED-ENGLISH PROFICIENT STUDENTS: THE LINGUISTIC BASES OF THE NONCOMPARABILITY OF DIFFERENT LANGUAGE PROFICIENCY ASSESSMENT MEASURES

by
Benji Wald
National Center for Bilingual Research

0. INTRODUCTION

This paper will explore the linguistic bases of the language proficiency assessment (LPA) tests widely used in the United States to classify primary school students as "limited-English proficient" (LEP) or "fluent."

The order of discussion is as follows: the historical context which led to studies of the comparability of various LPA tests, the linguistic bases of various LPA tests, and empirical work showing that the linguistic abilities tapped by various tests are non-comparable because different linguistic abilities have different patterns of development and use.

1.0 THE ISSUE OF TEST COMPARABILITY IN LANGUAGE PROFICIENCY

The concept of the LEP student is founded on the common-sense notion reflected in the Supreme Court's Law decision of 1974, that a *meaningful* education, an American civil right, means an education in a language that a speaker can understand. Since understanding, or comprehension, is a gradient notion of more or less, more often than all or none, the LEP student is one who fails to show adequate command of a language according to some chosen criteria.

What the chosen criteria are, is where the notion of oral LP comes into play. The legally sanctioned need to identify LEP students, i.e., to match actual students with the concept of LEP, led to a proliferation of LPA instruments available for district use. Literally hundreds of LPA instruments, especially for the initial classification of K-2 children, appeared either through commercial production or initiatives in individual districts (cf. Locks et al, 1978).

On the face of it, the criteria and implied notions of language proficiency are different for various tests. More deeply, the question has been raised whether various measures concentrating on different aspects of language are equivalent. Proposed answers to this question have varied from a claim that all aspects of language proficiency are equivalent, e.g., Oller's recent claim that there is a global language proficiency (glp) underlying all measures of language proficiency and even language achievement tests (Oller & Perkins, 1980), to extremely complex models factoring out mode and channel, e.g., spoken-written; production and comprehension; and domain, e.g., home, school, etc. Perhaps the golden mean is Cummins' claim that there are two types of language proficiency — one related to school achievement and another which is not (e.g., Cummins, 1980).

Confronting a potential chaos in the classification of students into LEPs and FEPs or LESs and FESs according to district choice, some state governments took an active interest in comparing instruments in order to see 1) how use of different instruments affected the LEP count, 2) whether or not an instrument had predictive value for school achievement.

In 1979, the Texas Education Agency supported a study reported by Gillmore and Dickenson (1979) to compare five LPA instruments in Houston, Texas.

Among the tests compared were three LPA instruments of further interest to us here; BINL, BSM, LAS. G & D gave pairs of LPA tests to 464 pupils between K and 12 in six Houston districts.

Among their findings:

1. Comparability was poor to poorly moderate. BINL/LAS closest for K-2 (Kendall's TAU .48), BSM/LAS closest for 3-6 (Kendall's Tau .52).

2. All together BINL was the hardest of the tests, classifying 73% of students as LESA (LEP), LAS was the middle (30%), and the BSM was easiest (19%).

3. Of the three pairs, BSM/LAS agreed the most 78% (N=40), BINL/BSM, 45% (N=51) and BINL/LAS, 49% (N=34).

4. Of the three tests, correlations (using Pearson's R) between the tests and achievement tests were only significant for LAS, but very modestly, e.g., reading .31, vocabulary .28.

*The research reported herein was supported by the National Center for Bilingual Research through funds from NIE (No NIE-R-79-0011). The opinions expressed do not necessarily reflect the position or policy of NIE, and no official endorsement should be inferred.

Paper presented at the Language Assessment Institute, National College of Education, Evanston, Illinois, June 18, 1981.

In 1980, the California State Department of Education commissioned a comparability study of the BINL, BSM and LAS. In that study, all three tests were given to over 1,100 students in grades 1, 3 and 5 in five schools throughout California, none having a majority Hispanic student body (40% the highest, in La Puente).

Their findings were similar to G & D in some ways, different in others (Ulibarri et al., 1980).

1. Different tests identified different percentages of the same population as LES, etc. (LEP) (As in G & D, 1979)
2. BSM was the hardest at each grade level, but BINL shifted from easiest to second place at grade 3.
3. BINL/LAS had the highest agreement, from 45% at grade 3, progressing to 65% at grade 5.
4. Despite 3, LAS and then BSM (except at grade 5) alone showed significant association with reading level.

Because of the comparability problem, each test having a different set of criteria for language proficiency and each producing different sets of LEP students from the same population, I have chosen to distinguish *language proficiency* from *language abilities*.

I define *language abilities* as the abilities that a speaker possesses to *use* a particular language, abilities which in performance reflect knowledge, whether conscious or unconscious, of a language. I define *language proficiency* as a quantitative measure of these abilities, or a subset of them. For my purposes, LP presupposes an LPA test. The test converts language behavior proceeding from language abilities into a measure of LP. Using this terminology, to talk about a LEP is unclear unless the criteria are given. On the other nand, *language abilities* can be discussed independently of such classification schemes. If we can identify what a speaker can or cannot do using a certain language, we then have a basis for discovering what the speaker has to learn or how the speaker has to use what he already knows, in order to achieve a meaningful education.

The following diagram schematizes the relationship between language abilities and language proficiency according to the above discussion

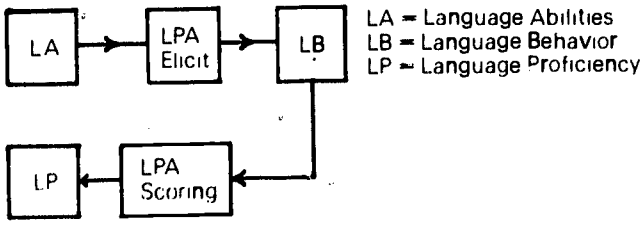


Diagram 1. Relation of Language Abilities to Language Proficiency

The scheme begins with the speaker's language abilities. In the first stage of language proficiency assessment, the LPA instrument is used by the tester to *elicit* language behavior. It is extremely important to recognize that the relationship between *language abilities* (LA) and *language behavior* (LB) is not direct, but is mediated by the *LPA elicitation*. The LPA elicitation creates a situation whose effects on the relationship between langage abilities and language behavior cannot be dismissed. Sociolinguistic research provides ample evidence that both quantity and quality of language is mediated by the social situation in which it occurs (cf. Labov, 1972; Mace-Matluck, 1980; Wald, 1980, 1981). This issue will become evident in the ensuing discussion of syntax. However, in this paper, the main focus of interest will not be immediately on the limitations of the tests, but the bases of the information that they provide about the speaker's language behavior. To this end, the criteria used for LPA scoring are focused on.

Continuing with the scheme, the LB which is produced through the LPA elicitation is then scored according to the criteria of the particular instrument used. This is the second stage in LPA. The output of this procedure results in the language proficiency (LP) classification.

Thus, the relation of LP to LA is quite indirect, depending first on elicitation and then on scoring.

2.0 THE LINGUISTIC BASES OF LPA TESTS

The most widely used LPA instruments, the LAS, BSM and BINL, and most others, focus on the *core linguistic components* of language. These components are:

1. **phonology**: the pronunciation and perception of linguistic sounds.
2. **morphology**: the processes of word formation; particulary, for English, the most frequently used inflectional suffixes.
3. **syntax**: the processes of sentence formation; the organization of words into sentences and intermediate units, i.e., clauses and phrases.
4. **lexicon**: vocabulary; the inventory of sound sequences and meanings paired into morphemes and words, especially nouns and verbs.

Each of the instruments is restricted to one or several of these components and devises a scoring system through which to quantify the results of elicitations, leading to assignment of the speaker to a classificatory category above or below a cut-off point for limited and proficient speakers.

What is important to us here is that each of the instruments has a different specific focus on just what core linguistic features it considers, and/or what value it assigns to any particular feature.

The following table shows the differing emphases of the three LPA tests.

Table 1

	BINL	BSM	LAS
phonology	–	–	+
morphology	–	+	(+)
syntax	+	–	(+)
lexicon	(+)	–	+

The BINL measures lexicon only insofar as it discounts non-English words used in an English-intended response in its use of a words-per-sentence count. The LAS measures morphology and syntax impressionistically, requiring the scorer to react in an impressionistic way to the frequency of deviations from standard English syntax and morphology in the speaker's retelling of a taped story. Special attention is drawn in Table 1 to the fact that the BINL and BSM take virtually diametrically opposed approaches to the evaluation of syntax.[1]

In considering the different content of each of these instruments, several questions come to mind. First, are the different components commensurate? That is, would we expect a score on one component to predict (have a direct relationship to) a score on another component? If so, why? If not, what are the bases for choosing or emphasizing one rather than another?

From a linguistic point of view, we might expect that within a given community most speakers acquiring English as a first language would show similar levels of development in all components, especially within the age range of 10-12, with which we will be dealing. Most features of phonology and especially morphology are relatively frequently used in speech, and most speakers would have adequate exposure to assimilate them. One would expect syntax and lexicon to be more matters of individual experience. To be sure, speakers of the same community should share the most obvious and usual syntactic patterns and typical vocabulary by this age. However, the more complex forms of syntax, found mostly in written English, e.g., the prepositional relative clause — the boy *to whom* I gave the book (rather than — the boy *(that/who)* I gave the book *to* — and specialized vocabulary, e.g., *distributor* (auto), *beside* (school-talk or archaic for *next to, by),* etc., we might expect to be differentially distributed

For acquirers of English as a second language, or even as a first langauge in a community in which the speaker is exposed from the outset to non-English speakers or many speakers of English as a second language, the situation may be quite different. The components of Table 1 are clearly separable. Thus, phonology varies greatly across English-first communities as well as among non-first speakers, regardless of the other components. Lexicon is also highly variable. The most extreme cases of separation of lexicon from syntax are shown by Creole languages, which may have an English vocabularly, but a syntax and morphology quite distinct from the mainstream varieties of English (cf. Hymes, 1971).

It is far from clear at what rates different components of a language, acquired as non-first, develop, or to what extent there is any predictable relationship among different components in second language development. There is general agreement, based on observation, that by early adolescence phonology begins to be acquired more slowly than other components. Some observations also show that social motivational factors may inhibit syntactic development among adults (cf. Schumann, 1978; Herdelberger, 1978).

Our interest at this point will focus on the relation of morphological to syntactic development among 42 late preadolescent Hispanics (aged 10-12) in the Los Angeles County area, as elicited by segments of the English versions of BSM-1 and LAS-1. We will be concerned with the value of the data elicited by these instruments as indicators of English language abilities.

3.0 EMPIRICAL STUDY OF THE LPA TESTS

As part of a comprehensive study of the effect of topic and situation on language behavior, in the final phase, individual students participated in a recorded language proficiency assessment interview with a bilingual female interviewer in her mid 20's. Each speaker was given sections of the BSM and then LAS, first in Spanish and then in English.

The BSM-E (English) segment consisted of page 6 of the Child Response Booklet. The speaker looks at three pictures which form a coherent story (pictures 5-7) and is requested to answer a number of questions. None of the speakers had any cognitive trouble interpreting the picture story or answering the questions in at least one language (Spanish or English). The responses are controlled to the point that, in many cases, precise cross-speaker morphological data were obtained.

The LAS-E segment consisted of listening to a recorded story following a traditional story-telling structure, although the content is original:

> Once upon a time there was a. . .one summer day. . .the next day. . . "I'm never going to drink pink ink again." (the end)

The speaker is requested to retell the story. This task elicits a unit of discourse from the speaker. Although it is not as controlled for precision as the BSM elicitation, it provides for a longer speech sample and a greater display of a variety of morphological and syntactic points of interest.

In examining the results of the BSM and LAS segments, we will not be directly concerned with the scoring applied to the samples, since these have already been shown to be noncomparable by the research discussed above, but rather with the more general questions of comparability of morphology and syntax on which these and other LPA tests are based.

3.1 BSM-E

The segment of the BSM-E (English) includes five questions (21-25) of which only the middle three are scored.

[1] Although the BSM calls itself a Basic *Syntax* Measure, it is virtually confined to *morphology* Morphology may be viewed as word-level, as opposed to sentence-level, syntax

However, for our purposes, the response to question 21 is also of interest. In a great many cases, the response to question 21 was along the lines of *the dog WANTS the food/to eat* The interesting morphological point is the *-S* marking subject-verb agreement. For speakers who created a response with *want*, the traditional rule of subject-verb agreement was recorded as present or absent. For some speakers, no data point was recorded because their responses did not contain an opportunity for subject-verb agreement, e.g., *the dog's hungry.*

The responses to questions 22 and 24 focus on irregular past tenses. The usual response to question 22 is *the dog ATE it* (= the food), and to question 24 is *it (= the apple) FELL*. Speakers who used the verbs *eat* and *fall* either used the irregular past form or did not inflect the verb for tense. Since, unlike subject-verb agreement (referred to hereafter as 3S), irregular past tenses must be learned item by item, it will be interesting to note that a specific relationship between the two past forms obtains in the data. Finally, question 23 elicits the most interesting data for variability. The question intends to elicit an unreal (contrary-to-fact) condition: what *would have happened* if the dog hadn't eaten the food? The crucial variable is the structure of the auxiliary: *would have*. In this analysis, we will attend especially to whether the auxiliary was a variant of *would have* (usually *woulda* or *would of)* or something else (e.g., *would, will, was gonna,* or nothing at all).

Figure 1 below shows a clear relationship between acquisition of the various structures and AOA (age of arrival). Each point gives the percentage of speakers using the form indicated out of the number of speakers using the structures, i.e., *eat, want, fall.*

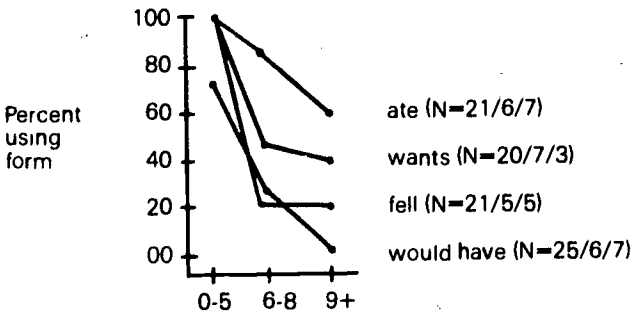


Figure 1: Percentage Using the Indicated Morphological Form, When Appropriate, by Age of Arrival

The figure shows that the unreal condition is the most discriminant structure of the four. Not even all of the earliest AOA group used *would have* in their responses. Of the two irregular pasts, *fell* is more discriminant than *ate*. The 3S inflection falls in between the two.

Diagram 2 below shows that there is actually a stable implicational hierarchy underlying the four features such that development of *would have* for the unreal condition implies development of the other three features, and so on until the indeterminacy between *wants* and *ate*.

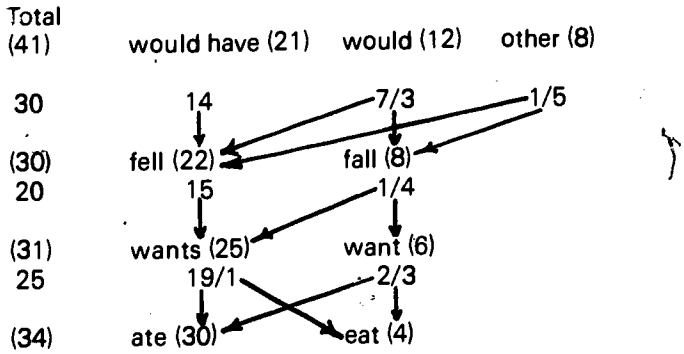


Diagram 2. (N) = Total number of forms in category
(N) = Total number of pairs of forms for adjacent categories

According to the implication in Diagram 2: *would have* implies *fell* implies *wants/ate*.

There is no obvious ordering between acquisition of 3S and the first irregular past verbs. For some reason, the past tense of *fall* is less accessible to speakers than the past of *eat* (frequency?). At this juncture, it is worth noting that the most frequently used irregular pasts — *was, went, came, got* — are used in the LAS story retelling by *all* speakers regardless of the implications of Diagram 2. This suggest that acquisition of some past forms is relatively early, earlier than the initial development of 3S. However, the irregular past forms are still spreading to other verbs as 3S develops.

In any event, it is quite generally the case that acquisition of the auxiliary structure *would + have* (henceforth *wda)* by a speaker indicates that 3S and the past forms of a great many irregular verbs have already been developed by the same speaker. It follows from this that at least for speakers of this age group, *wda* is a useful diagnostic of a relatively high level of morphological ability in English.

Before moving on to discussion of the LAS story retelling, it is worthwile to make one further observation about the structure of the unreal condition. This concerns the verb following *wda*. In the standard English of the classroom, *wda*, like the perfect, requires a following verb to be in participial form, e.g. *eaten*. For the verb *eat*, in our sample, this was *never* the case. Table 2 below shows that the norm for *wda + eat* is *wda ate*.

Table 2

	would have +		
AOA	ate	aten	eat
0-5	9	1	1
6-8	0	0	2
9+	0	0	0

In the communities of the speakers (LA Mexican-American), either the past participle of *eat* is *ate* (cf. standard *eaten)*, or speakers of this age group are still developing the adult norm *(eaten?)*, independently of first language background. This is an empirical issue. Until it is resolved, it is impossible to apply the BSM instructions for scoring to this form, since the BSM explicitly allows the use of nonstandard (local community based) features without penalty in evaluating language proficiency. The trouble is that the BSM gives no list of such features to help the scorer — nor can it, given the present state of our knowledge of the conventionally nonstandard English of Hispanic communities.

3.2 LAS-E

As noted above, the story retelling presents a wealth of linguistic data. Foremost, it elicits a coherent multisentence unit (discourse unit which can be compared both to its source (the recorded story) and to other replicated versions. For present purposes, we are interested in the syntactic structure of the story as produced by each speaker. Our aim is to *compare* how the *syntactic* abilities, revealed by our analysis of the LAS responses, relate to the *morphological* abilities evident in the BSM responses.

For analysis of the syntactic structure of the stories, two variables were selected which figure in the scoring procedures of the BINL:[2]

1) number of clauses/per story,
2) number of types of clauses/per story.

In Figure 2 below, the number of clauses in the stories are broken down by AOA group.

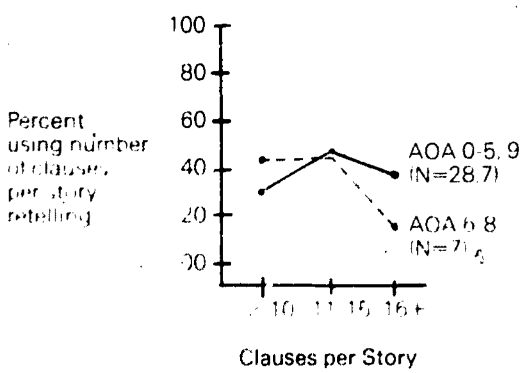


Figure 2 Distribution of Clauses per Story-Retelling by Age of Arrival

Although AOA showed a clear relationship to morphological development, it shows no clear pattern for story length. Most speakers of all AOA groups tell the story in 11-15 clauses before they return to silence.

We might suspect that syntactic complexity would be a better measure of linguistic development than clause length among speakers who are already accustomed to producing multisentence units in everyday situations. In the present analysis, our interest in syntactic complexity will be restricted to clause types. In the stories, we singled out seven different clause types used by at least two of the speakers, although the LAS source only used five (also represented among the seven).

1. **n + clause.** This simple type of clause is introduced by *n* representing any form of the conjunction *and)*, e.g.,

   . . .it was a giant *n* he like uh she like uh pink lemonade. . .(PQ 12f, AOA6)

2. **(n) then + cl.** Another simple clause type is introduced by either *then*, or *n then*, e.g.,

   . . .he ate uhm pink paint *n then* the next day he was sick. . .(CB 11f, AOA6) [not in LAS source]

3. **so (then) + cl.** Another common clause type is introduced by either *so* or *so then*, e.g.,

   . . .he liked pink lemonade a lot. *So* one day he went. . .(MC 11m, AOA4)

4. **after/when + cl.** A clause is introduced by *when* or *after* as a subordinator.

   . *when* he dri-drink the water, he said, that is not a lemonade. (LG 12f, AOA10)

5. **if + cl.** This clause type, introduced by *if*, only occurred in the story introducing an embedded question, e.g.,

   . he went n to go see *if* he find lem- pink lemonade. (MER 11f, AOA5)[3] [not in LAS source]

6. **relative clause.** This type of clause is introduced by *that* or *what* in the data, e.g.,

   Once upon a time there was a- a silly old monster *that* liked to eat pink lemonade (BR 12m, AOA9)

[2] The BINL has its own elicitation instrument in a series of pictures about which the subject is to utter a sentence consisting of one or more clauses. The scoring favors longer and more syntactically complex utterances. The BINL also has the variable of *instant de-* [illegible] that the visual image of the picture is required to be converted into a [illegible] story on the spot. This is a typical example of a *test* feature which rarely occurs in everyday communication but prefigures the requirement of *composition* as a writing skill. The story retelling is closer to everyday communication. Although the semantic content is highly determined, there is considerable syntactic creativity displayed by the speakers.

[3] Note the absence of a modal, e.g., *could* in the if-clause. This issue is separated from the *use* of this clause type.

. . .one day he saw *what* he thought was pink lemonade. . .(EP 11f, AOAO)

7. **any other clause.** The most common clauses of this type are those with no introducing marker, e.g.,

. . .he said ∅ he'll never drink pink ink again (JP 12m, AOAO)

or reported speech introduced by *that*:

. . .he said *that* he'll never eat pink ink again (AP 12m, AOAO)

or clause introduced by the conjunctions *but* or *because* (both rare):

. . .he went over to the ink *but* he didn't know (KR 11m, AOAO)

In Figure 3 below, AOA groups are plotted for clausal variety. There is no clear pattern discriminating the AOA groups. Generally, speakers tend to produce no more than five different clause types, without a clear preference between 2-3 and 4-5 types.

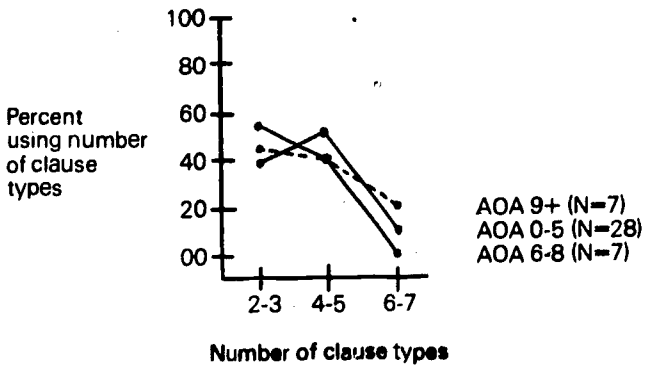


Figure 3. Distribution of Variety of Clause Types by Age of Arrival

A comparison of Figure 1 with Figures 2 and 3 shows that there is no clear relationship between morphological behavior and either length or clausal variety in discourse units. While AOA shows the development of morphological forms, it is indifferent to the types of syntactic behavior we have investigated so far.

It is important to note in this context that our approaches to the analysis of morphology and syntax are necessarily different. We can easily compare a morphological form with the standard form, but we cannot make such a comparison for syntactic variety since there is no clear standard for story length or syntactic variety. It is far from obvious that a speaker who uses only a few syntactic devices in the story retelling does so because he does not have adequate command of other syntactic devices. We certainly cannot assume that each speaker approached the story-retelling task with the same enthusiasm, concern for length or syntactic variety -- quite apart from the issue of English syntactic *ability*. In another paper, I will report on how length and clausal variety relate to situation by comparing behaviors on the LAS with the same types of behaviors in active competition for the floor in peer group sessions. This is beyond the scope of the present effort.

3.3 Further Observations on the Relation of Syntax to Morphology

There are some final observations concerning syntactic complexity that are of interest. Table 3 below shows the number of speakers using each of the seven clause types at least once.

Table 3

Other cl	n + cl	(n) then + cl	so (then) + cl	aft/whn +cl	if + cl	rel cl
42	39	32	18	9	4	19

The RC (relative clause) structure is used by a little less than half the speakers, although it occurred three times in the source. Again, we must note that absence of the structure in the LAS does not necessarily mean lack of knowledge of the structure. However, an interesting pattern of use is observed by either AOA or AGE. Observe Figure 4.

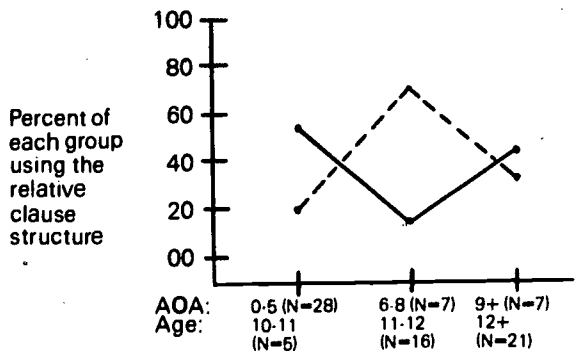


Figure 4. Percentage of Speakers Using the Relative Clause Structure in the Story Retelling by Age of Arrival (solid line) and Actual Age (broken line)

AOA shows a pattern of the depressed middle group (AOA 6-8). This pattern is reminiscent of a pattern reported for several studies, but most notably a Toronto study of immigrant students, by Cummins (1981). In the Toronto study, a mixed batch of immigrants of ages 11-12 showed the following characteristics: using the Peabody Vocabulary Test (PVT) of English vocabulary development, Cummins notes that those who arrived at ages 2-3 showed more vocabulary development than those who arrived later. However, those who arrived at ages 4-5 did *not* show more vocabulary development than those who arrived at ages 6-7 (in fact, they showed slightly less as a group). Thus, in that study, ages 4-5 is the *depressed middle group* with respect to the PVT since they do not show an advantage over peers who arrived at a later age. Cummins suggests that there is an interdependence between vocabulary development in L1 and L2 such that the more recent group (6-7) were able to transfer the skills underlying vocabulary acquisition in L1 to L2 and thus learn English vocabulary at a more rapid rate than the 4-5 AOA group, who one would expect had a less developed vocabulary in L1.

If a similar interdependence argument were adapted to explain the data of Figure 4, it would be that "something" (an unclear factor) in the development of the 6-8 AOA group's Spanish is either lacking or is not transferred to English, while for the later 9+ AOA group it is — and this "something" is responsible for the likelihood of using the relative clause in the story retelling.[4]

However, the interaction of RC with age is just as striking. The 11-year-olds, as a group, use the RC much more often than either the 10- or 12-year-olds. Whatever the explanation, there is no incremental developmental trend apparent, in contrast with morphology.

Table 4 below shows that using *wda* as a diagnostic almost evenly divides those who used the RC and those who did not.

Table 4

	+RC	-RC
+ wda	9	12
- wda	9	11

It is clear that, for the age and grade levels represented in this study, the syntactic and morphological measures derived from LPA tests are not comparable. Therefore, one should not expect any comparability of LPA instruments based on one and the other of these criteria.

The evidence indicates that development of morphology is relatively straightforward for preadolescents, depending on length of residence and/or exposure to American schools.[5]

On the other hand, the development of syntax and its use in test situations are two separate matters. In solving the problems of syntactic development and use, we will have to take into account situational variables of motivation as well as ability, and also the relation of the syntax of the first language to that of English.

In this context, we must note that the structure of the relative clause in Spanish is not appreciably different from that of English, owing to their common Indo-European heritage. In both cases the relative pronoun immediately precedes the relative clause, e.g.,

he didn't like *what* (he drank)

no le gustó *lo que* (comió
Neg to-him pleased it-that (he) ate
"he didn't like *what he ate*"

This is in contrast to a language like Chinese (Standard Mandarin) in which the equivalent clause precedes the qualified noun and is marked by a particle (P) at its *end:*

xǔduō (lái Zhōngguó fǎngwèn) *de* pengyou
many come China visit P friends
"many friends *who come visit China*"

or Japanese, with a similar structure, but without any marker:

(anata no tabeta) sakana wa
you subj. ate fish topic
"the fish (that) you ate"

or Swahili, in which the equivalent clause follows the modified noun, as in English and Spanish, but the marker of the relative clase (RM) is embedded in the verb phrase:

kitu (a-li-*cho*-kula msichana)
thing (she-Past-RM-eat girl
"the thing (that) the girl ate"

The similarity of the English and Spanish relative clauses may allow easy transfer from one language to the other. The same kind of transfer potential is not evident in morphology, e.g.,

com-e: com-ió = eat-s: ate
ca-e: ca-yó = fall-s: fell
hubiera (Aux. + Perfect + Subjunctive) — would have

We may expect that, except for those few cases of almost identical morphology, e.g., the English and Spanish regular plurals, the learning of English morphology is independent of language and shows an orderly pattern according to exposure (as reflected in our study in AOA). For syntactic structure, consideration of the first language is likely to be more important, but because of the apparent freedom of choice speakers have in selecting devices, it is difficult to judge a speaker's syntactic abilities from what he *doesn't* say in a test situation.

[4] Further research will show that this "something" is *not* failure to have developed the relative clause structure in either language for any of the speakers.

[5] I suspect the second plays a fairly minor role, but it is inseparable from the first in the data.

References

Cummins, J. "The Cross-Lingual Dimensions of Language Proficiency: Implications for Bilingual Education and the Optimal Age Issue." *TESOL Quarterly*, 14:2, (1980), 175-187.

_____. "The Role of Primary Language Development in Promoting Educational Success for Language Minority Students" (paper prepared for the California State Department of Education Compendium on Bilingual-Bicultural Education), 1981.

Gillmore, G. & Dickenson, A. "The Relationship between Instruments Used for Identifying Children of Limited English Speaking Ability in Texas." Region IV Education Service Center, Houston, TX, 1979.

Heidelberger, Spraus, Forschungs project "Pidgin Deutsch." "The Acquisition of German Syntax by Foreign Workers," in D. Sankoff, ed. *Linguistic Variation: Models & Methods*. Academic Press, New York, NY, 1978.

Hymes, D. "Pidginization and Creolization of Languages." Cambridge University Press, 1971.

Labov, W. "The Logic of Non-standard English." *Language in the Inner City: Studies in the Black English Vernacular*. Philadelphia, PA: University of Pennsylvania Press, 1972.

Locks, N. A.; Pletcher, B. & Reynolds, D. F. "Language Assessment Instruments for Limited-English-Speaking Students. A Needs Analysis." U.S. Department of HEW, NIE, October, 1978.

Mace-Matluck, B. J. "General Characteristics of the Children's Language Use in Three Environments." *A Longitudinal Study of the Oral Language Development of Texas Bilingual Children (Spanish-English): Findings from the First Year*. SEDL, 1980.

Oller, J. & Perkins, K. *Research in Language Testing*. Rowley, MA: Newbury House, 1980.

Schumann, J. H. *The Pidginization Process: A Model for Second Language Acquisition*. Rowley, MA: Newbury House, 1978.

Ulibarri, D., Spencer, M. L. & Rivas, G. A. "Comparability of Three Oral Language Proficiency Instruments and Thei Relationship to Achievement Variables." San Francisco, CA: Americas Behavioral Research Corp., 1980.

Wald, B. "Report on the Study of Limited Language Proficiency." National Center for Bilingual Research, Los Alamitos, CA, 1980.

_____. "The Relation of Topic/Situation Sensitivity to Language Proficiency." Paper presented at Ethnoperspectives in Bilingual Education Research. Eastern Michigan University, Ypsilanti, MI, 1981. (publication forthcoming under editorship of Raymond V. Padilla).

Part III
Research and Policy

EVALUATING SECOND-LANGUAGE SKILLS IN CANADIAN FRENCH-IMMERSION PROGRAMS

by
Sharon Lapkin
The Ontario Institute for Studies in Education

Paper presented at the Language Assessment Institute, National College of Education, Evanston, Illinois, June 17-20, 1981.

The Bilingual Education Project of the Ontario Institute for Studies in Education has been involved in evaluating French-immersion programs for the last ten years. The specific programs evaluated have been located in various parts of Ontario and share the following characteristics: (1) they are optional; (2) they serve an English-speaking or majority language population; (3) the initial years of the program involve from one-half to a full school day in which instruction occurs in the medium of the second language; and (4) in principle, the curriculum content of the immersion program is similar to that of the regular English program.

Canadian research studies on French-immersion programs have, for the most part, been based on comparisons of the academic achievement of immersion students with that of regular English program comparison groups. This emphasis can be explained by the need of educators to document that the educational experience of students in immersion programs was similar to that provided by the regular English program, and the desire of parents to be reassured in this regard.

For the purposes of this paper, the question of academic achievement in such subjects as science or mathematics is not addressed (but see Swain and Lapkin, 1981). The issue of the student's proficiency in their first language, English, is also not considered. Instead, the focus is on the second language proficiency: what is it? how has it been evaluated in immersion research studies? what kinds of conclusions can we draw from results to date?

Before addressing these questions, it is important to describe briefly two major French-immersion programs that we have evaluated. The most widespread immersion alternative available in all ten Canadian provinces is the early total French immersion program. The language of instruction in both the half-day Kindergarten program and early grades is French, with a daily period of English language arts being introduced in grade 2 or 3. At grade 5, usually from 60% to 80% of the school day is allocated to instruction in French, with this percentage dropping to approximately 50% at subsequent grades. At secondary school, early immersion follow-up programs typically consist of about three subject options (including French language arts) available in French for a total of 30% to 40% of the school day.

"Late" immersion programs may begin in grades 6, 7 or 8. The program evaluated by the Bilingual Education Project is located in Peel County (near Toronto) and begins at grade 8 with 55% to 70% of the instructional time in French. (Students entering the program have a background of short daily periods of French as a second language instruction in grades 6 and 7.) The intensive year of French at grade 8 is followed, at the secondary level, with several subjects available in French (history, geography and French language arts) at grades 9 through 11. In this particular late immersion program, no school subjects have been available in French at grades 12 and 13 except the traditional 40-minute French as a second language program in which post-immersion and non-immersion students have usually been combined in the same class.

In the early evaluations of both early and late immersion program alternatives, the French language skills of the immersion students were compared to those of their peers in the regular English program who were studying French as a second language in 20 to 40-minute daily periods. In retrospect, it seems naive to have compared the two groups, since, to take a specific example, after one year of late immersion program, Peel County immersion students performed as well on tests of French listening and reading comprehension as regular program students studying French as a second language in daily periods who were two to three grade levels ahead of them.

The comparison was also relatively meaningless in view of the objectives of French immersion programs with respect to second language skills. Educators and parents alike hoped that by the end of secondary school, immersion students (from both early and late programs) would be functionally bilingual. In 1974, the Ministry of Education for Ontario described, in functional terms, their expectations for second language use by immersion graduates. It was expected that these students would be able to continue their education in French at the post-secondary level, to accept employment where French was the working language and to participate easily in conversation. In other words, to a certain extent, the students' ability to speak, read, write and understand French would be native-like.

The relative comparison in assessing second language skills would therefore best be made in relation to the native language abilities of francophone students of the same age and grade level. The extent to which we have been successful in making such comparisons meaningful will be considered after we have reviewed several tests and procedures used in the evaluations and the results obtained.

When immersion education was first offered in Canada, there were no tailor-made assessment procedures (just as there were no suitable curriculum materials). One standardized French achievement test, normed on a population of native French-speaking students in Montreal, had been developed by the Commission des Ecoles Catholiques de Montréal. This *Test de rendement en Français* proved useful in the French immersion context, because it provided native speaker data at each grade level against which immersion students' progress could be assessed. The test measures such aspects of French achievement as identification of sounds, word knowledge (synonyms, antonynms, definitions), grammar (verb conjugations, number and gender agreement, recognition of parts of speech), spelling and reading comprehension.

The kind of information yielded by using the *Test de rendement en Français* can be exemplified by results from its use in evaluating the early French immersion program. By grade 1 or 2, immersion students score as well as about one third of their native French-speaking counterparts in Montreal. By grade 6, they score as well as one half of the Montreal comparison group. We have not analyzed the immersion test results diagnostically, that is, to identify aspects of achievement in which their performance is relatively strong or weak in comparison to that of the French-speaking peer group.

To complement the information from such discrete-point achievement tests which place a distinct emphasis on grammar, we have designed measures intended to be more communicative. These include the *Test de compréhension auditive* and the *Test de compréhension de l'écrit* to measure listening and reading comprehension respectively. In developing these tests, an attempt was made to identify a number of real-life situations in which immersion students might have contact with French speakers and to measure their understanding of the French used in each situation. Thus the listening test consists of a number of radio broadcasts including news items, sports items, weather forecasts, advertisements and radio drama, each followed by one or more questions with multiple-choice responses. Similarly, the reading test consists of a number of written passages taken from newspaper and magazine articles and advertisements, comic strips, horoscopes, television schedules, recipes, and poetic or prose literature.

We think of these tests as integrative measures, testing not so much knowledge of a specific vocabulary item or grammar point, as the students' ability to derive meaning from the spoken or written text as a whole. Can they grasp the global meaning of such passages in the same way as they would quite naturally in their own mother tongue?

A third integrative test used in the evaluations is a French cloze test, a written text from which every seventh word is deleted, with the initial and final paragraphs left intact to provide a context for the passage. This cloze technique is thought to provide an overall measure of proficiency since it reflects the students' intuitions and "educated guesses" about the second language (Lapkin and Swain, 1977, Oller, 1979).

Two of the tests just described have been administered to early and late immersion students at grade 8, for example, and to bilingual francophone students at the same grade level. The results suggest that by the end of elementary school, the performance of early French-immersion students on the cloze test and the listening comprehension test is similar to that of grade 8 classes of bilingual francophones. Comparative statements involving francophone students must always be qualified in view of possible IQ, socioeconomic and other difference between the groups.

The use of these tests has also enabled us to compare early and late French immersion programs at grade 8 when the late immersion students have had only one year of intensive exposure to French. As expected, late immersion students' scores are significantly lower than those of early immersion counterparts. In this case, the comparisons are statistical because we are confident that the relevant characteristics of the early and late immersion classes tested are similar as indicated, for example, by results from standardized IQ tests administered at grade 8 to both program groups.

Since the first groups of early immersion students in Ontario have not yet reached the end of secondary school (at the time of writing they have completed grade 10), it has not been possible, to date, to compare late immersion results with those of early immersion groups at the end of their respective programs. With respect to the question of how native-like the skills of late immersion students may be after several years in the program, we have been able to administer the listening test described earlier to bilingual francophone students at grades 10 and 12. The results indicate that the bilingual francophone groups score significantly higher than the immersion students. From such results we infer that the late immersion students' performance after several years in the program is not yet native-like.

Until this point, the test data presented have been intended to illustrate the kinds of comparative statements that can put immersion program results in perspective. With the exception of the cloze test, the tests that have been reviewed are measures of receptive skills. But how well can immersion students speak and write French? Further, the analyses reported are quantatitive, and no information has been presented on qualitative differences that may exist among the groups. Moreover, if we express the goals of the immersion program in "functional" terms — that is, as noted earlier, that immersion should enable its graduates to converse easily in French, for example — what do we know about the patterns of French language use of the two immersion program groups?

There are relatively few descriptive linguistic accounts of immersion students' spoken and written French. Those that are reported here relate almost exclusively to early French immersion programs. In one study of writing ability, Swain (1975) analyzed both English and French stories written by grade 3 immersion students. Comparing the French stories of the early immersion students to their English stories, she found the French

stories to be shorter but equally varied in the use of vocabulary items (except prepositions); there were more grammatical errors and fewer "technical" errors (in punctuation and capitalization), and more spelling errors. The French stories also tended to be more descriptive in nature than the English stories. It was clear that the French writing of the immersion students required more attention, since a second comparison between English stories written by the same immersion students and regular program comparison students showed few differences.

Using a French cloze test of the type described earlier, we also did a qualitative analysis of errors made by early immersion students at grade 5 relative to those made by francophone comparison groups, both bilingual and unilingual (Lapkin and Swain, 1977). The immersion students' scores were significantly lower than those of the unilingual comparison group. Some of the errors made were found among those made by unilingual French-speaking students. There was a striking similarity between the errors made by the bilingual francophone students and the immersion students, suggesting that the immersion responses were qualitatively similar, and that the processing strategies used by the immersion and bilingual francophone students might be similar.

Two studies of the total immersion students' spoken French have been conducted. The first (Harley and Swain, 1977, 1978) involved interviewing five randomly selected grade 5 immersion students, three bilingual grade 5 children of French-speaking home background attending a francophone school, and three unilingual French-speaking children from Quebec City. The interviews, conducted by a French-speaking adult who was not known to the children, were designed to elicit information on the verb system of the students. Among other things, the children were asked to narrate past experiences, explain current activities, and talk about the future: that is, they were expected to produce a variety of verb forms to realize specific semantic functions. A detailed linguistic analysis of the interviews revealed that the verb system used by the immersion students was simplified relative to that of the six comparison students. For example, immersion students were generally unable to produce conditional forms in French in response to such questions as: "Qu'est-ce que tu *ferais* si tu gagnais la loterie national?" Instead, the future might be used, thus removing the "hypothetical" nuance inherent in the conditional, or an adverb was used to convey tentativeness ("Je *peut-etre* acheter une voiture."). Thus immersion students were able to convey the essential message, but often used an inappropriate grammatical form to do so. The authors of the study concluded that there may be little need for the immersion students to acquire a completely native-like command of French: "Once the children have reached a point in their language development where they can make themselves understood to their teacher and classmates (as they clearly have), there is no strong social incentive to develop further towards native-speaker norms" (Harley and Swain, 1978:38).

It is important to bear in mind that the only source of native French input in an immersion classroom is the teacher and that sustained interaction with francophone peers may be necessary if the immersion children are to attain native-like speaking abilities (Swain, 1978c). In another study by Szamosi, Swain, and Lapkin (1979), recordings were made of the conversations of a small sample of grade 2 immersion pupils who, on an individual basis, spent two hours each week for a total of eight sessions playing in the French-speaking homes of grade 2 pupils who were attending a francophone school. A number of the play sessions were then transcribed and analyzed. The analysis was done in "functional" terms: that is, the purpose was to see how well grade 2 immersion students could use their spoken French in a natural context. It was found that the immersion students interacted with ease and naturalness and could joke, ask for clarification, issue orders, respond appropriately to the French-speaking playmate, and so on. The results of this study, along with those of the Harley and Swain (1978) study, suggest that early immersion students can convey a rich range of meanings in French, but that their way of doing so remains non-native.

The "non-nativeness" of immersion students' spoken and written French can be described, but these descriptions do not address the question of how acceptable their spoken and written French is to the native French speakers with whom they wish to interact. One recent empirical study (Lepicq, 1980) has examined the ability of early immersion students at grade 6 to communicate effectively in French with bilingual and unilingual French students and adults. The study involves presenting recorded speech samples from interviews with the immersion students and several unilingual French-speaking "control" students to a group of judges. The researcher interviewed the 96 judges (bilingual and unilingual, students and adults), to elicit views on the acceptability of the speech samples. The judges tended to apply different criteria in evaluating the immersion and francophone students, demonstrating that their expectations differed for the two groups. In general, the immersion students were assessed favorably. Lepicq concludes that immersion students have sufficient ability in spoken French to communicate with francophones and identifies errors which are more and less irritating to native speakers in the speech of the immersion students. It is this kind of information that is important for curriculum planning and syllabus design, since emphasis should be placed on those features of the second language which are considered most salient by native French speakers.

A small number of questionnaire studies have been conducted to determine whether immersion students actually put their French skills to use. Although relatively little self-initiated use of French (including listening to French on the radio, watching French television programs, using French on vacations etc.) is reported, immersion students are willing to use French in interpersonal communication. The extent to which French will be used in their adult lives in unknown, and it will be interesting some years from now to explore such questions as whether immersion students choose employment involving their second language skills to a greater extent than do non-immersion students.

The attempts to describe the second language proficiency of immersion students have been few, in part, because immersion is still a relatively recent phenomenon and because descriptive analyses of the sort reviewed above are time-consuming and expensive. The challenge to those of us working in this area is to cooperate more closely with mother-tongue specialists (linguists, psycholinguists and sociolinguists) so that theories of first-language development and information on aspects of first-language proficiency can be used to establish criteria against which second-language skills can be assessed in meaningful ways.

References

Harley, B., and Swain, M. "An Analysis of Verb Form and Function in the Speech of French Immersion Pupils." *Working Papers on Bilingualism, 14*, 1977, 31-46.

Harley, B., and Swain, M. "An Analysis of the Verb System Used by Young Learners of French." *Interlanguage Studies Bulletin, 3*, 1978, 35-79.

Lapkin, S., and Swain, M. "The Use of English and French Cloze Tests in a Bilingual Education Program Evaluation: Validity and Error Analysis." *Language Learning, 27*, 1977, 279-314.

Lepicq, D. "Aspects théoriques et empiriques de l'acceptabilité linguistique: Le cas du Français des élèves des classes d'immersion." Ph.D thesis. Toronto: University of Toronto, 1980.

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

Ontario Ministry of Education. "Report of the Ministerial Committee on the Teaching of French." Toronto: Ministry of Education, Ontario, 1974. (The Gillin Report)

Swain, M., and Lapkin, S. "Bilingual Education in Ontario: A Decade of Research." Toronto: The Ontario Institute for Studies in Education, 1981 (mimeo).

Szamosi, M., Swain M., and Lapkin, S. "Do Early Immersion Pupils "Know" French?" *Orbit, 49*, 1979, 20-23.

LANGUAGE ASSESSMENT PRACTICES AND POLICIES AT COLLEGES AND UNIVERSITIES:

Final Report of a National Survey of Bilingual Teacher Training Programs

by
Stanley S. Seidner
National College of Education

Introduction

A search of the literature regarding formalized bilingual teacher training programs in institutions of higher education will reveal that the area of language assessment of prospective candidates has remained a rich arena for research and model development. Virtually no studies have been conducted on the establishment of entry/exit-level criteria and assessment practices in bilingual teacher training programs (recent research has appeared by Seidner, 1981 a; Binkley, Johnson, et al., 1981). In contrast, an overwhelming emphasis of educational research has been placed upon these issues for limited-English proficient populations in grades pre-K through 12. For example: Cummins, 1979, 1980; Dulay and Burt, 1980; Matluck and Mace, 1973 (also Matluck, 1978); Estes and Estes, 1979; Maulden, 1979; Hardy, 1977; Cornejo, 1978; Curtis, Lignon, and Weibly, 1980; and Balasubramonian, 1979, among others). We begin this study with an overview of issues as they pertain to the topic. This will be followed by an analysis and discussion of relevant data from a national survey. Finally, the reader will be presented with models which pertain to the sample, conclusions, and recommendations.

A number of earlier investigations implied the need for such criteria, through the incorporation of particular language competencies by teacher training programs (for example: Sanchez, 1934; Blanco, 1975; Andersson and Boyer, 1970; Cordasco, 1976; U.S. Commission on Civil Rights, 1973; Flores, 1973; Valencia, 1970; Also Viera, Squires, and de Guevara, 1975, also see Michigan Department of Education, 1976; Illinois Board of Higher Education 1974; Texas Education Agency, n.d; National Consortium of Bilingual Education, 1979). The importance for considering the implementation of entry/exit level criteria for language assessment was underscored in Gaarder's less than veiled warning that the majority of bilingual teachers were inadequately prepared for classrooms (1970). Based upon earlier analysis of existing bilingual programs, the federally sponsored Maryland Conference of Bilingual Education offered a broad range of recommendations with implications for such criteria. Many of these recommendations were to remain unimplemented during following years (cited by McCaulay and Ramirez, 1977).

During 1975-76, a survey conducted by the National Center for Educational Statistics found 218 colleges or universities offering some form of bilingual teacher training. (Unfortunately, the second phase of this study as a comprehensive report remains on magnetic tape. For a discussion, see Waggoner, 1978). Basic federal grants during this fiscal year amounted to $53 million for 207 institutions of higher education.

An analysis by the former Department of Health, Education and Welfare of Title VII teacher training programs revealed needs in areas of second language acquisition and bilingual subject methodologies (Department of Health, Education and Welfare, 1976; also Dissemination and Assessment Center for Bilingual Education, 1978). However, the analysis failed to address areas of language assessment and development for potential teacher candidates, especially in the establishment of entry/exit-level criteria. The impression left from the onset was one of colleges and universities mobilizing their efforts to obtain coveted federal funds, rather than formulating long-range training programs, including systematic criteria for entry/exit levels of target language(s) assessment. The importance of this theme is revealed to the reader apropos to the topic through the course of this paper. In recent years, MacCaulay and Ramirez, (1977) underscored the urgency to carefully investigate the necessary requirements for teachers in bilingual education programs (also Ramirez, Gonzalez, et al., 1974; Sutman, 1979; Arciniega, 1977; Castillo, 1976; Carillo, 1977; Gonzalez, 1979; Rodriguez, 1980). As colleges and universities developed bilingual degrees, two concerns relative to this exigency were identified: 1. Higher Education Institutions simply relabeled existing courses as bilingual and regrouped them

The preliminary data of this study was presented earlier in an overly abbreviated form, due to editing constraints, in Padilla (1981). Unfortunately, the author was not consulted regarding substantial editing changes.

into a slightly different pattern. The inference, then, is one of a short-term economic expediency perhaps taking precedence over a long range of moral commitment to the program.[2] Seldom were new programs hiring new faculty to teach courses which were appropriate to bilingual teacher preparation. Furthermore, potential chasms existed between suggested remedies offered by scholars, and the lack of implementation by colleges and universities. Formulations by other researchers regarding target competencies for training programs, articulated desirable linguistic skills, but also neglected to investigate the nature of entry/exit-level criteria in language assessment by institutions of higher education. For example, the guidelines promoted by the Center for Applied Linguistics (1974) suggest a variety of competencies, but neglect to suggest a manner of implementation by colleges and universities in establishing points of entry and exit. Similarly, competencies were offered in a specific or more general fashion from researchers and organizations without benefit of an appropriate data base regarding ongoing practices by institutions (Palmer, 1975; Blanco, 1977; Martinez, 1975, Castillo, 1976; Brisk, 1978; Board of Directors, N.Y.S.U.T., 1975, Bilingual Education Service Center, 1977). Suggested competencies from Acosta and Blanco (1978) strongly recommended an exit-level language proficiency test which would include the evaluation of pedagogic competencies "in both languages." Inherent in the recommendation was some assessment process which would logically begin with potential students' onset of studies; hence, the need for entry-level criteria. In contrast then, with its definitions and applications to limited-English proficient students in public school, the terminology of entry/exit-level criteria for language assessment will be generally viewed for teacher training institutions in terms of overall preparation processes. The two important aspects for our purposes of this viewpoint, include the actual acquisition and manipulation of target language skills in subject areas (contextual basis). Both features of remedial and contextual basis will be discussed in another subsection relative to suggested model characteristics. (It is important to note that a lack of consensus in agreement among colleges and universities pertaining to the establishment of criteria and assessment practices is evident with recent data from a study conducted by the RMC Corporation (Binkley, Johnson, et al., 1981).

Particular emphasis is warranted regarding the nature of entry/exit-level criteria and quality of assessment processes, given recent studies on the relationships of teacher language to academic achievement (Merino, Politzer, and Ramirez, 1979; Merino and Politzer, 1977; Ramirez, 1979, 1980; Seidner, 1981b., also Bolger, 1967). It would follow that the lack of substantial criteria levels and assessment processes by training institutions would ultimately have some degree of negative bearing on the overall quality of teacher language. We find substantiation for this statement in the National Assessment Survey of Title VII ESEA basic programs which provided additional data through responses of surveyed teachers and directors (Dissemination and Assessment Center for Bilingual Education, 1978). The greatest incidence of need for teachers and directors corresponded at very high levels in areas of teaching content areas bilingually (Directors (n = 154), at 89; Teachers (n = 521), at 245) and bilingual teaching strategies (Teachers (n = 534), 242; Directors (n = 153), 85). Consistent with these responses, the National Center for Education Statistics reported a majority of teachers who were rated as Spanish/English bilinguals in their national survey, with language skills not derived from professional training programs. Only approximately one third of this number (n = 14,000) were rated as having crucial academic preparation to teach subject content areas to limited-English proficient students (Waggoner, 1979; also 1978; in addition, see National Advisory Council for Bilingual Education, 1981). The National Center for Educational Statistics is currently sponsoring another nationwide survey, similar to the one of 1977. (On this, see Chamot, 1981.) An earlier study by Frank Ramirez (1974), found that 20% of bilingual teacher applicants in the Los Angeles Unified School District failed the language fluency test (also Migdail, 1976). A survey was conducted by the bilingual teacher training unit of the New Mexico State Department of Education the year before. It was found that 10% of their sampled 136 bilingual teachers and aides could function at a third grade level when administered a reading and writing test in Spanish. None were able to show acceptable proficiency at the fourth grade level (Pascual, 1979).

This brief overview of issues underscores the need to raise a number of questions, especially in regard to entry/exit-level criteria in language assessment, which remain largely unanswered. The recent study conducted by the RMC Corporation provides us with a valuable general overview of current practices. Nonetheless, no comprehensive data base has been assembled in response to some of the following concerns: Who makes decisions regarding the establishment of criteria for entry/exit-level language assessment in training programs and choices of instrumentation? Who conducts the actual assessment of target languages? Are there relationships between the experience of decision makers in programs and choices? To what extent are public school, community sectors and faculty other than immediate program personnel included in the decision-making process? To what extent are opportunities offered to upgrade target language skills? What are the relationships between these opportunities to use target languages in educational experiences? To answer these and other concerns, we will continue with an examination of data from a national survey of bilingual teacher training programs in the United States.[2]

I would like to thank Dr. Donna Johnson from the RMC Corporation and Dr. Richard Baecher from Fordham University for sharing this data.

---

[2] Listings of bilingual teacher training institutions, both public and private, from which the sample was randomly selected originated from compilations by government and other sectors (for example, Dissemination and Assessment Center, 1979; U.S. Department of Education, 1981; National Clearinghouse for Bilingual Education, 1980 a, 1980 b, etc.). A distinction is made in this study between institutions and programs, since a number of colleges and universities offer more than one on graduate and undergraduate levels. (see footnote 8 of this paper for additional information.)

Overview of the Study

A survey of 127 items was mailed to 93 randomly selected teacher training institutions in the United States.[1] The final report of this study reflects a return of 69% (n=64). In addition, approximately 10% of respondents (n=9) indicated that their programs were discontinued, accounting for a total of 79% of the mailings. The percentage of return is considered more than adequate for potential analysis (See, for example, Babbie, 1979; Shannon, 1948). Responses arrived from programs located in 23 states, Guam and Puerto Rico. The survey was addressed to Program Directors with provisions for collaborative answers by other program personnel.[4] The final analysis showed that 90.6% (n=58), were completed by single individuals. The remaining 9.4% (n=6) were completed by more than one person. Only 12.5% (n=8) of responses completed by single individuals were done by personnel other than the program director. General data describing years of program existence and the number of immediate faculty are listed below in Tables 1 and 2. The mean number of years for program existence was approximately five-and-a-half years.

Table 1

PROGRAM EXISTENCE

ITEM	n=64	%	$\bar{X}$	s
			5.391	2.646
No. of years				
1-4	22	34.4		
5-8	33	51.5		
9-13	9	14.1		

Program mode for respondents was 5 years of existence (n=23), 35.9%. Only a small percentage of respondents (n=2) cited 1 year or less of existence for their programs. The program mean for years of existence corresponds with those from the aforementioned RMC study of 56 institutions. Prior months in compilation of their data account for the minimal difference in RMC and current statistical descriptions. This brings us to the number of faculty members serving in programs. Due to difficulties of interpretations of full-time and part-time equivalents, variations in definitions from one institution of higher learning to another, and differences brought about by semester, quarters and trimester divisions, immediate bilingual personnel were equated as such against individual institutional standards. (See Table 2.)

Table 2

PROGRAM FACULTY

ITEM	n=64	%	$\bar{X}$	s
			4.500	3.086
No. of Faculty*				
1-4	35	54.7		
5-8	24	37.3		
10-18**	5	8		

*Immediate program personnel/full and part-time equivalents equated against individual institutional standards.

**Reported numbers were: 10 Faculty (n=2), 12 (n=1), 15 (n=1), and 18 (n=1). Secretarial and similar support personnel were excluded.

Program mode for faculty was 5 employees for respondents (n=19, 29.6%). Another 20.3% (n=13) cited 2 faculty members, while 28% (n=18) responded with 3 to 4 personnel (14% for each category). The mean number of faculty per program was four-and-a-half (given the marginal division of a human being). The large standard deviation(s) reflected the wide range of numbers away from $\bar{X}$. In contrast, the sample from the RMC study revealed that the majority of bilingual courses were taught in most programs by the combined categories of 3-6 faculty members. The overwhelming number of personnel were drawn from the education department of colleges and universities throughout undergraduate and graduate levels, followed by foreign language sectors and the social sciences (Binkley, Johnson, et al., 1981). Of the 64 responses, 18 private and 30 public institutions offered bilingual teacher training programs for bachelor's degrees and undergraduate level credentials. (See Table 3)

Table 3

BILINGUAL TEACHER TRAINING PROGRAMS OF RESPONDENTS* (n=64)

Program	1. Assoc. (2 yrs.) %	2. Bachelor's Credentials %	3. Master's %	4. Post %	5. Doctoral %
Private Institutions 39% (n=25)		32 (n=20)	32 (n=20)	4.8 (n=3)	4.8 (n=3)
Public Institutions 61% (n=39)	6.4 (n=4)	52.8 (n=33)	43.2 (n=27)	1.6 (n=1)	16.0 (n=10)

*Columns 3, 4, 5 include programs offered by respondents cited in Column 2 (Bachelor's/Credentials level). As a result, the number of programs is apparently larger than the number of training institutions. Percentages (%) are based upon the sample of 64 responses (100%).

[1]Weaknesses of questionnaire research have been pointed out by Ruekmick who grudgingly conceded the value of this approach in education. Franzen and Lazarsfield (1945) concluded that the use of mailed questionnaires yielded more data which was more reliable than interviews (also see Topp and McGrath, 1950, Phillips, 1951, Conrad, 1960, Young, 1966, Rosenberg, 1968; Babbie, 1979).

[4]A study by See (1957) relates that a greater number of questionnaire returns were obtained when the original request was directed to the administrative head of an organization.

Three of the respondents representing private institutions cited current teacher training programs only on master's and doctoral levels. (Another private institution responded with a Master's in Education program). One public institution claimed only a master's program while two referred to master's and doctoral programs. There were no discernible patterns which emerged in private as opposed to public training institutions. Of the responses, 65.5% of the sample (n=42) were subsidized by federal and college/university funds. Only 12.5% received exclusively federal funding (n=8). In comparison, 3.2% (n=2) were completely funded by the respective college/university. Another 1.6% (n=1) were completely subsidized by the state. The remaining 17.2% (n=11) received funds from other sources (e.g. combinations of the state/federal, state/college, university/private sources, etc.) Private institutions characterized the last three categories, while the majority of public institutions offering undergraduate programs typified that of exclusive federal funding. The majority of programs receiving exclusive or combined categories of federal funding and colleges/universities were situated in states with special requirements for bilingual certification, the largest percentage of respondents, including California (15.6%), Texas (10.9%), Illinois (10.9%), New Jersey (7.8%), and New York (15.6%). Although New York State lacks a particular mandate, the majority of the responses arrived from New York City which conducts bilingual licensing of personnel.

As part of our overview, the 64 respondents reported the following categories of langauge assessed in their training programs (See Table 4)

Table 4

LANGUAGES ASSESSED BY PROGRAMS

CATEGORY	n=64	%	Cum. Freq. %	Mode=3.000
1 English Only	1	1.6	1.6	
2 Non Eng. Lang.	13	20.3	20.3	
3 Eng. and one other target language	42	65.6	65.6	
4 More than one non-English target language	4	6.3	6.3	
5 None	4	6.3	6.3	

The none (no assessment procedures) category included a Chamorro English training program and a private institution receiving no federal funds. Although category 3 reflected an overwhelming inclusion of Spanish, other languages such as Yupik were included in the data tabulation. The majority of the institutions receiving exclusive or partial federal funding, assessed one target language in addition to English in their bilingual teacher training programs. The second largest category, the assessment of one target language exclusive of English, reflected a majority of programs receiving some degree of federal funding. This brings up a question, how did assessment procedures of programs with more than two target languages differ from those with only two? Differences were apparent in assessment approaches, processes and choices of instruments by programs with more than two languages. In brief, the analysis of the data shows an increasingly heterogeneous pattern of assessment approaches for programs with more than two languages. More observation, interview and questionnaire techniques were employed in combination as the number of target languages increased. There was a greater reliance upon sectors other than the immediate program with the increase in languages, as well as the participation of public school personnel and community resources. It is interesting to note that the RMC sample revealed a smaller enrollment and reduced curriculum for target languages other than Spanish and English. These developments may be reflective of an overwhelming emphasis of funding, resources and program development for Hispanic ethnolinguistic groups. Having presented a general overview, we will now turn to some other specific aspects of the study, appropriate to questions raised in the introduction.

DECISION MAKERS

Who makes the actual decisions in determining entry/exit-level criteria for assessment of target languages? Only 18.8% (n=12) of the respondents cited the director of the program as the sole decision maker. (It is appropriate to note that the RMC investigation found that 28% of their sites reflected one or two key individuals who significantly affected the operation of the program). The largest incidence occurred with 40.6% of programs reporting decision making by immediate faculty. The next highest incidence was found at 26.6% of programs in which bilingual faculty from sectors other than their immediate program (e.g. foreign languages, language testing centers, etc.) were involved. The remaining 14% consisted of shared decision making by these others in conjunction with program directors. (One program which reported no assessment procedures responded with the director as decision maker in a hypothetical situation of near-future need.) Only 15.6% (n=10) of the respondents reported better than average participation of public school personnel in their decision-making process other than English. At the same time 15.6% (n=10) reported average participation. Finally, 54.6% (n=35) reported no participation, while another 11% cited involvement of public school personnel as low (based upon a total n=64 with 2(3.2%) missing cases). A similar pattern of responses evolved for the processes of involvement in establishing English language criteria.

Better than average community involvement (very high/highest ratings) accounted for 25% (n=16) of the respondents; 28.1% (n=18) reported average participation. A remaining 17.2% (n=11) claimed low participation, while another 20.3% (n=13) reported no involvement; 9.4% (n=6) of the cases were incomplete for consideration. The perceptually higher percentage of community may be explained, in part, by the existence of advisory boards from a number of institutions. Advisory boards, however, may also be mostly token representatives of participation in order to meet any funding commitments. Some writers in the field have

suggested benefits of community involvement in training programs. (For example, Brisk, 1978; also implications from Resnick, 1976; Carrillo, 1977; Barrera, 1973; De Inclan, 1976, and Pascual, 1979). It is interesting to note that 15.6% (n=10) of the target programs reported a better than average student participation in establishing language assessment criteria for entry/exit levels. Some 21.9% (n=14) cited average, while another 20.3% (n=13) were low; 40.6%(n=26) responded with no participation and 1.6%(n=1) of these cases failed to respond. Student participation was higher than anticipated and may be due to opportunities of internships, scholarship and fellowship roles. This brings us to the question of who conducted the actual assessment of target languages?

Less than 10% of program directors unilaterally conducted assessment of target languages. The incidence constituted less than half of the respondents reporting unilateral director decision making regarding entry/exit-level criteria. This type of assessment conduct was found in programs characterized by two target languages, less than the reported mean number of years of program existence. Some association was found between assessor(s) of languages and the decision makers for entry/exit-level criteria ($p<.05$, $r_s=.38$), nonetheless, not high enough to make an absolute judgment. The largest number of respondents occurred where assessment was conducted by all bilingual program faculty. (See Table 5)

Table 5

CONDUCTORS OF LANGUAGE ASSESSMENT

CATEGORY	n=62*	%	Cum. Freq. %	Mode=5.000
1 Program Director	6	9.4	9.4	
2. All Bilingual Program Faculty	22	34.4	35.5	
3. Some Bilingual Program Faculty	4	6.3	6.5	
4. Faculty from Other Sectors	7	10.9	11.3	
5. Bilingual and Other Faculty	23	35.9	37.1	

*Two missing responses refer to two programs reporting no formal assessment processes.

The second greatest occurrence consisted of bilingual and other faculty. Respondents largely identified "other faculty" in terms of personnel from English language centers and institutes. A number of respondents (6.3%) mentioned community resources participating in the actual assessment processes. Another 3.1% reported involvement of public school personnel. The brief occurrences were mostly found in programs reporting more than two target languages. Faculty from other sectors reflected involvement from foreign language departments and English language centers. The domination of assessment processes for students in bilingual programs by these sectors may be a considerable concern, ascertained from a number of comments volunteered y respondents. We also find evidence that the poorest interdepartmental relations arose between bilingual and foreign language sectors (Binkley, Johnson, et al., 1981). This was attributed to the incompatability of philosophies regarding the importance and nature of student proficiency in target languages and/or competition for student populations. In support of this premise, only 8% of American colleges and universities required a foreign language for admission during 1979, compared with 34% in 1966. Federal support for foreign language programs during this year (only 10%) was lower in real terms than in the 1960's. Private foundations have withdrawn almost completely from the field (President's Commission of Foreign Language and International Study, 1979). As a consequence, potential struggles between bilingual and foreign language sectors may transcend real issues of economic need and surface as the symbolic control of testing processes. This control would portend eventual funneling of students into one series of courses by one sector as opposed to another. Respondents for this study who indicated a large degree of institutional control over target languages mentioned such instruments as the M.L.A. test and the TOEFL which are popularly utilized by Foreign and English Language sectors. However, it was difficult in other conditions of control over assessment processes as reported by respondents, to distinguish between individual, program and institutional choice since a high degree of probable coincidence existed.

Given this pattern of involvement, what were selection choices of decision makers regarding assessment? In ascertaining entry/exit-level criteria for English and other languages, respondents chose the following kinds of assessment for program preference. (See Table 6)

Table 6

PREFERENCE FOR LANGUAGE ASSESSMENT

ASSESSMENT	I. English (n=64)	%	Adj.* (n=62)	II. Other Lang. (n=62)	%	Adj.** (n=61)
1. Commercial Instruments	9	14.1	14.5	6	9.4	9.8
2. Noncommercial Field-Tested Instruments	4	6.3	6.5	4	6.3	6.4
3. Informal Interviews	5	7.8	8.1	3	4.7	4.9
4. Observations	2	3.1	3.2	2	3.1	3.3
5. Other***	42	65.6	67.7	46	71.9	75.4

*2 cases Missing Mode I = 5.000
**3 cases Missing Mode II = 5.000
***Includes combinations of the above or other instruments

Compared to these responses, the RMC preliminary data indicated that the interview was the most often utilized by their sample for target languages other than English. Category Other reflected a majority of combinations which included informal interviews and/or observations. The largest incidence of program preference in the higher than average categories for languages other than English (n=37, or 57.8%) also occurred with interview scales. Comparative data from respondents for discrete point tests show a higher than average program

preference of 47.4% (n=27), and a 34.4% (n=22) for criterion-referenced tests. Questionnaires rating 22.1% (n=13) patterns of program preference were somewhat similar in the assessment of English. Some indication of particular use of instrument emerged, such as the FSI and the M.L.A. test (on these and other instruments, see Clark, 1978; Oller and Perkins, 1980). The contention that these patterns may result from locally developed or state developed procedures finds some strength in corroboration in RMC data. The RMC study further found that their sample failed to establish specific criteria for assessing student's language proficiency, accordingly, students were expected to be fluent in target languages, although no definitions were established in regard to the nature of fluency. Most programs saw no need for English requirements beyond those of the university as a whole (since the majority of students were educated in the United States and considered fluent in English). Assessment of students at exit levels occurred more frequently before their graduation from institutions in states with certification requirements in bilingual education. The reverse occurred in states without certification requirements. No general pattern appeared among respondents of this study among states with certification requirements. The use of a measure such as the FSI appeared stronger in one state, for example, while it was weaker as a recommended approach in another. A forthcoming discussion will provide a recapitulation and summary of these and other points. For the moment, we will examine relevant issues of remediation provisions and opportunities to utilize target languages.

Turning to concerns regarding availability of remediation courses in target languages, we find that only 3.1% (n=2) of respondents ranked available opportunities in English as low or none, as compared to cumulative 7.8% (n=5) for other languages. Another 67.8% of respondents (n=43) reported better than average opportunities for English language remediation, in comparison to 62.7% for other languages.

The highest opportunities for utilizing languages in academic settings went to English in contrast to other target languages (84.24% (n=54) in English compared to 46.8% (n=30) for other target languages). In order to ascertain relationships of language remediation opportunities to those of utilization, both Kendall's Tau

$$T = \frac{S}{\sqrt{\frac{1}{2}N(N-1) - T_x}\sqrt{\frac{1}{2}N(N-1) - T_y}}$$

and Spearman's Rho

$$r_s = \frac{T_x + T_y - \sum_{i=1}^{N} d_i^2}{2(T_x T_y)^{1/2}}$$

were adopted as rank order coefficients. The level of confidence $p<.05$ was selected a priori. Spearman's Rho yielded a closer approximation to Pearson's product-moment correlation coefficients as the data was more or less continuous. Although Kendall's Tau is usually more meaningful with a larger number of tied ranks, a correction for ties overcomes the potential shortcoming.

Table 7

RELATIONSHIP OF LANGUAGE REMEDIATION OPPORTUNITIES TO LANGUAGE UTILIZATION

RANK/TIES	N	r_s	sig.*	T	sig.*
English Language					
A. Opportunities					
B. Use	60	.3574	.048	.3046	.021
Other Languages					
A. Opportunities					
B. Use	58	.3612	.006	.3187	.001

*$p<.05$

Given the levels of confidence, a substantial degree of correlation is apparent between English/other language remediation opportunities and their counterpart ties for utilization. (See Table 7 above.)

Levels for remediation opportunities, ranked from none through highest, paralleled those for utilization of target languages in subject classes. Opportunities for upgrading language proficiency materialized in the following manner. Remedial coursework was provided in the target language to upgrade competencies. Subject-area coursework was presented in target languages. Field experiences were provided for students with the opportunity to utilize target languages. Programs which exceded two target languages, appeared to rely less upon utilization of target languages in formal class surroundings and more upon field experiences. The data substantiates the contention from the RMC study that the utilization of languages in classrooms depended largely upon the instructor's facility with the languages and with the number of target languages in the program. As the number of target languages increase, across-the-board opportunities for remediation seemed less than provided by programs with only two. It is appropriate to note that 34% of all programs in the RMC sample required their students to take one or more remedial language courses. Another 26% of the programs recommended coursework, but did not require the student to take it. As the academic level (i.e. masters, doctorate) of the program increased, the number of required courses was found to decrease. Given the nature of this study, this author's survey could offer only general substantiation of this tendency in this instance, rather than a definitive statistical comparison. No pattern appeared in the preliminary tabulations to indicate correlations between opportunities for remediation and types of assessment instruments. However, a higher correlation existed between the nature of the ultimate decision maker(s) for establishing language assessment criteria and opportunities in utilizing target languages other than English language skills (n=57, $r_s = .5459$, $p<.05$, reported sig. .001). The high incidence of group decision making (re. the immediate program with outside faculty), corresponded with high occurrences of opportunities for utilization of these languages.

Experience, Academic Preparation and Institutional Commitment

An analysis of data shows the directors of programs averaging a little over 7 years of teacher training experience on college and university levels. Interestingly, the average number of these years in bilingual teacher training for directors was less than 5 1/2 years. Throughout, the standard deviation indicates a rather large distribution spread regarding experience and academic preparation. Overall, teaching in public schools for directors averaged less than 5 1/2 years, of which less than 5 were in bilingual education. (See Table 8.) In comparison, faculty in bilingual programs averaged less than 5 1/2 years in teacher training. (Note mode of 2 years overall experience and only 1 year in bilingual teacher training.) Teaching in public schools averaged about 2 years for faculty, of which most experience was in bilingual education. (Here again, the mode for general teaching was 1 year compared to 1/2 year in bilingual education.)

Table 8

ACADEMIC AND EXPERIENTIAL BACKGROUND OF SAMPLED PROGRAMS (n=64)

	Directors			Program Faculty		
CRITERIA	$\bar{X}$	s	Mode	$\bar{X}$	s	Mode
Teacher Training (n=64)	7001	4110	5000	5377	4388	2.000
Bilingual Teacher Training (n=64)	5401	4414	2000	24896	3369	1000
Teaching (Public School) (n=60)	5216	5213	1000	2121	5912	1.000
Bilingual Teaching (n=60)	4704	5910	1000	1997	4274	0050
Academic Prep (n=64)	5252	4012	5000	4120	5062	2600
Assessment Univ Level (n=64)	4587	3465	4000	15202	2221	1000
Assessment Public School (n=62)	2782	3552	1000	1821	5122	1000

In terms of academic preparation, directors averaged less than 5 1/2 years of formal (e.g. college, university) and informal (e.g. workshops, seminar, etc.) training in second language acquisition and language assessment subject areas. Faculty compared with a little over 4 years. As for an actual language assessment experience in college and university training programs, directors averaged a little more than 4 1/2 years compared to a little more than a year-and-a-half for faculty. Language assessment experience in public school averaged a little less than 3 years for directors and a little more than a year-and-a-half for faculty.*

Data from the RMC sample reveals that 90% of the program directors and 85% of the education faculty have or had at one time a teaching credential. This in itself does not denote any large-scale experience on behalf of these individuals since such credentials in most instances can be obtained upon graduation from an accredited teacher training institution. Data from the RMC sample indicated the larger numbers of program faculty with doctoral degrees. (Of the doctoral specializations, 23% were in curriculum and instruction, another 15% in other educational areas, and 24% in language and linguistics.) About half the faculty had experience in a classroom with limited-English proficient students. Less than half of either group had experience of any sort in a bilingual classroom. Unfortunately, data from the RMC study regarding exact numbers of years experience were not available for specific comparisons with this author's sample. However, in general, the patterns of both studies appeared to corroborate each other as far as general trends.

The ramifications of the aforementioned data are such to suggest that current faculty may have much less overall teaching experience in public school sections, as well as less specific experience with limited-English proficient students, than actual teachers in public schools. Doctoral degrees might not offer a substantial substitution since, as one of my eminent colleagues so eloquently put it, the trainers of the teachers and the professionals in the college or university, who write about limited-English speaking students and offer potential directions, "may never have seen a kid in their entire lives within a classroom setting." The situation is further compounded by above-cited evidence which shows possible reluctance and/or inability of colleges and universities to involve experienced personnel from public schools in the decision-making process. The argument has been advanced that the relative "newness" of the field has, in essence, shaped the dilemma (Seidner and Seidner, 1981, b; 1982). We will attempt to offer some insight through our data based upon approaching the issue of institutionalization. Henry Pascual (1979) reminds us of our enthusiasm to "apply severe criticism to public school programs," or pressure upon publishers for better materials. Yet, we somehow neglect to question colleges and universities which remain sacrosanct and appear to "do no wrong."

Inherent in this statement is the ultimate concern of institutional commitment to programs and to the actual philosophical premises of bilingualism and bilingual education. For purposes of this paper, two particular concerns have been identified which relate directly to this study: a. commitment of resources to the program, and b. commitment to continue the program beyond the span of time dictated by funding sources other than from the university. Obviously, both concerns will ultimately influence the establishment and continuity of entrance/exit criteria for language assessment.

*The survey queried respondents along similar lines of related experience for other involved faculty and public school personnel. The returns were too miniscule to make any type of overall judgments. The small return indicating involvement from public schools in the assessment process yielded the following completed responses (n=6): Average no. years bilingual teaching: 4-4 1/2 years (40%), and 6 years (20%).

As shown earlier, a miniscule percentage of respondents indicated the total support of programs by the respective institution. On the surface, some commitment of institutional resources appears with the observation of the RMC sample that the proportion of program directors and faculty funded partially or entirely from institutional sources was quite high (Binkley, Johnson, et al., 1981). A secondary observation was the acknowledgement of other bachelor's and master's level programs alongside Title VII funded programs in at least ten institutions. It would be a naive assumption that colleges and universities are in the education business for sheer altruism.[6] An examination of a number of training programs, which claimed some or all absorption of the director's and faculty salaries, shows various distributions of funding among student reimbursements (tuition, materials, etc.), faculty costs (including trips), institutional overhead, and other operational costs. If Title VII funding were to cease, would training institutions continue with commitment of programs? Given the data in Table 7, how significant are institutional payments of directors' salaries in part or whole?

Will this apply to other program faculty? The RMC study offers the beliefs of interviewed directors that the institutions would consume funding programs beyond fiscal expectations. How valid are these kinds of remarks from respondents who are seeking funding, when interviewed by quasi-representatives of the ultimate funding sources?.[7] An important piece of the puzzle was cited earlier, namely the 15% of respondents who took the time to indicate that their programs were no longer in existence. Another study would be welcome to examine the extent of this phenomenon, in relation to particular sources of funding and institutional commitment.[8]

Another important characteristic of institutional commitment would appear to be one of retraining. If institutions rely upon personnel from departments other than bilingual, claim a shortage in trained professionals (Binkley, Johnson, et al., 1981) or stress the relative incipiency of such training programs, what is the commitment, then, to provide adequate staffing? Part of the answer may be found perhaps through inservice retraining. At least one national task force recommended a locus of support for inservice training in colleges of education (Kersh, 1978). No evidence has surfaced to substantiate an independent initiative or commitment of this nature. In fact, a recent study subsidized by the American Association of Colleges for Teacher Education examined the commitment of 17 institutions to inservice education. The potential position of the Office of Bilingual Education and Language Minority Affairs in securing funds and credentials was deemed a key factor in institutionalization (Carey and Marsh, 1980). Additional light may be shed on the nature of commitment (or lack therof) by examining the actual profile of immediate faculty within bilingual programs.

As shown above, an analysis of data collected from our sample revealed that directors of programs are the most experienced personnel. A further examination supported the premise that a large number of individuals were ensconced in areas other than bilingual at the onset of Title VII funding. However, the mean years of program existence exceeds the mean years of faculty teacher training experience. The large distances between directors and immediate program faculty suggested a lack of long-range commitment by training institutions to maintain the latter. Additional information from the RMC study shows that only 31% of all program directors had tenure, while another 34% were in tenure-track positions. The remainder were employed along non-tenure lines. Over 60% of faculty in bachelors/credentials programs were on tenure-tracks, the percentage increasing (for directors as well) as the academic level (i.e., master's, doctoral) increased. The citation of close to 60% of the tenure-track individuals, as tenured, points to their incorporation in definition from other departments as immediate program faculty. (Otherwise, the inconsistency of mean years for program existence and the average of seven years for tenure would remain unresolved.) The highest overall evidence of academic rank for directors occurred on assistant and associate professor levels, compared to instructor and assistant professor for faculty. In short, university hiring and retention policy may be a critical factor affecting alternative choices for selecting processes and entry/exit-level criteria for language assessment of potential bilingual teachers. Perhaps any reluctance vis-a-vis commitments by institutions is due in some part to the relatively short nature of existence of bilingual training programs. Colleges and universities seem to favor a philosophic framework, expressed by one institution as the incorporation of faculty which are "academically respectable and appeal to the scholarly" (Burke, 1972; See also Broudy, 1962; and Weber, 1972; Arends, 1973). Although the goal may be praiseworthy, the interpretation of terms by a select guardianship might at least prove troublesome. This is especially so for a relatively new program in its quest

[6]The Carnegie Council on Policy Studies in Higher Education predicts that enrollment will fall even more than the size of the historic college-age cohort of 18-21 (1980), and colleges and universities will compete for even more scarce students in destructive way. Institutions of higher education ultimately are seeking other sources of funding to venture beyond, what may be considered in numerous incidences, survival levels (Kramer, 1980; Mayhew, 1979; U.S. National Center for Educational Statistics, GPO Annual (1979, 1980); Academy for Educational Development, 1971; Carnegie Council on Policy Studies in Higher Education, 1979).

[7]A number of respondents from the sample when spoken to after the survey in an informal setting expressed fears that the institution would not continue. . .the programs beyond the expiration date of federal programs. They describe their respective institutes with such descriptive terms as "cannibalistic," and "prostitute." Also consider Franzen's and Lazarsfeld's study on conditions and validity of interviewing.

[8]During fiscal year 1980-1981, schools of education at 27 institutions received over $1 million in Title VII funds to design first-year training programs. One hundred and thirty colleges/universities and 38 other applicants received $18.5 million to improve teacher training programs and to prepare bilingual personnel. Forty-four colleges and universities received more than $4.5 million in fellowships to post-graduate students. Twelve universities, four school districts and three nonprofit organizations received $9 million to operate 19 bilingual education service centers. Eight universities, four school districts and five materials development centers received $6.5 million to develop curriculum materials (U.S. Department of Education, 1981).

for academic credibility, which seeks to withstand the test of time. The factor of continuity is obviously critical and portentously mortal for the implementation of these processes and contingent upon institutional commitment of resources and maintaining experienced personnel.

Discussion and Summary:

Before proceeding to a description of our models, in brief summation:

1. A pattern of decision making emerged with a *substantial though modest* number of directors making decisions for entry/exit level criteria in language assessment. As predicted, involvement of public school personnel would be minimal.

   One of the expected outcomes was the larger rate of participation of community personnel. Also as anticipated, the largest incidence occurred with programs sharing decisions between director and bilingual faculty, followed by combinations with faculty from other sectors than the immediate program. An interesting high correlation between student and community participation was an expected outcome ($n=63$, $r_s = .409$, $p<.05$).

2. An analysis of data revealed a tendency for director-made decisions to modestly correlate with a preference for commercial instruments, particularly discrete-point tests in other languages ($n=51$, $r_s = .3634$, $p<.05$). As predicted, assessment conducted by faculty members from sectors other than the immediate bilingual program showed similar selection of instruments for English and other target languages. Selections of other combinations of instruments, including observation measures and methods, were favored by respondents with shared decision making which involved combinations of sectors.

3. It was hypothesized that bilingual faculty would have a low number of years of experience in the public sector. Unexpectedly, an analysis of data showed a significant difference between experience in public schools, colleges/universities and academic preparation between directors of programs and bilingual faculty. The differences were large enough to lend to speculation as to the nature of institutional commitment to bilingual training programs.

   Significant data regarding the period of time of programs' existence and the number of years of teacher training experience for bilingual faculty raise a number of serious issues regarding the intentions by colleges and universities to maintain programs and retain faculty. Ultimately, decisions by institutions on commitment of resources and retention of experienced faculty will effect the nature of entry/exit-level criteria and assessment processes in target languages. A secondary source of data in support of these issues emerged from a 15% return which claimed discontinuation of programs by colleges/universities due to the cessation of Title VII funding. It is questionable whether programs will continue on institutional funds following termination of Title VII funding.

4. A strong correlation was observed between opportunities for remediation and those for language utilization. However, a distinction in entry/exit-level criteria was obvious in language assessment for remediation purposes and for language in the context of academic subject areas.

   Director-made decisions supported a trend away from contextual assessment, emphasizing remediation. A unilateral emphasis upon remedial assessment, as hypothesized, was observed with decision and assessment processes being made by faculty, other than bilingual, for English and foreign languages.

5. Although not uniform, to provide a general pattern, the data revealed some degree of influence, by a number of states with certification requirements in bilingual education, upon the assessment practices of respective respondents. In one of the most concerned and flexible of these states, representatives from bilingual training programs served on an advisory board to recommend assessment procedures and certification requirements. (Illinois Association of College and University Educators in Bilingual Programs, 1981; Seidner and Seidner, 1981, a; 1982).

6. The highest percentage of respondents cited assessment of target languages conducted by all bilingual faculty, followed by bilingual and other faculty, supporting the contention that sectors other than the immediate bilingual program within the university influenced assessment processes. Less than 10% of program directors unilaterally conducted program assessment. Instruments selected by foreign language departments and English language centers were mostly commercially produced. The utilization of commercial instruments suggested a potential subordination of program objectives and entry/exit-level criteria to that originally established in the construction of the instruments. In a few instances, the assessor for target languages was drawn from the community and/or public school.

7. As originally hypothesized, a greater heterogenous approach to assessment occurred in programs with increasing numbers of target languages. More combinations of observation, interview, and questionnaire techniques were observed to be employed. This may be partly due to a comparable lack of instruments in a number of target languages.

The Model

We now turn attention to theoretical models which reflect observations regarding our sample of respondents. I would like to emphasize that any observations are solely confined to our immediate sample. Three models are offered: (1) The first is centered around a homogeneous director-made decisions model for entry/exit level

criteria in language assessment. (2) The second reflects varieties of heterogeneous combinations of decision making. (3) The third typifies homogeneous sectors, other than director-made decisions.

Seven characteristics are offered in descriptions of these models which are limited to the sample. These include characteristics of: *1. Quantification/Observation, 2. Context of Approaches, 3. Flexibility of Instruments, 4. Criteria Influence, 5. Assessment Sectors, 6. Numbers of Languages, and 7. Duration of Program.*

Characteristics

1. **Quantification/Observation** (a): The choice of working for this particular characteristic emanates from Shuy's (1978) observations in designing the iceberg effect. *The selection of discrete-point instruments* (e.g., TOEFL) by some of the respondents in isolation or combination suggests more quantifiable and testable surface aspects but less so for what Shuy termed critical features (Semantic meaning and functional meaning). *Integrative testing* approaches the critical features more so by not isolating frequencies as much from natural context or reducing items into such small components for measurement as discrete-point tests. Although they may give a more global picture of language ability (e.g., "CLOZE procedures"), integrative tests are still difficult to measure in natural context (i.e., a context in which the language is not artificially elicited). Moreover, clarification is wanting as to what integrative tests actually meansure (also see Shuy, 1978). *A third selection, one of direct rating methods* has been regarded as having a high degree of content validity, assessing a full range of pertinent skills (also see Clark, 1978). *Problems with direct rating procedures*, such as those developed by the Foreign Service Institute, occur with questions of interreliability among raters. *The fourth selection of self-report ratings* raises questions as to reliability of responses from reporters. Theoretically, the subject acts from a base of honesty and accurately responds to the items. *The fifth selection of direct observation* appeared to correspond the most, in terms of Shuy, to critical features and was least quantifiable for surface structures (including videotaping, classroom observation, etc.)

   The first model reflected quantification/observation choices homogeneously with discrete-point or integrative testing. The second heterogeneous model represents more random combinations. The third, reflected homogeneous choices favoring the so-called critical areas.

2. **Context of Approaches** (b) Respondents were characterized by assessment processes either geared to language remediation (English/target language skills and communicative competencies), and/or bilingual academic subject area approaches. As noted earlier, a substantial association existed between respondents' provision for remediation opportunities and available academic experiences for target language utilization. Contextual testing, in the manner expressed by Bondaruk, Child and Terrault (Spolsky and Jones, 1975) refers to attempted measurement of an individual's ability to apply knowledge of grammar-lexis systems to specific points within discourse. Context of approach in the bilingual assessment process suggests linguistic applications, in contrast, to particular subject area points corresponding to selection of observation techniques by sampled respondents. A significant correlation emerged between director-decision making in entry/exit-level criteria and selection of discrete-point instruments for other target languages ($n=56$, $r_s=.3307$, sig.$=.013$, $p<.05$).

   In comparison, the highest incidence of observation techniques, included in combination with other approaches, accrued more readily with program faculty decision making.

   Here again, Model I was characterized by a homogenous preference for remediation-type approaches, Model II reflected combinations by respondents of approaches, while Model III favored homogeneous remediation type, on academic subject context observation techniques.

3. **Flexibility of Instruments** (c) This characteristic differs from the one previously cited in that it indicates to what degree of combination, respondents provide for instruments appropriate to described approaches. Spolsky and Jones (1975), Oller and Perkins (1980) among others, endorsed the premise of using combinations in assessment approaches. The heterogeneous model reflected the tendency toward the greatest flexibility of approaches with over 70% of the sample respondents. The remaining percentages characterized one type of instrument, when utilized, at both ends of the spectrum for Model I and for Model III.

4. **Criteria Influence** (d) An important consideration for our model is the degree of influence from sectors outside the program upon the determination of language assessment criteria. Internal administrative pressure may be a consideration in institutions for such programs as the one for Chamorro, in Guam. At the same time, a varying degree of influence was experienced with faculty other than the bilingual program in determining English and other target languages. An interesting pattern, cited above, was observed with the selection of a particular instrument (e.g., FSI, MLAT, etc.) by programs in states with particular mandates. The factor of Title VII requirements was another consideration this study was unable to ascertain regarding degrees of influence.

   The limitations of interpretations for the particular sample are once again underscored, given the changing nature of the Title VII regulations. The data failed to place criteria influence as tangential to one model to the exclusion of others. They appeared to some degree as characteristics for all three models.

5. **Assessment Sectors** (e) Our fifth characteristic directly relates to previously cited concerns as to sources of decision making and assessment of languages. In brief summation, director-made decisions characterized the first homogeneous model, while combinations of sources typified the heterogeneous second model. The third model was characterized by homogeneous sectors, other than the director.

6. **Number of Languages** (f) Another prevailing feature for our models was the number of languages. Most programs cited one language in addition to English. The general tendency in this category was toward Spanish, although an occasional target language such as Yupik would appear in combination with English. As with the characteristic of criteria influence, this characteristic was found tangential to one model, to the exclusion of others. However, the director-made decision homogenous model tended to subscribe to the Spanish/English category.

7. **Duration of Program** (g) The last characteristic refers to the number of years of existence for sampled programs. Aristotle tells us that "even men of experience succeed more than those who have theory without experience" (cited in Broudy, 1962). It would seem to follow that longevity, in the sense of consistency of program existence, would yield some quality of experience. A general tendency emerged in which programs with increasing years of experience beyond the mean exhibited a parallel preference for combinations of instruments and more flexible choices for assessment approaches. The highest frequency of these responses occurred with the heterogeneous Model II, which was characterized by a variety of assessment approaches. The data also inferred that the majority of the two homogeneous models were in existence for a few years, in comparison to the heterogeneous one.

Diagram 1 presents us with a summarized overview of characteristics pertaining to our three theoretical models. Clearly these are not to be considered absolute within their theoretical framework, in view of aforementioned date which substantiates ranges of variation between stipulated models. These variations are therefore implicit within the given structure.

Conclusions and Recommendations

1. Any assumptions drawn from this data must be ultimately limited to the samples of respondents. This is largely due to:

   a. a lack of substantial data base which is also consistent. Unfortunately, the literature has yielded little or no research to make definite comparisons. An obvious need exists for further research in areas covered within this paper and related aspects.

   b. limitations of investigation approaches for this study. The nature of survey research whether questionnaire and/or interview techniques are still subject to question, especially in view of the lack of substantiating verification from other studies. However, these approaches remain the most effective existing means for gathering this type of data.

   c. a potential shifting of population mean (u). A replication of this study will be most welcome to provide much needed additional data. The shifting tides of federal funding and university responsiveness to a need for training bilingual educational personnel may, for example, contribute to altering the nature of existing programs.

Summarized Overview of Characteristics for Theoretical Models

Model/Characteristic	1	2	3	4	5	6	7
I	Preference for discrete-point and integrative tests	Preference for remedial context	Preference for one instument	Relative number/degree of influence sectors	Homogeneous director-made decisions	Variable degree of influence	Usually less than mean number of years
II	Variety of combinations of instruments	Preference for variety of context	Preference for combination	Relative number/degree of influence sectors	Heterogeneous combinations of decision makers	Variable degree of influence	Usually exceeded mean number of years
III	Preference for observation-type instruments	Preference for academic subject context	Preference for one instrument*	Relative number/degree of influence sectors	Homogeneous sectors of decision makers other than directors	Variable degree of influence	Usually less than mean number of years or more

*when utilized

2. The observed discrepancy between program faculty and directors in experience and academic background suggests some very strong concerns regarding the caliber of individuals establishing entry/exit-level criteria. Future studies might perhaps explore these factors in relation to other training concerns. This would appear critical since a sizeable number of the sample expressed between a low to average satisfaction with their own assessment processes (54.7%). Similar indices might be revealed in future studies of teacher training practices.

3. Whether the result of a relatively short span of existence, deliberate institutional policy, or a combination thereof, respondent programs revealed a serious deficit in trained personnel. Mandatory institutes, such as the one for leadership training, should be augmented on a regional level with performance-based sessions on comprehensive language assessment approaches.

4. The federal grant to institutions should be treated in its strictest sense as a contractual obligation. As such, long-range commitments by the college or university, beyond the federal funding scope, would be a desirable inclusion, with the consequence of total repayment by the institution of expenditures upon breach of contract. This business-as-usual approach is not unrealistic and would assist funding sources in their efforts to determine those institutions with sincere commitments from the short-range expediency of others.

5. The assessment processes, which are outlined in the contract, should include (along the same lines of reassuring in point 4) the mandatory participation of school personnel and/or community resources. A distinction between "real" and "token" participation can be established through a separate evaluation of these sectors which would serve as a check and balance.

6. More effective federal monitoring of training programs and their environments seems desirable from several vantage points. Investment in short-term increases in visits and cross checks from independent sources may in the long run prove cost-effective in justifying expenditures for programs. At the same time, these visits may serve as an effective deterrent to potential abuses.

7. Attention has been given to describing target competencies for teachers of limited-English proficient populations in public schools. A need exists to delineate similar competencies for their teacher trainers in colleges and universities. As the locus of funding power in the nation, the Office of Education might initiate this process in collaboration with national and/or regional accrediting bodies. Incorporation of flexible criteria into federal guidelines could ultimately contribute to providing target personnel who would better serve limited-English proficient populations, rather than catering to administrative structures and priorities of institutions.

B. Training programs with increasing numbers of target languages may pose problems in the acquisition of personnel and materials, the structure and schedule of academic coursework and remedial classes, as well as the establishment of language entry/exit-level criteria. The commitment of federal, institutional and perhaps in a number of instances, state resources is essential toward providing the personnel and environment.

The need to focus our attentions upon the college/university sector is increasingly imperative, given the myriad of complex political, social and economic pressures on the horizon, which continue to threaten the existence of programs for limited-English proficient students currently attending public services.

With this in mind, it is hoped that this study is a beginning in a series of others which will efficaciously increase our current knowledge regarding teacher training processes in colleges and universities, promote educationally sound decisions and improve upon current practices.

Select Bibliography

Academy for Education Development. *319 Ways Colleges and Universities Are Meeting the Financial Pinch.* New York: A.E.D., 1971.

Acosta, Robert (Kelly) and Blanco, George. *Competencies for University Programs in Bilingual Education.* Washington, D.C.: Government Printing Office, 1978. H.E.W., OE 78-07903.

Andersson, Theodore and Boyer, Mildred. *Bilingual Schooling in the United States.* 2 vols. Austin: Southwest Educational Development Laboratory, 1970.

Arciniega, Tomas. *Preparing Teachers for Mexican Americans.* Nestor, Texas: National Education Laboratory, Publishers, 1977.

Arends, Robert. *Handbook for the Development of Instructional Modules in Competency-Based Teacher Education Programs.* Buffalo: Authors, 1973.

Babbie, Earl R. *The Practices of Social Research.* Belmont, California; Wadsworth, 1979.

Balasubramonian, K. *Measurement, Evaluation and Accountability in Bilingual Education Programs.* Rosslyn, Va.: N.A.C.B.E., 1979.

Barrera, Maria. *Bilingual Education.* Austin: Region XIVI Education Service Center, 1973. (Report presented to the Texas Legislature.)

Bilingual Education Service Center. *Implementing Bilingual Education Teacher Training: Report of a Conference.* Racine, Wisconsin, 1977.

Binkley, Joanne L. and Johnson, Donna, et al. *A Study of Teacher Training Programs in Bilingual Education Vol. I: Program Descriptions.* Mountain View, California: R.M.C. Research Corporation, 1981 (preliminary report).

Blanco, George. "Competencies Needed by Bilingual Education Teachers," *Educational Leadership,* no. 35, 1977, pp. 123-127.

_____ "The Education Perspective," Rudolph Troike, ed., *Bilingual Education Current Perspectives,* vol. 4. Arlington: C.A.L., 1977. pp. 1-66.

_____ "Language Assessment in a Bilingual Teacher Education Program." Paper presented at the annual meeting of A.E.R.A., San Francisco, California: April 20, 1976.

Bolger, Philip Albert. "The Effect of Teacher Spanish Language Fluency upon Student Achievement in a Bilingual Science Program." Unpublished Ph.D dissertation, St. John's University, 1967.

Brisk, Maria Estela. *Bilingual Higher Education Programs: Their Impact on Institutions and Community.* Los Angeles: National Dissemination and Assessment Center, 1978.

Broudy, Harry S. *A Critique of Performance-Based Teacher Education.* American Associaton of Colleges for Teacher Education, 1962, p. 7.

Burke, Cassel. *The Individualized Competency-Based System of Teacher Education at Weber State College.* Washington, D.C.: American Association of Colleges for Teacher Education, 1972.

California State Department of Education. *Criteria for Bilingual Teacher Competencies.* Sacramento: State Department of Education, 1977.

Carey, Lou and Marsh, David. *University Roles in Inservice Education: Planning for Change.* Washington, D.C.: A.A.C.T.E., 1980.

Carnegie Council on Policy Studies in Higher Education. *Three Thousand Futures: The Next Twenty Years in Higher Education.* San Francisco: Jossey-Bass, 1980.

_____ *Fair Practices in Higher Education: Rights and Responsibilities of Students and Their Colleges in a Period of Intensified Competition for Enrollments.* San Francisco: Jossey-Bass, 1979.

Carrillo, Frederico. *The Development of a Rationale and Model Program to Prepare Teachers for the Bilingual-Bicultural Secondary School Programs.* San Francisco: R and E Research Association, 1977.

Castillo, Max. *Developing Competency-Based Bilingual Teacher Education.* Houston: University of Houston, 1976.

_____ "A Study of Competency Behaviors for Elementary Teachers in Bilingual Environments." Unpublished Ph.D. dissertation, University of Houston, 1975.

Center for Applied Linguistics. *Guidelines for the Preparation and Certification of Teachers of Bilingual/Bicultural Education.* Arlington: C.A.L., 1974.

Chamot, Anna Uhl. "Teachers Urged to Respond to Survey," *Forum,* IV, 1, January, 1981, p. 3.

Charters, W. W. "Beyond the Survey in School Board Research", *Educational Supervision and Administration,* no. 41, 1955, pp. 449-452.

Clark, John. "Direct Testing of Speaking Proficiency: Theory and Application." Proceedings of two-day conferences at Georgetown University, March 14-15, 1978, ERIC AN-BE00106B.

Conrad, Herbert S. "Contributions of the Federal Government to Educational Research Methodology," in Frank Banghart, ed., *Educational Research.* Bloomington, Ind.: Phi Delta Kappa, 1960, pp. 23-41.

Cordasco, Francesco. *Bilingual Schooling in the United States.* New York: McGraw-Hill, 1976.

Cornejo, Ricardo. "The California Language Census Survey: Methodology Issues," paper presented at American Education Research Association, March, 1978. ERIC, ED 153486.

_____ *A Synthesis of Theories in Research on the Effects of Teaching in First and Second Languages.* Austin: National Educational Laboratory Publishers, Inc., 1974.

Cummins, Jim. "Entry and Exit Fallacy in Bilingual Education", *NACBE Journal,* IV (3), 1980. pp. 25-59.

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

Curtis, Johnathan; Lignon, Glynn D. and Weibly, Gary W. *When Is a L.E.P. Student No Longer L.E.P.?* Los Angeles: N.D.A.C., 1980.

De Inclan, Rosa G. "School-Community Relations," *NABE Journal,* I(2), 1976, pp. 73-77.

Department of Health Education and Welfare. *First Report on the Condition of Bilingual Education in the Nation.* Washington, D.C.: Government Printing Office, 1976. OE 77-01704.

Dissemination and Assessment Center for Bilingual Education. *Guide to Title VII ESEA Bilingual Bicultural Programs.* Austin: DACBE, 1979.

Dissemination Center for Bilingual, Bicultural Education. *Procedures: National Conference on Bilingual Education.* Austin: DACBE, 1973.

Duker, Sam. "The Questionnaire Is Questionable." *Phi Delta Kappan*, no. 29, 1948, pp. 386-392.

Dulay, Heidi and Burt, Marina. "The Relative Proficiency of Limited English Proficient Students," *NABE Journal*, IV(3), 1980, pp. 1-23.

Estes, Gary C. and Estes, Carole. "What Is the Criterion of Interest in Identifying Limited English-Speaking Students: Language Dominance or Proficiency?" Paper presented at American Educational Association, San Francisco, April 1979, ERIC, EO 178613.

Flores, S. "The Training of Bilingual-Bicultural Personnel," *Bilingual Education* (Proceedings from Aspira National Bilingual Education Think Tank) March 1973, pp. 49-65.

Franzen, Raymond and Lazarsfeld, Paul. "Mail Questionnaire as a Research Problem," *Journal of Psychology*, no. 20, 1945, pp. 293-320.

Freytes, Celeste E. and Rivera, Charlene. *Inservice Education Models for Classroom Educators to Assess Language Proficiency in Bilingual Students.* Los Angeles: N.D.A.C., 1979.

Gaarder, A. Bruce. "The First Seventy-Six Bilingual Education Project," James E. Alatis, ed., *Bilingualism and Language Contact.* Washington, D.C.: Georgetown University Press, 1970.

Gilbert, A. Jarvis and Adams, Shirley J. *Evaluating a Second Language Program.* Arlington: Center for Applied Linguistics, 1979.

Gonzalez, Josue. *Towards Quality in Bilingual Education and Bilingual Education in the Integrated Schools.* Rosslyn, Va.: N.C.B.E., 1979.

Hardy, Roy. "Should Jose Be Tested in Spanish?" Paper presented at National Council on Measurement in Education, April, 1977, ERIC, ED 163 043.

Hunter, William. *Multicultural Education through Competency-Based Teacher Education.* Washington: American Association of Colleges for Teacher Education, 1974.

Illinois Association of College and University Educators. "Bilingual Agenda for Teacher Certification Meeting", Woodstock Center, May 19-20, 1981 (mimeo).

Illinois Board of Higher Education. *Bilingual Teacher Preparation Conference Report.* Springfield: Board of Higher Education, 1974.

Kersh, B. Y. *Faculty Development for Inservice Education in the Schools.* Washington, D.C.: A.A.C.T.E., 1978.

Kramer, Martin. *The Venture Capital of Higher Education.* San Francisco: Carnegie Foundation, 1980.

MacCaulay, Ronald S. and Ramirez III, Manuel. *Spanish-English Bilingual Education in the United States: Current Issues, Resources, and Recommended Funding Priorities for Research.* National Institute of Education: N.I.E., C-74-0151, 1975.

Martinez, Federico. "The Development of a Rationale and Model Program to Prepare Teachers for the Bilingual-Bicultural Secondary School Programs." Unpublished Ph.D. dissertation, The University of New Mexico, 1975.

Matluck, Joseph H. "The Multiple Role of Oral Language Assessment: Language Status, Research and Curriculum Development," paper presented at International Reading Association, May, 1973, ERIC, ED 161058.

_____ and Mace-Matluck, Betty J. "Language Characteristics of Mexican-American Children: Implications for Assessment," *Journal of School Psychology*, (11(4), 1973, pp. 365-386.

Maulden, Michael, et al., "Language Assessment Procedures for Identifying LESA Students," paper presented at American Educational Research Association, April, 1979, ERIC, ED 175933.

Mayhew, Louis B. *Surviving the Eighties: Strategies and Procedures for Solving Fiscal and Enrollment Problems.* San Francisco: Jossey-Bass 1979.

Merino, Barbara J.; Politzer, Robert and Ramirez, Arnulfo. "The Relationship of Teachers' Spanish Proficiency to Pupils' Achievement," *NABE Journal*, I:I (2), 1979, pp. 21-37.

Merino, Barbara J. and Politzer, Robert. *Toward Validation of Tests for Teachers in Spanish/English Bilingual Education Programs.* Research and Development Memorandum No 151, Stanford Center for Research and Development in Teaching, 1977.

Michigan State Department of Education. *A Position Statement on Bilingual Instruction in Michigan.* Michigan Department of Education, 1976.

_____. "Guidelines for Selecting Test Instruments and Procedures for Assessing the Needs of Bilingual Children and Youth," 1976. ERIC, ED 161951.

Migdail, Sherry Resnick. "An Analysis of Current Select Teacher Training Programs in Bilingual-Bicultural Education and the Development of New Teacher Training Designs," unpublished Ed.D. dissertation, American University, 1976.

Mouly, George J. *Educational Research: The Art and Science of Investigation.* Boston: Allyn and Bacon, Inc., 1978.

National Clearinghouse for Education. *Guide to Title VII ESEA Bilingual Education Programs 1979-80.* Rosslyn, Va.: N.C.B.E., 19B0a.

_____ *Guide to State Agencies.* Rosslyn, Va.: N.C.B.E., 19B0.

National Consortium for Bilingual Education. *Report on Survey Findings: Assessment of Bilingual Education Programs.* Fort Worth: N.C.B.E., 1971.

National Dissemination Center for Bilingual Education. *Teacher Education Programs for U.S. Colleges and Universities.* Austin: DACBE, 1975, 1976.

National Foundation for the Improvement of Education. *Public Education Strategies: An Information Dissemination Program for Bilingual Education.* Rosslyn, Va.: National Clearinghouse for Bilingual Education, 19B1.

N.Y.S.U.T., Board of Directors, "Position Statement of Bilingual Education," New York State United Teachers, November 22, 1975.

Okada, Mashito and Sancedo, Thomas; Chairpersons. "Development of Entry/Exit Criteria and Associated Assessment Procedures for Bilingual Education Projects." Symposia Sessions, abstracted in American Education Research Association Annual Meeting, 19B0 (Abstract-ISSN-00B4-634).

Oller, John and Perkins, Kyle. *Research in Language Testing.* Rowley, Mass.: Newbury House, 19B0.

_____ *Language Testing: Testing the Tests.* Rowley, Mass.: Newbury House, 197B.

Palmer, Judith Walker. *Developing a Competency-Based Bilingual Teacher Training Program.* Arlington: C.A.L., 1975, EDRS, ED 10B525.

Pascual, Henry. *Clients and Teachers in Bilingual Education.* Los Angeles: N.D.A.C., 1979.

Phillips, William M. "Weakness of the Mail Questionnaire: A Methodological Study", Sociology and Social Research, no. 35, 1951, pp. 260-267.

President's Commission on Foreign Language and International Studies. *Strength through Wisdom: A Critique of U.S. Capability.* Washington, D.C.: Government Printing Office, 1979.

Ramirez, Arnulfo G. "Teaching Reading in Spanish — A Study of Teacher Effectiveness," *Reading Improvement,* 16 (4), 1979, pp. 304-313.

_____ "Language in Bilingual Classrooms," *NABE Journal,* IV (3), 19B0, Pp. 61-79.

Ramirez, Frank. "Marketability of Bicultural/Bilingual Teachers in 1974 in the Los Angeles Unified School District," unpublished Ph.D. dissertation, Claremont Graduate School, 1974.

Ramirez III, Manuel; Gonzales, A., et al. "Research Priorities in Bilingual Education in the United States: Challenges for the Present and Future," William Mackey, Jacob Ornstein, et al., *The Bilingual Education Movement.* El Paso: University of Texas, 1977, pp. 36-47.

Rhode, R. "Bilingual Education for Teachers," *Educational Forum.*, no. 3B, 1974, pp. 203-209.

Ringawa, Marcel. "Cultural Pedagogy: The Effects of Teacher Attitudes and News in Selected Bilingual Bicultural Environments," Ray Padilla, ed. *Ethnoperspectives in Bilingual Education Research. Vol. II: Theory in Bilingual Education.* Ypsilanti: Eastern Michigan University, 19B0, pp. 372-3B7.

Rosenberg, Morris, *The Logic of Survey Analysis.* New York: Basic Books, 196B.

Ruekmick, C.A. "The Uses and Abuses of the Questionnaire Procedure," *Journal of Applied Psychology,* no. 14, 1930, pp. 32-41.

Sanchez, G. I. "The Education of Bilinguals in a State School System," unpublished Ph.D. dissertation, University of California at Berkeley, 1934.

Saville, Muriel R. and Troike, Rudolph C. *A Handbook of Bilingual Education.* Washington, D.C.: TESOL, 1971.

See, Harold W. "Send It to the President," *Phi Delta Kappan,* no. 3B, 1957 p. 130.

Seidner, Stanley S. "Language Assessment at Post-Secondary Institutions," Raymond Padilla, ed., *Ethnoperspectives in Bilingual Education Research: Vol. III.* Ypsilanti: Eastern Michigan University, pp. 36B-3B0, 19B1a.

_____ "Discrimination in Non-Discriminatory Testing," Bilingual Multicultural Symposium, University of Illinois at Chicago Circle, July 24, 19B1b.

_____ "New Implications for Bilingual Testing," Paper delivered at the National Association for Bilingual Education, Anaheim, California, April 21, 19B0.

_____ "Correlates of Language, Ethnicity and Power," Paper delivered in New York Association for Bilingual Education, February 14, 19B0b.

_____ "Emergence of Bilingual Education: An Overview of Curriculum Developments," *Thresholds in Education,* IV (1), February, 197B, pp. 13-16.

_____ and Seidner, Maria Medina. "In the Wake of Conservative Reaction, Two Parallel Analyses", Raymond Padilla, ed., *Bilingual Education Research:* Vol. IV. Washington, D.C., N.C.B.E., 19B2 (in press).

_____ "A Review of Criteria for Teacher Training and State Certification." Paper delivered at 10th Annual International Bilingual Bicultural Education Conference, Boston, May 29, 1981.

_____ "Surviving the 1980's: Bilingual Education in a Politically Hostile Environment." Presented at TESOL Conference, Champaign-Urbana, April 10, 1981b. (To be published in proceedings.)

Shannon, John R. "Percentage of Returns of Questionnaires in Reputable Educational Research," *Journal of Educational Research*, no. 42, 1948, pp. 138-141.

Shuy, Roger. "Problems in Assessing Language Ability in Bilingual Programs," in Herman La. Fontaine, ed. *Bilingual Educator*. Wayne: New Jersey 1978, pp. 376-380.

Spolsky, Bernard and Jones, Randall, eds. *Testing Language Proficiency*. Arlington, Va: Center for Applied Linguistics, 1975.

Stanier, John Donald. "A Model of a Bicultural School-Community Linkage Component of Instructional Coordinators." Unpublished Ph.D. dissertation, University of Pittsburgh, 1971.

Sutmin, Frances. *Educating Personnel for Bilingual Settings: Present and Future*. Washington: American Association of Colleges, 1979.

Swanson (Seidner), Maria. *Bilingual Education, The National Perspective*. Arlington Heights: Illinois B.E.S.C., 1974 (reprinted from Gilbert A. Jarvis, ed., *A.C.T.F.L. Review*, vol. 5. Skokie: National Textbook Co., 1974).

Texas Education Agency. *A Statewide Design for Bilingual Education*. Austin: Texas Education Agency, n.d.

Topp, Robert F. and McGrath, G. D. "About the Questionnaire-Answer It" *School Executive*, no. 70. 1950, pp. 59-60.

U.S. Civil Rights Commission. *Teacher and Student: Differences in Teacher Interaction with Mexican-American and Anglo Students*. Washington, D.C.: Government Printing Office, 1973.

_____. *Teachers and Students, Report V: Mexican-American Education Study*. Washington, D.C.: Government Printing Office, 1973.

U.S. Department of Education. *News Release*, February 4, 1980.

U.S. National Center for Statistics. *Financial Statistics of Higher Education: Current Funds, Resources and Expenditures*. Washington, D.C.: Government Printing Office, 1979, 1980.

Valencia, Atilano. *The Effects of a College Teacher Training Project with Emphasis on Mexican-American Cultural Characteristics* (evaluation report). Sacramento, California: Sacramento State College, 1970.

Vazquez, Jose A. "Competency-Based Teacher Education and the Bilingual Teacher: New York State," *NABE Journal*, I(2), 1976, pp. 79-81.

Viera, Silvia; Squires, Leslie; and de Guevara, Gloria. *A Framework for the Training of Bilingual/ESL Teachers in the School of Holyoke, Massachusetts*. Amherst: University of Massachusetts, 1975.

Waggoner, Dorothy. "Teacher Resources in Bilingual Education: A National Survey," *NABE Journal*, Vol. III(2), 1979, pp. 53-60.

_____. "Non-English Background Persons. Three U.S. Surveys", *TESOL Quarterly*, Vol. XII(3), 1978, pp. 247-262.

_____ *State Certification Requirements for Teachers of Bilingual-Bicultural Education Programs*. Washington, D.C.: Government Printing Office, 1977.

Walsh, Sister Marie Andre. "The Development of a Rationale for a Program to Prepare Teachers for Spanish-Speaking Children in Bilingual-Bicultural Elementary School." Unpublished Ph.D. dissertation, University of Texas at Austin, 1971.

Weber, Wilford, *Competency-Based Teacher Education: A Scenario*. Washington, D.C.: American Association of Colleges for Teacher Education, 1972.

Young, Pauline V. *Scientific Social Surveys and Research*. Englewood Cliffs, New Jersey, Prentice-Hall, 1966.

ISSUES OF SPECIAL EDUCATION AND LANGUAGE ASSESSMENT FOR THE LIMITED-ENGLISH PROFICIENT STUDENT

by
Antônio Simões, Jr.
Rhode Island College
Providence, Rhode Island

When I was asked to write about special education or the special needs of the linguistic minority, I was quite reluctant to lay claim to an area that was out of my expertise (special education). I did not respond to the invitation until I was sure that my expertise, curriculum theory-bilingual education, would be a valuable contribution to the field of bilingual-special education. I finally accepted this challenge when I realized that for me to talk about special education from a technical point of view would be foolish, if not outright dishonest, but still, I could analyze the concept of "special needs" if this domain were analyzed from a sociopolitical paradigm.

The purpose of this paper is, then, to analyze the concept of special needs as described from a sociopolitical level. This paper will *not* deal with specific treatments in cases of mental retardation, hyperactivity or other conditions that may have some medical history. I will leave these issues to the medical field where proper and competent professionals are familiar with the issues.

The Concept of Category in a Special Needs Situation

The concept of category or the process of categorization is a logical place to begin our analysis. Some of the literature assumes that there is always a direct relationship between special education and bilingual education. For example, Chapter 71A; the Transitional Bilingual Education Act of Massachusetts, the Courts' Declaration of Policy was the following:

> Section 1. Declaration of Policy
>
> "The General Court finds that there are large numbers of children in the Commonwealth who come from environments where the primary language is other than English. Experience has shown that public school classes in which instruction is given only in English are often inadequate for the education of children whose native tongue is another language. The General Court believes that a compensatory program of transitional bilingual education can meet the needs of these children and facilitate their integration into the regular public school curriculum. Therefore, pursuant to the policy of the Commonwealth to insure equal educational opportunity to every child, and in recognition of the needs of children of limited English-speaking ability, it is the purpose of this act to provide for the establishment of transitional bilingual education programs in the public schools and to provide supplemental financial assistance to help local school districts to meet the extra costs of such programs."

The concept of categorization through this act is in that transitional bilingual education becomes compensatory in nature, in which the bilingual population must eventually become part of the "regular" public school curriculum. Through this act public schools were now mandated to develop transitional bilingual education under specific definitions. They were:

1. Classrooms in this domain were compensatory in nature.
2. They were not part of the regular program.
3. The transition toward English was the objective of the program. Native language instruction was eventually eliminated.

This political-educational pardigm becomes now a "special" mission or a "special education" process for a specific population, the linguistic minority. It does not accept these children as *regular* children who just happen to speak another language. The political/pedagogical consequences of these kinds of definitions may take different avenues. In this case, bilingualism is not an asset, but a deficit, for a special population.

Federal legislation also views bilingualism from the point of view that the limited-English proficient child has a deficit (not speaking English) and is thus categorized from a negative point of view. The Bilingual Education Act of 1968 (popularly known as Title VII) is a classic example. Although it is believed that federal legislation was passed to preserve language and culture outside the English language, the federal legislation was passed under different assumptions. First, the long-range goal was not bilingualism, but proficiency in English (National Institute of Education 1975:6) Other stipulations such as poverty criteria or percentages of proficient English-speaking children in a classroom, classify a specific population outside mainstream society. Again, it is important to note how society at large and the educational enterprise view a special population with "special problems."

The concept of categorization becomes more complex when one questions the idea of "special," "abnormal," limited-English proficient" or other concepts that are applied to educational theory and practice. If one assumes that within the human range of behavior and conditions there is a diversity of differences, then all behavior or conditions should be treated as "normal" education. That is, there is nothing special or different outside the so-called deferred norm, but all behaviors, including speaking a different language, are conditions that have pedagogical or educational solutions. Once a behavior

is identified as a concept that belongs to a "special" category, then hegenomy may take place and the behavior is placed outside the norm. Bergin cites the process of categorization when she analyzes the Riverside Study. She states that:

1. Public schools were the major labelers.
2. Public schools shared their labels widely with other agencies.
3. Black and Spanish-surnamed children were more likely to score seventy-nine or below on an I.Q. test than Anglo children.
4. Among those scoring below seventy-nine on an I.Q. test, children who were Spanish-surnamed and who were from low socioeconomic levels were more likely to be placed in special classes.
5. Only 19 percent of the children placed in classes for the mentally retarded ever returned to the mainstream school program.
6. Black and Spanish-surnamed children were "over-labeled" as mentally retarded and Anglos were "under-labeled."

These conclusions are frightening if one follows through the logic that it is essential to label or categorize for educational purposes. One could easily conclude that children who are Black or who have a Spanish surname tend to have low I.Q.'s and very rarely return to "mainstream" education. On the other hand, Anglos tend to have less special education problems than Blacks and Hispanics. The literature of incorrect labeling and the School's legitimizing of "special" problems is substantial. Beeghley and Butler talk about the consequences of intelligence testing in the public schools before and after desegregation. They state that the schools functioned to facilitate differential labeling of mental retardates. The labeling process has resulted in a form of institutional racism in special education programs regardless of integration in the regular classroom (p. 746). In other words, the desegregation process in itself did not alleviate labeling children and the schools continued to categorize racial or linguistic minorities as "mentally retarded" or place them in "special" education classes. Beeghley and Butler cite the following statistics in their research. (p. 746)

Table I

PROGRAM PARTICIPATION AND ETHNICITY: 1968

Type of Program	Ethnicity Anglo	Black	Chicano
Educable Mentally Retarded	54%	82%	87%
Trainable Mentally Retarded	9%	5%	2%
Educationally Handicapped	19%	7%	5%
Physically Handicapped	18%	7%	7%
	100% (347)	100% (126)	101%* (201)

χ^2 equals 78.018 P .001
*Rounding Error

If one analyzes these statistics carefully, Black and Chicano children have a high percentage of "educables," while Anglo children are less categorized as "educables." When any school system or society accepts that a specific population has more "mental retardation" than another race or group, one must take a second look.

Kamin studies this issue from a sociopolitical perspective. His research illustrates that I.Q. tests and their use as classification instruments are not neutral. His analysis suggests that biases are not only based in the instrument but also in our "deep culture," present and past, which helps to support misclassification (p.317).

The literature of the misuse of instruments and the misclassification of students is abundant to conclude that the concept of categorization, especially in the field of bilingual-special education, must be used with extreme caution. This concept should be further examined when one approaches the idea of statistical norming in a special-bilingual education population.

We will now look at this issue of what might be categorized as "normal" and what is categorized as "special."

The Normal — Special Education Paradigm: A Look from a Minority Perspective

I believe it is correct to suggest that not everyone has the same abilities to perform general and specific tasks. There is a range of human behavior and human ability in humanity, regardless of the language use of the individuals. It may be possible that a statistical norm may belong to *each* language group because each group may have its own "norming effect" toward a sociopolitical norm. That is, what is "standard" or what is a "norm" is a political categorization even if one can legitimize a norm through statistical means. Haugen talks about norm from another perspective. When he talks about language use, 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, a kind of 'old shoe.' (p. 91)

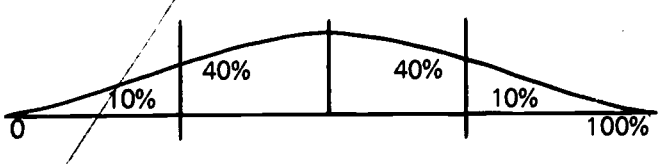
Here, the concept of norm becomes problematical in that social class language becomes a valued norm, but it still can be used as a statistical norm to legitimate a curve to differentiate between "normal" or "standard" language and "nonstandard" language. In other words, if one uses a norm to divide dialects from "standard language" from the point of view of categorization, many of our children may be incorrectly placed in "special" situations. Although this logic seems obvious, there are still

many complex issues that revolve around language use in a monolingual and bilingual environment. Placing children in a statistical norm does not address the linguistic issues in a so-called bilingual-special education situation.

It is true, however, that educators must use some definitions to prescribe educational environments for students. As stated previously, within any linguistic population, there is a range of abilities — what might be categorized from "mental retardation" to the "exceptional" child. We could accept behaviors and languages as "states of being" during one's lifetime and this concept could lessen the impact of the category of "special." That is, within any specific language population, the so-called bell-shaped curve describes different "states of being" during one's lifetime, and this concept could lessen the impact of the category of "special." Within any specific language population, the so-called bell-shaped curve describes different "states of being" that identify specific needs and prescribe specific educational environments. One may now argue that the problem of special education and the problem of states of being is only in semantics and the concept of categorization has not been eliminated. The key to solve this problem may be "norming" specific language groups and prescribing educational environments within each language group. To make this problem logical, let us examine Figure I.

Figure I

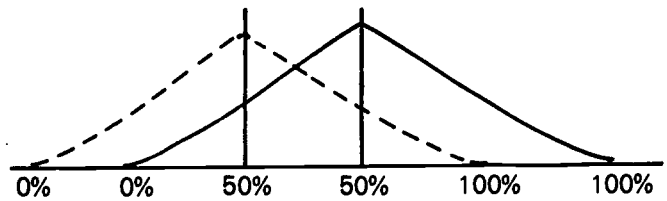
A STATISTICAL NORM FOR ABILITIES IN THE UNITED STATES



In Figure I, the curve may be used to identify placement of individuals with reference to general and specific abilities. Children are usually classified in percentiles with reference to categories, i.e., reading comprehension, math ability, social studies and so on. If a child is below the 50th percentile, he or she is considered "below the norm," and, of course, if the child is above the 50th percentile, he/she is considered above the norm. It may be correct to suggest that this curve is more or less accurate when one is speaking only of a particular linguistic community, but Table I implies that more Blacks and Chicanos are below the "norm" by the percentages of educable mentally retarded students in each population. If one recalls, there were significantly more Black and Chicano children placed in "special" education classes than Anglos. The data seems to suggest that we must rethink the statistical norm to fit the linguistic language minorities in the United States. Let us assume that each linguistic population has its own bell-shaped curve within its own cultural or political context. If one accepts this paradigm ar if one accepts the current research data that racia and linguistic minorities have a higher placement in "special education" classes, Figure II may illustrate a visual description of our problem.

Figure II

NORMING FOR DIFFERENT LINGUISTIC POPULATIONS



– – – – – Language Minority Norm
——— English or Anglo Norm

One may note that the "English-Anglo" curve or norm and the "Language Minority Norm" come from two different "standards," and both curves have their own statistical "mean." Hence, if one acquires the skills and expertise to identify "special" situations in each language group, the placement of minority children will be less political.

It is important to note that the author of this paper is *not* suggesting that each population should be segregated totally within its own "special education" situation. (Which, by the way, becomes a monolingual situation instead of a bilingual environment) The author is implying that the process of categorization as it now takes place in special education within the context of language is now labeling and misplacing language minority children in "special" situations in which they do not belong. Simply said and accepting all the complexities, some of the children do not have the "knowledge goods" or "cultural capital" (English) to compete with another population. The present process seems to segregate more children into "special" situations because of the lack of instruments for the language minority populations. This is a process that is unacceptable in our educational institutions.

Some Comments and Conclusions on the Process of Categorization in a Special-Bilingual Education Situation

We must confront the issue that most paradigms that categorize human beings are political in nature, except where severe physical handicaps are involved. The "liberalism" of the sixties and of the seventies which produced new systems or categories to legitimately acquire a process of equal opportunity for all groups may have failed because it failed to take into account the "deep culture" of its own value system. It seems to be correct when some social scientists and educators stated that in the final analysis, the new sociology fails in spite of its desire for radical and fundamental change with reference to the concept of egalitarianism. Its failure is in the inability to illuminate how social and political structures function to mask reality and promote ideological hegemony (Gramsci, 1971; Entwistle, 1978; Giroux, Penna, 1979). To justify this situation is to continue segregation, deny equal access to the minority and eventually maintain two societies, separate and unequal. I hope this is not the case.

Bibliography

Beeghley, Leonard and Butler, Edgar. "The Consequences of Intelligence Testing in the Public Schools Before and After Desegregation," *Social Problems*, #21 p. 740-754, 1974.

Bergin, Victoria. *Special Education Needs in Bilingual Programs*, Rosslyn, Virginia: National Clearinghouse for Bilingual Education, 1980.

Haugen, Einor. "Norm and Deviation in Bilingual Communities," *Bilingualism* (Hornby, Peter, Editor) New York: Academic Press, 1977.

Kamin, Leon J. "Social and Legal Consequences of I.Q. Tests as Classification Instruments: Some Warnings from Our Past," *Journal of School Psychology*, Vol. 13, No. 4, 1975.

Massachusetts, Commonwealth of, *Two Way*, Boston: Bureau of Transitional Bilingual Education, 1976.

National Institute of Education, *Spanish-English Bilingual Education in the United States: Current Issues, Resources and Recommended Funding Priorities for Research*, Ms, 1975.

THE EFFECTS OF BILINGUALISM ON INTELLIGENCE: A CRITICAL REVIEW

by
Neal Kirschenbaum
National College of Education

Dr. Simoes (1981) noted and gave evidence of the sociopolitical perception of the bilingual child. In this paradigm, he highlighted the negative view of the bilingual, and in particular, the perception that the bilingual is deficient in his abilities as well as in need of compensatory education. Perhaps at this point, further elucidation of the integration of psychology, bilinguaism, and the historical roots of this merger would be appropriate. This paper will focus on the effects of bilingualism on intelligence, an area heavily investigated in psychology and eluded to in Simoes' presentation (1981).

Numerous studies determining the effects of bilingualism on intelligence present contradictory results, ranging along the gamut from detrimental effects to the enhancement of intelligence. The bulk of the earlier studies indicated bilingualism to have negative effects on intelligence. However, the ability to generalize from the results of these studies is seriously curtailed by the lack of adequate experimental controls and methodological practices that bring the validity of their conclusions into question.

An early study in this area was conducted by Saer (1923) employing 1,400 bilingual urban and rural Welch children. The results of a Stanford Binet intelligence test (translation was provided), vocabulary, composition and dextrality tests indicated the intellectual superiority of monolingual children and was interpreted in terms of "mental confusion" experienced by bilinguals. Similarly, Smith (1923), Yoshioka (1929), Wang (1926), Mead (1927) and Rigg (1928) conducted studies with a variety of bilingual groups and indicated the inferior intellectual ability of bilinguals. However, the lack of experimental controls for vital factors as age, sex, and socioeconomic status and subjective scoring procedures (Smith, 1923) and measures of bilingualism (Wang, 1926; Mead, 1927; Rigg, 1928) limit the generalizability of these results. Additionally, many of the tests administered were verbal in nature which could artificially depress the IQ scores of bilinguals who may have a language disadvantage because of limited proficiency in the test language.

Several authors addressed this latter issue by administering nonverbal intelligence tests. A minority of these studies (Jones and Stewart, 1951; Anastasi and Cordova, 1953; Anastasi and De Jesus, 1953) indicated the intellectual superiority of monolinguals over bilinguals. It should be noted that some of the aforementioned methodological concerns are in question in these studies as well as the lack of testing sophistication in the bilingual group. Interestingly, in a later study (Jones, 1959) when the two linguistic groups were equated for socioeconomic status, it was found that there no longer was a significant difference in IQ scores obtained on the nonverbal test, dramatically accenting the importance of controlling this frequently overlooked criteria. The majority of studies (Pinter and Keller, 1922; Barke, 1933; Arsenian, 1937; Seidl, 1937; Darcy, 1946; Johnson, 1953; Carrow, 1957) indicated no statistically significant difference between monolingual and bilingual IQ scores as measured by nonverbal tests. Adding to the confusion, a number of studies (Hill, 1936; Spoerl, 1944; Kolaska, 1954) employing *verbal* intelligence tests found no significant difference between monolingual and bilingual IQ scores, and a few studies (Davis and Hughes, 1927; Stark, 1940; Peal and Lambert, 1962) using verbal as well as nonverbal IQ tests noted a positive correlation.

To better understand the interaction between intelligence and bilingualism, one must critically evaluate the past research. First, the lack of uniformity in defining bilingualism limits the comparison between studies and the ability to generalize. Only two studies employed objective criteria and measures of bilingualism (Johnson, 1953; Peal and Lambert, 1962). Other studies provided either subjective measures such as questionnaires or no control over this variable. For example, Pinter (1932) used the child's surname to determine bilingualism, Pinter and Keller (1922) determined bilingualism based on the parents' nationality, and Davies and Hughes (1927) assumed bilingualism based on the Jewish ethnicity of the subjects.

In determining the bilingualism of subjects, several criteria must be considered. First, O'Doherty (1958) notes the distinction between the balanced and pseudo-bilingual. The pseudo-bilingual is one who is far more proficient in one language than another and does not use his second language as a means of communication. A balanced bilingual is proficient in both languages and has the ability to use either one as a means of communication. Much of the research indicating the detrimental effects of bilingualism on intellectual functioning employed pseudo-bilinguals, a highly questionable practice in view of O'Doherty's (1957, p. 285) claim. "The pseudo-bilingual is the real problem, since very often he fails to master either language, while the bilingual by definition has mastered both." The importance is further supported by Peal and Lambert (1962) who employed balanced bilinguals and in fact showed bilingualism to have a positive effect on cognitive functioning.

A second significant differentiation, which is dependent upon the acquisition method of the two languages, is between compound and coordinate bilinguals. Osgood and Ervin (1965) describe compound bilingualism as acquiring both languages in the same context of learning one through the medium of the other. Thus, the two languages comprise a single system. The coordinate bilingual acquires the two languages in two different contexts. These two languages are acquired at different times, places, or concurrent life situations, and thus,

parallel language structures evolve. These differing linguistic structures in bilingual individuals may affect the nature of their bilingualism.

A third consideration is the age of onset of the second language. This issue is much debated in the research literature with some researchers claiming linguistic superiority of early onset bilingualism (Judd, 1927; Castillejo, 1933; Arsenian, 1937; Penfield, 1953) while others vehemently argue, "It is suggested that bilingualism in young children is a hardship and devoid of apparent advantage, because bilingualism appears to require a certain degree of mental maturation for its successful mastery." (Yoshioka, 1929, p. 476). Although the optimal age of second language learning is disputed, it is agreed that "the time at which the second language is introduced to the young child may be one of the critical dimensions of the ultimate cognitive effects of being bilingual." (John and Horner, 1971, p. 171)

Sociocultural and psychological factors can affect bilingualism and thus should be accounted for in the experimental paradigm. Weinrich (1970) notes several extra-linguistic factors that affect the nature of bilingualism such as size and sociocultural homogeneity of the bilingual group, as well as idiosyncratic and stereotyped attitudes towards the languages, associated culture, and bilingualism itself. Several studies (Christophersen, 1948; Anisfeld, Bogo, and Lambert, 1961; Levine, 1969) point to affective associations with a language which affect bilingualism. Fishman (1968) underscores the necessity of looking at a comprehensive sociolinguistic model in comprehending the functioning of the bilingual.

Thus, the scrutiny of many complex variables is a mandatory step which must precede the evolution of a uniform definition of bilingualism. This definition is mandatory if research is to generate broad psycholinguistic principles regarding the effects of bilingualism on intelligence, yet, to date, this criteria remains unfulfilled.

Additionally, one can be critical of past research due to the types of intelligence tests employed. As previously noted, many early studies (Smith, 1923; Portenier, 1947) used only verbal IQ tests which may artificially depress bilingual IQ scores because of language disadvantage. Translations of verbal IQ tests were also employed (Saer, 1922; Keston and Jiminez, 1954); however, methodologically, this is a highly questionable practice since the translation had not been subjected to standardization procedures. In general, Darcy (1963) criticized the practice of using solely verbal IQ tests citing their lower validity as an intelligence measure.

The lack of controls over variables such as age, sex, number and socioeconomic status was previously noted in much of the research. In addition, several studies employed IQ tests involving time limits which have been shown to penalize the IQ scores of bilinguals (Knapp, 1960; Lewis, 1959).

A final criticism stems from the lack of longitudinal data on the cognitive development of bilinguals. To obtain a comprehensive, holistic view of the bilingual's cognitive development and intellectual functioning, one must evaluate the child over a period of time, rather than at one point in time.

In conclusion, the effects of bilingualism on intelligence have been heavily researched, yet contradictions and ambiguity have resulted. To some degree, this confusion may have unwittingly aided the negative perception of the bilingual child documented by Simoes (1980). Perhaps more rigorous, methodologically sounder research is mandatory to clarify the relationship between IQ and bilingualism and dispell unwarranted perceptions of the bilingual child.

References

Anisfeld, K., Bogo, N., & Lambert, W. E. *Evaluation/Reactions to Accented English Speech.* McGill University, 1961.

Anastasi, A., & Cordova, F. A. "Some of the Effects of Bilingualism upon the Intelligence Test Performance of Puerto Rican Children in New York City." *Journal of Educational Psychology*, --, (1953), 1-19.

Anastasi, A. & De Jesus, C. "Language Development and Nonverbal IQ of Puerto Rican Preschool Children." *Journal of Abnormal and Social Psychology*, 48, (1953), 357-366.

Arsenian, S. *Bilingualism and Mental Development.* New York Teachers College, Columbia University, 1937.

Barke, E. M. "A Study of Comparative Intelligence of the Children in Certain Bilingual and Monoglot Schools in South Wales." *British Journal of Educational Psychology*, 3, (1933), 237-250.

Carrow, M. A. "Linguistic Functioning of Bilingual and Monolingual Children." *Journal of Speech and Hearing Disorder*, 22, (1957), 371-380.

Castillejo, J. "Modern Language in an International School." *The New Era*, 1933.

Christophersen, P. *Bilingualism.* London: Methuen, 1948.

Darcy, N. T. "Bilingualism and the Measurement of Intelligence. Review of a Decade of Research." *Journal of Genetic Psychology*, 103, (1963), 257-282.

Darcy, N. T. "The Effect of Bilingualism upon the Measurement of the Intelligence of Children of Preschool Age." *Journal of Educational Psychology*, 37 (1946), 21-44.

Davies, J. & Hughes, H. G." An Investigation into the Comparative Intelligence and Attainments of Jewish and Non-Jewish School Children." *British Journal of Psychology*, 1B, (1927), 134-146.

Fishman, J. A. "Sociolinguistic Perspectives on the Study of Bilingualism." In J. A. Fishman et al. *Bilingualism in the Barrio*. Washington, D.C.: Department of Health, Education, and Welfare, 196B.

Hill, H. S. "Correlation between IQ's of Bilinguals at Different Ages on Different Intelligence Tests." *School and Society*, 44, (1936), 59-90.

John, V. & Herner, V. *Early Childhood Bilingual Education*. New York: Modern Language Association, 1971.

Johnson, G. B. "Bilingualism as Measured by a Reaction-Time and the Relationship between a Language and a Nonlanguage Intelligence Quotient." *Journal of Genetic Psychology*, B2, (1953), 3-9.

Jones, W. R. *Bilingualism and Intelligence*. Cardiff: University Waks Press, 1959.

Jones, W. R. & Stewart, W. A. "Bilingualism and Verbal Intelligence." *British Journal of Psychology*, 4, (1951), 3-B.

Judd, C. H. *The Psychology of Social Institutions*. New York: Macmillan Co., 1927.

Keston, J. J. & Jiminez, C. A. "A Study of the Performance on English and Spanish Editions of the Stanford-Binet Intelligence Test by Spanish-American Children." *Journal of Genetic Psychology*, B5, (1954), 263-269.

Knapp, R. R. "Effects of Time Limits on the Intelligence Test Performance of Mexican and American Subjects. *Journal of Education Psychology*, 1, (1960), 1, 14-20.

Kolaska, B. J. "The Relationship between Bilingualism and Performance on a Linguistic Intelligence Test." *Dissertation Abstract*, 14, (1954), 23-96.

Levine, H. "Bilingualism: Its Effects on Emotional and Social Development." *Journal of Secondary Education*, 44, 1969, 69-73.

Lewis, D. G. "Bilingualism and Nonverbal Intelligence: A Further Study of Test Results." *British Journal of Educational Psychology*, 29, 1959, 17-22.

Mead, J. "Group Intelligence and Linguistic Disability among Italian Children." *School and Society*, 25, 1927, 465-46B.

O'Doherty, E. F. "Bilingualism: Educational Aspects." *Advancement of Science*, 56, (195B), 2B2-2B6.

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

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

Penfield, W. "A Consideration of the Neurophysiological Mechanism of Speech and Some Educational Consequences." *Proc. American Academy of Arts and Science*, B2, 1953, 20B-209.

Pinter, R. "The Influence of Language Background on Intelligence Tests." *Journal of Social Psychology*, B, (1922), 214-222.

Pinter, R. & Keller, R. "Intelligence of Foreign Children." *Journal of Educational Psychology*, B, (1922), 214-222.

Portenier, I. G. "Abilities and Interests of Japanese-American High School Seniors." *Journal of Social Psychology*, 25, (1947), 53-61.

Rigg, M. "Some Further Data on the Language Handicap." *Journal of Educational Psychology*, 19, (192B), 252-256.

Saer, D. J. "An Inquiry into the Effect of Bilingualism upon the Intelligence of Young Children." *Journal of Experimental Pediatrics*, 6, (1922), 232-240.

Seidl, J. C. G. "The Effect of Bilingualism on the Measurement of Intelligence." Unpublished doctoral dissertation, Fordham University, 1937.

Simoes, A. "Issues of Special Education and Language Assessment for the Limited-English Proficient Student." Paper presented at the meeting of the Bilingual Conference of National College of Education, Evanston, IL. June 19B1.

Smith, F. "Bilingualism and Mental Development." *British Journal of Psychology*, 13, (1923), 270-2B0.

Spoerl, D. T. "The Academic and Verbal Adjustment of College-Age Bilingual Students." *Journal of Genetic Psychology*, 64, 1944, 139-157.

Stark, W. A. "The Effect of Bilingualism on General Intelligence: An Investigation Carried Out in Certain Dublin Primary Schools." *British Journal of Educational Psychology*, 10, 1940, 102-106.

Yoshioka, J. G. "A Study of Bilingualism." *Journal of Genetic Psychology*, 36, 1929, 473-479.

DATA COLLECTION AND LANGUAGE ASSESSMENT POLICY

by
Susan Duron
Illinois State Board of Education

The data collection and language assessment policy in Illinois has been established through a dynamic process. This process involves a variety of individuals whose interests and expertise vary greatly. The resulting policy represents a consciously directed effort toward a uniform data collection procedure which is currently utilized by Illinois' Transitional Bilingual Education Programs serving over 42,000 students.

This paper will be divided into three sections. The first is a historical perspective of the evolution of bilingual program evaluation efforts. Included in this section are current activities and special projects that are provided to Transitional Bilingual Education Programs.

The second section of the paper deals with Illinois' uniform data collection procedures. These procedures provide the framework for the longitudinal studies that began with the first annual *Program Summary and Evaluation Report* (1980). Sections of this report, prepared and published by the Illinois State Board of Education, are included within this paper.

The third section of the paper outlines the information that was obtained as a result of the data collection and language assessment policy in Illinois. Primary evaluation questions were asked of the Illinois data. These questions dealt mainly with issues such as transition and exit rates, program participation information and student identification information. Further analysis of achievement as a measure of program exit are also included. This section closes with information on testing instruments that are used to assess language proficiency, dominance and achievement.

1. **Historical Perspective:**

   a. Bilingual Education in Illinois Prior to the 1976 Mandate

      Prior to 1971 when the Bilingual Section of the Illinois State Board of Education (known then as OSPI) was established, local districts provided programs for non-English speaking or limited-English proficient students at their own expense or through funds received from Title I, ESEA, earmarked for English as a Second Language.

      Bilingual Education in Illinois was implemented, for the most part, in districts which had high concentrations of limited-English proficient students. Local district personnel who were willing to write a proposal describing and documenting legitimate local needs received discretionary state funds. Long narratives were submitted addressing the newly developed six goals of Bilingual Education:

Goal 1: Students in the bilingual-bicultural program will achieve fluency and literacy in two languages.

Goal 2: Students in the bilingual-bicultural program will achieve at a rate commensurate with their own age, ability, and grade level in all school subject areas.

Goal 3: Students in the bilingual-bicultural program will demonstrate growth in self-esteem.

Goal 4: Students in the bilingual-bicultural program will be provided with a coordinated and integrated learning environment through effective coordination with the regular school program.

Goal 5: All teachers and staff members of participating schools will be involved in a comprehensive inservice training program.

Goal 6: Parents and other community members will be involved in the planning, implementation, and evaluation of the bilingual-bicultural program.

These goals were expected to be implemented in order to address the needs of bilingual students who were classified on the basis of language and/or performance. The following classifications which were tied to funding of bilingual programs from 1975-1979 were developed. Districts were reimbursed by the bilingual section for students Levels I-IV who were receiving program services.

Level I: The student does not speak, understand, or write English, but may know a few isolated words or expressions.

Level II: The student understands simple sentences in English, especially if spoken slowly, but does not speak English, except isolated words or expressions.

Level III: The student speaks and understands English with hesitancy and difficulty. With effort and help, the student can carry on a conversation in English, understand at least parts of lessons, and follow simple directions.

Level IV: The student speaks and understands English without apparent difficulty, but displays low achievement indicating some language or cultural interference with learning.

Level V: The student speaks and understands both English and the home language without difficulty and displays normal academic achievement for grade level.

Level VI: The student (of non-English background) either predominantly or exclusively speaks English.

Bilingual programs in local districts were evaluated during the 1974-1977 school years utilizing a type of goal-centered evaluation model known to districts as On-Site Evaluations. Districts were not selected at random for evaluations. They were, however, selected to allow "representation of a cross-section of the programs in terms of language, school level, program model and strong and weak programs as judged by bilingual education program specialists from the state office." (Park, p. 3)

The goals and strategies of the On-Site Evaluation included the following:

- The first was to assist the local bilingual education program by providing immediate feedback to the program staff, especially while the program was in the developmental stage.
- The second goal of the On-Site Evaluation was to provide a vehicle for involving and inservicing a cadre of people interested in bilingual education to enhance their information and expertise (Bilingual Education Service Center, 1976).
- The third goal of the On-Site Evaluation was to assist the State Board of education in identifying overall strengths and weaknesses in bilingual programs across the state in order to influence program and funding priorities.

b. Bilingual Education in Illinois in Recent Years (Post '76 Mandate)

In 1978, the bilingual section abandoned its efforts to conduct internal evaluations of its programs after the Transitional Bilingual Education mandate (Article 14-C of the School Code) was enacted. Educational specialists were employed to monitor local programs and to assure cooperative efforts in achieving compliance with Article 14-C and the Rules and Regulations for Transitional Bilingual Education (1976).

The Illinois State Board of Education in June, 1979, received a final report on an evaluation of the state-funded Transitional Bilingual Education Program which was conducted by L. Miranda and Associates, a third-party evaluator.

Although the findings of the evaluation were limited, the results, nevertheless, pointed toward some specific weaknesses in the program that were in need of remediation:

—The need to establish a management information system to maintain complete, accurate, uniform and comprehensive program data;

—The need to establish a framework to guide planning, operation and evaluation of Transitional Bilingual Education Programs;

—The need to establish a goal-based evaluation model supplemented by special studies to determine context and design variables that affect student performance;

—The need to focus on answering evaluation questions that can provide guidance for program effectiveness and system design.

As a result of the recommendations by the third party evaluators, the Illinois State Board of Education, in June, 1979, required that standardized procedures be developed for the determination of student eligibility and program participation in Transitional Bilingual Education Programs.

In response to this recommendation, advice from various levels was elicited. Meetings were held with decisionmakers at the state and local levels. Program Evaluation and Assessment staff and individuals representing both downstate and Chicago programs were involved.

In August, 1979, two outside contractors were hired to provide technical assistance to the State Board of Education by coordinating and organizing a state-wide inservice workshop on evaluation procedures for Illinois Transitional Bilingual Education Program Coordinators and Directors. In addition, the contractors were expected to address the areas of local district and statewide data collection procedures in an attempt to refine the information procedures and to draft guidelines and appropriate data collection instruments.

On September 21, 1979, the state-wide workshop on bilingual education evaluation and reporting procedures for FY-80 was held. At this meeting the bilingual program directors were presented information on the recently developed uniform state-wide procedures for data collection and program administration.

In December, 1979, the Illinois State Board of Education advertised for the newly designed position of Bilingual Program Evaluator. The specific duties of the evaluator included:

—Develop format procedures for evaluation and reporting;

—Implement and coordinate the state-wide evaluation system;

—Deliver inservice and technical assistance to state and local educational agency personnel regarding bilingual student assessment and bilingual program evaluation;

—Provide assistance to state education agency staff in drafting state plans, grant applications and responses to RFPs with respect to implementation of the evaluation component;

—Prepare an annual Illinois State Board of Education report based on a synthesis of multiple sources of data which will be disseminated to the members of the General Assembly, to local districts, and to concerned individuals.

On January 7, 1980, the Bilingual Program Evaluator was hired and assigned to the Program Evaluation and Assessment Section. Since that time, a number of major evaluation activities have occurred. These activities can be summarized as follows:

1) Development of evaluation reporting forms and procedures as part of a state-wide uniform data collection procedure;

2) Organization and presentation of five regional workshops on bilingual education;

3) Provision of on-site technical assistance to state and local district personnel;

4) Dissemination of information regarding bilingual evaluation to local, state and national agencies;

5) Collection of statewide data to be used in preparation of an annual report.

c. Current Activities and Special Projects

Local districts have received inservice and technical assistance through the following activities and programs which provided an interdependent network of support services.

i. Regional Evaluation Workshops on Bilingual Evaluation

In March and April of 1979, five regional workshops were held. The workshops included an overview of the new state-wide data collection procedures with specific information provided on form completion. Sessions were also presented on writing or revising District Assessment Procedures, English language evaluation and English language proficiency instruments.

ii. Title IV National Origin Desegregation Project

The Title IV National Origin Desegregation project, a grant award operating under the auspices of the bilingual section, has provided services to requesting districts throughout the state during the 79-80 school year. These services have included:

- On-site assistance in developing national origin desegregation plans,
- Inservice training for staff and community as a result of the implementation of a national origin desegregation plan,
- Inservice training for staff providing special education services including counseling and scheduling practices for national origin minority high school students,
- Inservice training for parents of national origin minority children who are receiving special education services,
- Assistance in planning appropriate special education curriculum for national origin minority students,
- Assistance in identifying appropriate placement language tests and assessment instruments for student placement,
- Assistance in planning meaningful programs of instruction and appropriate program models for national origin minority high school students.

iii. Title VII SEA Technical Assistance Project

The federally funded Title VII SEA Technical Assistance project, a grant award operating under the auspices of the bilingual section, has provided technical assistance to school districts in cooperation with the Bilingual Education Service Center and other outside agencies. Among the services provided are:

- Visitation and monitoring of all Title VII projects;
- Sponsorship of a conference on bilingual/multicultural materials;
- Sponsorship of state-wide meetings for bilingual program directors;
- Sponsorship of two regional parent conferences held at East Moline and Elgin and a state-wide parent conference held in Chicago;
- Sponsorship of workshops on crosscultural communication, Asian curriculum adaptation, and proposal writing;
- Organization of a colloquium series on special issues related to the education of limited-English proficient children;
- Organization of meetings for administrative staff on Rules and Regulations and special education;
- Coordination of the State and institutions of higher education efforts on bilingual teacher certification appeals.

A total of 120 bilingual program directors, coordinators, teachers and aides attended the workshops. In addition, persons attended who were responsible for bilingual student assessment at the local level, including principals, superintendents, regular classroom teachers, school secretaries, and special service personnel. Forty-two of the state-funded districts sent staff to the regional workshops.

iv. On-Site Technical Assistance Provided to State and Local Educational Agency Personnel

Information dissemination on the new state-wide data collection procedures has been very effective through direct contact and consultation with state and local agency personnel. Meetings have occurred between the bilingual evaluator and SEA/LEA staff. Fifty-two of the 70 funded districts were visited. In addition, there were numerous meetings and phone conversations with Chicago District #299 program, technical services and Research and Evaluation staff.

Annual student report data from 70 local districts began arriving at the State Board in June. Data clarification and editing were completed after contact was made with source districts. Many site visitations and phone conversations were made in order to enable the necessary linkage to provide accurate data reporting.

The cooperation between local education agency personnel and the bilingual evaluator has been good. The technical assistance sessions have resulted in product-oriented outcomes usually focusing on drafting of district-assessment procedures and/or development of student-assessment strategies that are local-district specific.

The interactions between state education agency personnel and the bilingual program evaluator have been of two types: a) informational overview and b) inservice of bilingual educational specialists who serve in a consultative role to local districts. Packets of information regarding the state-wide evaluation were distributed to bilingual education specialists and to the Illinois Resource Center consultants to aid them in their provision of technical assistance to Illinois local districts.

v. Indochina Refugee Children Assistance Program

The federally funded Indochina Refugee Children Assistance Program has operated under the auspices of the bilingual section. Supplemental educational services provided for eligible Indochinese students by the State Board have included the following:

- Administration of the Indochina Refugee Children Assistance Program and the initial phases of the Transition Program for Refugee Children and Educational Services for Cuban and Haitian Entrant Children Program;
- Coordination and presentation of workshops on topic of Indochinese languages, cultures and materials for teachers, administrators, aides, etc.;
- Participation in planning sessions for the forthcoming state-wide conference "Meeting Education Needs of Refugee Children in Illinois";

- Dissemination of packets of Indochinese material and information to over 60 different school districts and agencies;

- Establishment of comprehensive Indochinese Resource Library with over 300 holdings including bilingual texts, teacher's manuals, cultural materials, maps, cassettes, dictionaries, area handbooks, bibliographies, etc. (Over 100 agencies were contacted during the development of this collection.);

- Act as liaison between the State Board and U.S. Department of Education on issues related to the Indochina Refugee Children Assistance Program;

- Provision of Information and Referral Services, e.g., assisting school districts in identifying available resources to supplement local efforts to provide services to refugee students.

vi. Illinois Resource Center (IRC)

The Illinois Resource Center (IRC) was established in 1972 and until 1975, operated as the Bilingual Education Service Center to assist school personnel in the proper identification and educational assessment of students of limited-English proficiency and the establishment of appropriate instructional programs of remediation, based on the needs of individual school districts. A professional staff of highly trained educators provided direct assistance to teachers through workshops, consultations, and materials dissemination. IRC staff have developed and housed the most comprehensive collection of instructional materials, resource documents, and research-based literature treating the education of limited-English proficient students in the nation.

Throughout the year, several hundred teachers received assistance fom the IRC in a variety of education areas which included:

- A diagnostic/prescriptive interpretation of individual assessment of the language minority student for effective instruction,

- A fostering of oral language and literacy development of the LEP student,

- Classroom management techniques for the multilingual classroom,

- Adaptation of content area instruction for the LEP student,

- Identification and assessment of LEP exceptional children and appropriate instructional techniques of remediation.

2. **Uniform Data Collection** Procedures

The administrative procedures for bilingual education programs have undergone significant modifications over the past years. The following components were developed in establishing a uniform statewide procedure for program administration and data collection.

The components are:

a. PUBLIC SCHOOL BILINGUAL CENSUS
b. PROGRAM APPLICATION
c. STUDENT CUMULATIVE RECORD
d. ANNUAL STUDENT REPORT

Figure 1 shows a graphic representation of the interrelationship between the program administration and data collection procedures. While each component is reported independently, the interdependency among them is firmly established. Because the district assessment procedure is an integral part of each component, it is depicted as such.

a. The Public School Bilingual Census

The Public School Bilingual Census (see Figure 2) has been designed to identify all non-English language background students and to distinguish which of these students are eligible for program participation at the attendance center. Students are recorded, according to language, in one of three columns.

Column A is provided for students with a non-English background who are attending classes at the attendance center. Column B is provided to distinguish those students reported in Column A whose English proficiency level is below average in aural comprehension, speaking, reading *or* writing in English. Column C is provided to distinguish those students reported in Column A whose English proficiency level in aural comprehension, speaking, reading *and* writing is equal to or above average.

In order to conduct the census, each local education agency must have developed a district assessment procedure. The quality of the district assessment procedure reflects the degree to which districts are able to appropriately identify and evaluate students from non-English backgrounds.

b. The Bilingual Program Application

The Application for Transitional Bilingual Education Program (see Figure 3) is designed to provide program descriptions at the attendance

center level. It is the mechanism whereby districts may request state reimbursement for services provided to students identified in Column B of the Bilingual Census. It also provides local districts with the opportunity to request that some eligible students (as identified in Census Category B) be exempted from program participation. This process is known as Waiver of Program Participation. It is checked against the results of the needs assessment that was performed by the district, juxtaposed with the instructional program which would be provided the student in lieu of bilingual education.

Information to be recorded includes a personnel summary and a program summary by language group for each attendance center in the district. Program budget information is listed according to funding source for each expenditure category. The application also contains the district assessment procedure, which is a description of the overall decision-making process through which program entry and exit decisions are made.

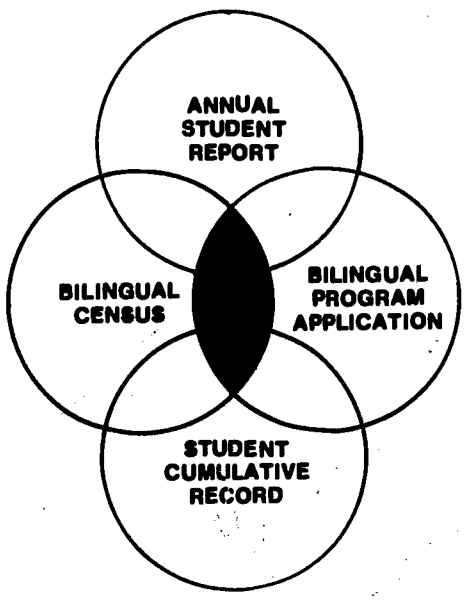
c. Student Cumulative Record

The Student Cumulative Record (SCR) (see Figure 4) is designed to provide local districts with a vehicle for collecting information about students at each attendance center. The Student Cumulative Record remains in the district and serves as the basis for required end-of-the-year reporting. This document has been designed to follow the student during a three-year period. Recorded on this form are data on individual student progress in English, home language, and subject matter skills. Also recorded are data on program entry and exit.

The Student Cumulative Record is a form that districts may modify according to local needs. The Chicago Public School System, for example, collects and reports more information than that contained on the SCR and chooses to utilize a cumulative record idiosyncratic to their district.

FIG. 1

STATEWIDE DATA COLLECTION PROCEDURES



● = **District Assessment Procedure**

d. Annual Student Report

The Annual Student Report (ASR) (see figure 5) is a computer-generated turnaround document which is submitted to the State Board of Education. Local districts may compile student data from their student cumulative records. This report provides the basis for state-wide evaluation activities to assess individual and total program effectiveness. Individual student information is reported on the ASR by student identification number. It includes scores on English language proficiency instruments, attendance, and other programmatic information.

Some large districts, whose computer facilities are easily accessible, have chosen to submit annual student report data on magnetic tape or disc for their large numbers of students. A reporting format is provided those districts along with data edit criteria. Special provisions are made between the Data Control Section of the State Board and districts submitting via tape or disc.

In the spring of 1981, all districts received preprinted Annual Student Report Forms based on student information reported during the previous year. To expedite current-year reporting, identification numbers, language code, entry data years in other programs, and grade level incremented by one year were preprinted. Program staff needed only to add current, updated information, rather than to duplicate the previous efforts.

Concurrent with receipt of the preprinted forms, districts received an aggregate summary of their district's bilingual program information. This turnaround data is expected to be useful in allowing districts to present information to local decision makers requesting an accurate statistical summary.

3. Major Findings

This capsule summary highlights the major findings from the data collected on students enrolled in downstate and Chicago transitional bilingual education programs in Illinois. The majority of Illinois' 143,471 students from non-English backgrounds come from Spanish-speaking backgrounds. Bilingual Census figures indicate that 93,124, or 65% of students, are of Spanish backgrounds. The next largest language background groups represented by Illinois' students are Korean (4,740, or 3.3%); Arabic (2,405, or 1.7%); Vietnamese (1,637, or 1.1%); and Assyrian (1,143, or .B%).

A total of 49,645 students from non-English backgrounds were identified in Illinois as achieving below age or grade level in listening, speaking, reading or writing in English, based on local district assessment procedures. This figure, representing 34.6% of Illinois' students from non-English backgrounds, is reflective of the students eligible for transitional bilingual education program services.

The number of students participating in transitional bilingual education programs in Illinois totalled 41,966. Sixty-nine downstate programs accounted for 6,B73 students, or 16.4% of the total; the Chicago program accounted for 35,093 or B3.6% of the total students participating in programs.

Of those students who were eligible for program services, Chicago and downstate programs served 84.5% from non-English backgrounds. Those not participating in transitional bilingual education programs were not included due to one of the following reasons: denial of parental permission, placement in another program of instruction deemed more appropriate for the child (such as special education), and attendance center location factors. A transitional bilingual education program is mandated only for students from a non-English background who share a common language with 20 or more students in an attendance center.

Students in Illinois' transitional bilingual education programs were identified and assessed for program participation, placement and exit based on 24 different language proficiency and dominance tests, 11 standardized reading tests and 7 standardized achievement tests. Twenty-two percent of all students (9,532) were administered pre- and post tests, while 19,841 or 47.3% were administered at least one language or achievement instrument. A list of instruments recommended for reporting transitional bilingual education student assessment data has been compiled by the State Board of Education and disseminated to local districts. District staff have participated in numerous inservice opportunities related to testing and assessment. It is expected that the number of students assessed will increase in subsequent years as technical assistance, inservice and district assessment procedures improve.

Illinois transitional bilingual education programs reported daily minutes of instruction. State data illustrate that these students spent an average of 33 minutes in English as a second language instruction, 10B minutes in native language instruction, and 150 minutes in English instruction.

Positive achievement gains were reported for students participating in Illinois' transitional bilingual education programs. On English language proficiency and dominance tests, downstate students were found to have raised their proficiency level by one category (on a language proficiency scale of 1-5), a considerable gain in one program year. Scores based on the *Iowa Test of Basic Skills*, a measure of reading achievement, indicate that Chicago's transitional bilingual education students scored better than expected compared with English-speaking cohorts in compensatory programs. Furthermore, Chicago's students were acquiring English comprehension and speaking skills while developing basic reading skills.

A total of 7,236, or 17.2% of the total number of students participating in transitional bilingual programs, exited from the program. These students left programs by successfully transitioning, dropping out of school, voluntary or involuntary program termination, or for other reasons. Exit rate, unlike successful transition rate, reflects the yearly student turnover. The downstate exit rate was found to be 19.7%, while the Chicago exit rate was calculated at 17.0%.

Transitional bilingual education programs in Illinois successfully transitioned 6,11B students. This figure represents a successful transition rate of 14.6%, state-wide. Downstate programs successfully transitioned B.1% of the students participating in programs, while Chicago successfully transitioned 15.9% of their students.

a. Testing Instruments:

Language Proficiency and Dominance

Implicit in the assessment process is a mechanism for decisionmaking regarding student identification and evaluation in the seventy State-approved bilingual programs. While the request for information needed for district assessment procedures did not specifically require utilization of commercially developed testing instruments, all but five districts used at least one commercially developed instrument. Approximately 20% of all districts utilized teacher-made vocabulary, syntax, comprehension or production tests.

The commercially developed tests for assessment of English language skills were divided into three categories: a) language proficiency tests, b) standardized achievement tests, and c) standardized reading tests. A list of test instru-

ments, by test code, was provided to districts and included in the instructions for completing the Annual Student Report.

Of the thirty-three language proficiency tests listed on the instruction sheets, twenty-three were utilized by the seventy districts. The fact that twenty-three different language proficiency tests were used complicated the procedure for determining language proficiency gains because of the variability of the instruments. However, of the twenty-three instruments utilized, six instruments accounted for students in approximately 75% of the districts:

Boehm Test of Basic Concepts (BTBC)
Functional Language Survey (FLS)
Interamerican Tests
James Language Dominance Test (JLDT)
Language Assessment Scales (LAS)
Peabody Picture Vocabulary Test (PPVT)

Because language proficiency tests are not often used by general public school educators, a brief description as found in *A Guide to Assessment Instruments for Limited-English-Speaking Students* (Reynolds & Sisson, 1978) follows:

i. Boehm Test of Basic Concepts:

A diagnostic instrument used to measure the child's mastery of concepts related to space, time, and quantity. There are 25 pictorial items which are arranged in order of increasing difficulty. The examiner makes a brief statement about each item and asks the students to choose the picture which best corresponds. Students respond by marking answers in their test booklets. The test may be individual or group administered and is appropriate for grades K-2. Approximately 20 minutes is required for test administration.

ii. Functional Language Survey:

An individually administered survey of students' ability in comprehension and production of English. It consists of a total of 15 questions in 3 sections: 1) Comprehension, 2) Production/Repetition, and 3) Comprehension/Production. The FLS is appropriate for 1st through 12th grades and takes 7-15 minutes to administer.

iii. Interamerican Tests:

A norm-referenced instrument which assesses vocabulary and reading comprehension skills. There are 3 levels: 80 items on Level I, 110 items on Level II, and 125 items on Level III. For Levels I and II, students respond by marking answers in test booklets. For Level III, students use a separate answer sheet. This test is group administered. Administration time is 20-25 minutes for Levels I and II and 41 minutes for Level III. The grade range is 1-13.

iv. James Language Dominance Test:

A diagnostic instrument used to measure a students' receptive and expressive vocabulary ability. A total of 40 parallel items are included in Spanish and English. Students respond orally. This test is individually administered and administration time is approximately 10 minutes. The test is appropriate for grades K-1.

v. Language Assessment Scales:

A diagnostic instrument containing 100 items designed to assess phonemic production and discrimination, vocabulary production, sentence comprehension, oral production skills, and a student's ability to use language to attain specific goals. Instructions are given orally, and item stimuli are either taped or pictured in the test booklet. Students respond orally or by pointing. The test is individually administered. A language arts supplement containing follow-up learning activities and language games related to each test item is available. The test takes 20 minutes to administer in English and 20 minutes to administer in Spanish and is appropriate for grades K-12.

vi. Peabody Picture Vocabulary Test:

An instrument used to provide an estimate of the subject's understanding of English vocabulary words. The examiner says a word for each of the 150 picture sets (4 pictures in each set) in the series, and the student responds by pointing to the correct picture. The test takes approximately 15 minutes to administer and is appropriate for grades pre-K through adult.

THE ASSESSMENT OF LANGUAGE MINORITY STUDENTS: CURRENT TRENDS IN THE STATE OF ILLINOIS

by
Maria Medina-Seidner
Illinois State Board of Education

Background

Illinois has actively supported bilingual education programs for almost a decade. Beginning with $200,000 appropriated by the General Assembly in 1970-1971 to fund an experimental program in five Chicago schools, the state-funded experimental program (mandated since 1976) presently provides bilingual instruction in over 30 languages to 40,000 children of limited-English proficiency (LEP) in the Chicago public schools and 83 school districts throughout the state. The appropriation level in FY 80 is $16.5 million.

As in most states mandating bilingual education, the Illinois law requires the establishment of a *transitional* bilingual program. Student participation in the program is limited to three years "or until such time as (the student) achieves a level of English language skills which will enable him to perform successfully in classes in which instruction is given only in English, whichever shall first occur" (School Code of Illinois, Article 14-C-3).

Questioning the steady increase in the number of students enrolled annually in transitional bilingual education programs and the corresponding increase in the annual appropriation request, the General Assembly has begun to question the effectiveness of the program. Specifically, legislators want to know:

1. Are students in bilingual education programs learning English?

2. Are students "transitioning" out of the bilingual program into all-English classes?

3. How many students exit the program each year and for what reasons?

These concerns led the General Assembly to request an outside evaluation of the Illinois Transitional Bilingual Education Program. In June, 1979, L. Miranda and Associates presented their findings to the Illinois State Board of Education. The evaluation determined that bilingual programs were meeting their objectives and that between 17 percent and 24 percent of students enrolled in bilingual programs exited the program each year. The evaluation also pointed to the need to establish a management information system to maintain complete, accurate, uniform, and comprehensive program data at the state level.

In response to these recommendations as well as to the increasing demands for program accountability on the part of the General Assembly, the Illinois State Board of Education has undertaken a complete revision of its annual census procedures (including the determination of student eligibility and program participation in Transitional Bilingual Education Programs) and has developed statewide procedures for data collection and program administration.

Student Assessment and Data-Gathering Procedures

During the 1979-80 school year, the Bilingual Education Section, Program Evaluation and Assessment Section, Research and Statistics Section, and Data Management Section of the Illinois State Board of Education have worked cooperatively to develop and implement new procedures for data collection and program administration. These procedures consist of the following components: (a) Public School Bilingual Census, (b) Program Application, (c) Student Cumulative Record, and (d) Annual Program Report. (Sample forms are available from: Illinois State Board of Education, 100 North First Street, Springfield, Illinois 62777)

Public School Bilingual Census. According to the School Code (Art. 14-C-3):

> Each school district shall ascertain not later than the first day of March, under regulations prescribed by the Superintendent's Office, the number of children of limited English-speaking ability within the school district and shall classify them according to the language of which they possess a primary speaking ability and their grade level, age, or achievement level.

The Public School Bilingual Census has been designed to identify all non-English language background students and to distinguish which of these students are eligible for program participation. Census forms are distributed annually to all school districts in January and are due at the Illinois State Board by March 1. Each attendance center must fill out a form. Students are recorded according to language in one of three columns. Column A is provided for students with a non-English background who are attending classes at the attendance center. Column B is provided for students reported in column A whose English proficiency level is below average in aural comprehension, speaking, reading, or writing in English as determined by district per-

Reprinted by permission of the editor from *Georgetown University Roundtable on Languages and Linguistics 1980*, edited by James E. Alatis, Washington D.C.: Georgetown University Press, 1980.

sonnel in accordance with a state-approved district assessment procedure. Column C is provided for students reported in column A whose English proficiency level is equal to the state-approved norms or average.

In order to conduct the census, each LEA must have developed a district assessment procedure. This procedure must contain:

(a) A description of the district procedure for the identification of students with a non-English background specifying the instruments or other assessment strategies used, including the individual(s) responsible for implementing the procedure and the training to be received by the person(s) who will perform the identification. For example, STUDENT IDENTIFICATION must include: PROCEDURE for identification of non-English background students, INSTRUMENT specification, WHO assesses students from non-English backgrounds, TRAINING of the assessor.

(b) A description of the district procedures for evaluating the English proficiency of students whose first or native language is English, specifying the instruments and/or procedures to be used, including the person(s) responsible for the English proficiency evaluation and the time the evaluation will occur. For example, EVALUATION must include: PROCEDURE for evaluation of native English-speaking students' English proficiency, INSTRUMENT specification, WHO assesses English language proficiency of native English speakers, TIME of evaluation.

(c) The average English proficiency, performance, or achievement level by grade or age equivalent for students whose first, or native language is English. For example, PROFICIENCY LEVELS must include: STATEMENT of local average proficiency levels for EACH age or grade level for native English-speaking children.

This procedure is reviewed by the Bilingual Section and Evaluation Section staff at the Illinois State Board of Education. If the district assessment procedure is unsatisfactory, technical assistance is provided. Since this was the first year of this procedure, five regional technical assistance workshops for district evaluation staff were conducted prior to the census. These sessions helped districts not only in the preparation of their assessment procedures, but also in improving their knowledge of assessment instruments.

The new student assessment requirements also reflect the State Board's position as stated in their newly adopted Policy Statement on Bilingual Education:

> The State Board of Education believes that the educational needs of *each* student limited in English proficiency should be met. The Board shall approve the standards by which the district determines the eligibility of its non-English background students for transitional bilingual education programs and services. The required primary criterion for eligibility and successful program completion for each non-English background student shall be an acceptable measure of English proficiency as compared with peers whose first or native language is English.

The *Program Application* is prepared by each school district planning to conduct a bilingual education program the following school year. It is due May 1. The Program Application is designed to provide program descriptions at the attendance center level. The number of students must correspond to the number given on census column B; otherwise, an explanation must be attached. The application provides the LEA with the opportunity to request that some eligible students be exempted from program participation. However, an explanation must be attached including a description of the needs assessment that was performed and the instructional program which will be provided instead of bilingual education. Information is recorded which includes a personnel summary and a program summary by language group for each attendance center in the LEA. The application also contains the district assessment procedure.

The *Student Cumulative Record* is designed to provide local districts with a vehicle for collecting information about students at each attendance center. The Student Cumulative Record remains in the district and serves as the basis for required end-of-the-year reporting. This document has been designed to follow the student over a three-year period even if the student transfers to another school. Provisions are also made for a fourth or follow-up year. Recorded on this form are data on individual student progress in English, home language, and subject-matter skills. Important program information, such as minutes of English as a second language (ESL) instruction per day, minutes of instruction per day using the native language, minutes of English instruction per day (excluding ESL), and total days present during the year, is also recorded. Also recorded are data on program entry, e.g. entry date, years in other programs and exit (e.g. exit date and exit code). The bilingual program exit codes are as follows:

1. *Transitioned:* Student able to perform successfully in an all-English classroom as determined by district exit criteria.

2. *Involuntary program termination:* Student has moved or been promoted within the district to an attendance area which does not require or offer a bilingual program.

3. *Voluntary program termination:* Student has withdrawn from the program at the request of the parents.

4. *Dropped out of school:* Student has dropped out of school, but still resides within the school district.

5. *Other:* Student has either withdrawn from school and moved to another district or his/her status is unknown.

The *Annual Student Report* is a computer-generated turn-around document which is submitted to the State Board of Education in July. LEAs are expected to compile student data which will provide a basis for statewide evaluation activities to assess overall program effectiveness. Individual student information is reported by student ID number and includes scores on English language proficiency, instruments used, attendance and other programmatic information, e.g. the amount of time spent daily on ESL instruction, native language, and English language instruction.

A comprehensive evaluation report based on information gathered from the Bilingual Census, the Program Application, and the Annual Student Report is to be prepared annually and disseminated to members of the General Assembly, local education agencies, institutions of higher education, parents, community representatives and other concerned citizens. The report will include a descriptive section which will emphasize the unique programmatic characteristics of each bilingual program in Illinois and a quantitative section which will address these evaluation questions:

1. What is the total number of identified LEP students by language who were adequately and appropriately served through state-funded transitional bilingual education programs during the preceding school year? In order to demonstrate the extent to which students were served or underserved, one or more of the following will be included: pupil/teacher ratio, full-time equivalency ratios, type of teacher endorsement, program model type (self-contained, pull-out, etc.).
2. What is the total number of identified LEP students by language who were underserved through state-funded Transitional Bilingual Education Programs during the preceding school year?
3. What is the total number of identified LEP students by language who were not served through state-funded Transitional Bilingual Education Programs during the previous school year?
4. What entrance criteria are utilized to determine student participation in the Transitional Bilingual Education Program?
5. How many students left the Transitional Bilingual Education Program during the preceding school year?
6. Of the students leaving the Transitional Bilingual Education Program, what were the reasons?
7. How many students were exited (successfully transitioned) and found able to perform successfully in an all English classroom as determined by district exit criteria?
8. To what extent do students in the Transitional Bilingual Education Program show evidence of progress in English language skills?
9. How many minutes per day of ESL instruction did students in Transitional Bilingual Education Programs receive?

POLITICAL EXPEDIENCE OR EDUCATIONAL RESEARCH?

An Analysis of Baker and deKanter's Review of the Literature of Bilingual Education

by
Stanley S. Seidner, Ph.D.
National College of Education

Introduction

Baker and deKanter's study appeared in a flurry of controversy, as at least one newspaper declared in bold headlines "Studies Disavow U.S. Focus on Bilingual Education" (*Washington Post*, September 29, 1981). The staff writer began the article by stating "Controversial new Department of Education studies" concluded that "the federal government no longer should focus its efforts to aid non-English speaking students primarily in (*sic*) bilingual education because there is little evidence those programs work..." (*Ibid.*, p. A6). Curious to determine if this was yet another case of misrepresentation by the media of sound educational research, I obtained a copy of the study (Baker and deKanter, 1981). After a careful analysis, I sadly conclude that the study reflects a trend of prostituting educational research for possible political and economic expediency.

The problems which plague the potential development and execution of a research design are well-known and too extensive to sustain a worthy recapitulation (See Best, 1977; Stanley, 1967; Thorndike, 1971, 1968; Ebel, 1967; Popham, 1975, Ennis, 1973, among others). The actual development and implementation of a research design according to scientific canon implies the greatest possible reduction of bias and a "fair hearing" of data. Difficult as it becomes to do so with primary data, the application of this type of approach is more so in approaching data from a secondary perspective. A hidden agenda of a political nature further compromises this approach to the point where the attempt at rigorous research metamorphoses into statements of support for public policy recommendations. Examining with care the authors' methodology, I am convinced that Baker and deKanter's review is an example of this process. This paper presents an analysis of their review, with particular attention given to the authors' methodology and conclusions. We begin with an overview of their research design.

General Research Approach

Baker and deKanter pose two research questions which they claim are "derived from the principal intent of Federal policy..." (Introduction, p. 1)."*

1. Does transitional bilingual education lead to better performance in English?
2. Does transitional bilingual education lead to better performance in nonlanguage subject areas?

The authors state that they intended to limit the discussion of their review to the two questions (ch. 1, p. 8). They contend to have reviewed "more than 300 documents," of which 150 were program evaluations. Most of the studies were Title VII evaluations, supposedly representing every region in the country. They also claim that studies were identified by means of an "ERIC search, by consultation with experts in the field, from prior reviews, and from lists of studies." Although they insist on having "covered the major studies," Baker and deKanter acknowledge that they were unable to obtain copies of pre-1978 evaluations from the Office of Bilingual Education. Nevertheless, the authors express the belief that their review is the most comprehensive undertaken to date. Only 28 studies were judged by the authors to apply to their concerns and methodological criteria. In essence, their conclusions and recommendations for national policy are based upon these 28 selected studies. It is unfortunate that the authors failed to fully develop the application of their criteria by briefly citing their reasons for rejecting each of their reportedly examined 300 or so studies. Although they claim limitations of time, it becomes difficult to determine the quality of attention Baker and deKanter allegedly expended in their total review. Their criteria for rejection included:

i. Rejected studies lacked random assignment to treatment and comparison groups, as well as

This paper was written as an independent analysis of a report by Keith A. Baker and Adriana A. deKanter, who are employed by the Office of Planning, Budget and Evaluation, U.S. Department of Education.

The author would like to extend his appreciation and gratitude to Dr. Robert L. Thorndike, Professor Emeritus, Teachers College, Columbia University; Dr. Meyer Dwass, Director of Statistics and Evaluation, Northwestern University; Dr. Calvin Claus, Chair, Psychology, National College of Education for their consultation, suggestions and encouragement in the preparation of this paper.

*The citations of Baker and deKanter are consistent with their confusing pagination.

control for potential initial differences between groups.

ii. Studies failed to address the issues which Baker and deKanter claimed were under consideration.

iii. The studies employed a norm-referenced design.

iv. Studies used comparison of post-test scores only with nonrandom assignment.

v. There was a reliance upon school-year gains for the program group without a control group.

vi. There was a reliance upon grade-equivalent scores.

A number of criteria (for example, vi) can be debated with arguments for and against their inclusion. On the surface, it would seem as if the authors are adhering to rigorous research methodologies. However, a closer look shows otherwise. An analysis of Baker and deKanter's paper has elicited the following concerns:

1. The first concern is the authors' improper definitions of terms. Through their glossary of selected terms and some explanations in the text, Baker and deKanter attempt to set the frame of reference which would aid them in approaching the research questions. Serious problems develop, as we will presently see, with a number of terms. The authors present their definitions of terms such as Transitional Bilingual Education (TBE), English as a Second Language (ESL), Structured Immersion, and Submersion (Baker and deKanter, Introduction, pp. 2-3; ch. 1, pp. 5-8). In their methodology of creating operational definitions, the authors would have to assume that their terms are equally applicable to the selected studies. It would appear that Baker and deKanter failed to check the validity of their definitions with the reality of practices in programs which they selected or to correct for any potential inconsistencies. (Their definitions fail to account for such differences as time and treatment, for example.) The authors' definitions of TBE, for instance, informs us that subject matter "is at least partially taught" in L^1 until a successful transition can be made to L^2 (ch. 1, p. 6). Unfortunately, the vagueness attached to their interpretation of the essential nature of this definition is useless for rigorous regard and consistent measurement approaches. Balasubramonian, et. al. (1973), may have addressed a number of schools with models corresponding more readily to maintenance bilingual/bicultural education. (Illinois implemented their mandate for TBE in 1976). Baker and deKanter fail to reconcile potential differences in their definition of such variables as time and treatment from this study and others, like Carsrud and Curtis (1980), Cohen (1975), Covey (1973), Kaufman (1968), or Ramos, et al. (1967) among others.

A similar problem arises with Baker and deKanter's definition of immersion. The authors overlook that the St. Lambert project (Lambert and Tucker, 1972) was designed to promote functional bilingualism through the vehicle of a home-school language switch (also Bruck, Lambert and Tucker, 1974). It would also appear that the authors of the Canadian studies, as well as others associated with the program, have defined immersion in terms of bilingual education, with the goal of promoting dual language facility (for example, Swain, 1972; Tucker, Lambert and D'Anglejan, 1973; Lambert, 1974; Cohen and Swain, 1976; Swain and Barik, 1973). Here again, faulty research design is apparent in Baker and deKanter's reconcilation of their definition of "structured immersion" and their selection of studies (ch. 1, p. 6, ch. 2, pp. 62- 72). Again, there is no evidence that the authors considered differences in approaches to the implementations of programs by Lambert and Tucker (1972), Barik and Swain (1975), Barik, Swain and Nwaninobe (1977), and Pena-Hughes and Solis (1980). If the authors desired to propose an operational definition, they were unsuccessful in convincingly describing procedures.

Another instance of potential problems arises with the authors' definition of ESL. There is no evidence to indicate, here again, that Baker and deKanter took under consideration similar concerns. Had they considered the distinctions in studies between students learning English as a second language from the availability of ESL methodologies? Which ESL methodologies, if any, are under consideration (i.e. behavioral, cognitive, eclectic)? Again, we are given no indication. In short summarization, Baker and deKanter's approach to their definitions is less than acceptable as a durable component of their research design. Young and Comtois relate the following (1979):

> Research utilization is an extraordinarily complex phenomenon. Yet it can be analyzed like any other social phenomenon through a process of conceptualizing variables, defining them, and operationalizing them in a research setting. And as with other phenomena, clear definitions are necessary to make possible communication on the topic. Without such clarity, one cannot make sense out of the otherwise conflicting assessments of such things as the extent of program evaluation use and what factors seem associated with that use (p. 64).

The authors' definitions could not fail to influence, in turn, the extent of their own research (contingent upon what Caplan, 1977 believed to be the conceptualizations of use and research; also Babbie, 1979, and Weiss, 1977, 1978, Thorndike and Hagen, 1969). It would then follow that the remainder of Baker and deKanter's methodology might also be subject to scrutiny.

2. A second concern is the authors' attempt to play the role of rigorous researchers. Baker and deKanter apparently feel in a general sense that technically flawed studies are worse than no studies at all. There is a general lack of consensus regarding evaluation methodology. Some evaluators adhere to principles of rigorous research, while others contend that such processes exclude, in their rigidity, approaches to gathering other information (See Young and Comtois, 1979). The application, in general, of a rigorous research approach to a secondary study becomes problematic. This approach to a review of literature subordinates, in premise, the methodologies and findings of selected and rejected studies to strict adherence of scientific canon. The principle in itself would appear somewhat palatable, if Baker and deKanter had taken the time and care to equitably apply their criteria. Their approach would seem convenient in eliminating a number of studies which would potentially refute their arguments. Consider the statement made by the primary researcher of a rejected study that Title VII Basic Programs "are primarily geared to equalize educational opportunities." Accordingly, such programs "are not designed to, nor controlled by the strictest canons of scientific research" (Leyba, 197B, p. 6). The resulting few studies which survive the authors' gauntlet become too paltry in number and diverse in methodology and research intent to even afford serious conservation of Baker and deKanter's research questions. One of the basic tenets of educational research was expounded by Thorndike (1973):

> We must always remember that any test, or any other type of behavior observation represents only a limited sample from some domain or behavior. It represents the domain imperfectly, and the score it produces is only an approximation to the score that the individual would get for the whole domain. . .(P.5B)

The observations of individual researchers represent limited domains imperfectly and should be viewed within the stated designs of the particular studies. The face-value argument of the authors might run along the lines that policy making dictates current answers based upon available data. The role of rigorous researcher, as it is selectively applied by Baker and deKanter, is ludicrous by virtue of their failure to recognize differences of research purposes and methodology among studies. (We will see evidence of this in the following subsections.)

3. An outgrowth of the preceding subsection is a concern with the authors' methodology in attempting to reconcile the selection of data to their research questions. Baker and deKanter select Kaufman's study (196B), for example, to test the effectiveness of TBE presumably against submersion. The purpose of Kaufman's study is generally conveyed by the title, "Will Instruction in Reading Spanish Affect Ability in Reading English?" There is no evidence to suggest that the study was designed to test the general effectiveness of TBE in comparison to the authors' definition of submersion or to specifically answer Baker and deKanter's research questions. An analysis of their secondary review, however, shows attempts by the authors to "bend" the data to fit their research questions. At best, the collection of studies indicates a review of literature which touches upon their research questions by selective interpretation and inference — not by substantial research which has been collected by original and careful design. For example, Baker and deKanter cite one group in Kaufman's study (group A), which received more instruction in Spanish than a comparison group, as having "showed no English reading improvement" (p. 18. By their own admission, more time was alloted to primary speakers of Spanish in their native language than in English.) It would seem that Baker and deKanter classify Kaufman's study within their definition of TBE because of the use of Spanish. Why would the authors prefer this categorization, instead of perhaps immersion? One would appear as arbitrary as the other. A less generous reviewer might categorize Baker and deKanter's attempts as "charlatanic," rather than reflecting misinformation.

It is also intriguing that Baker and deKanter praise as a strength of Kaufman's study "the use of covariance to control for pre-existing differences" in verbal and non-verbal IQ, among other factors (p.1B). Discerning the authors' rationale is further confusing, given their citation on I.Q. testing and "innate language ability" (ch. 1, p. 16, f. 9; also see their definition of "correlation", p. 2). The authors are too preoccupied with statistical tests to consider any valid relationship of intelligence testing to Kaufman's primary intentions or to their own research questions. (See my concerns below on statistical misuse.) If the authors contend there is a valid relationship to either, they fail to provide them. If there isn't, why would Baker and deKanter mention the utilization of covariance as a strength of the study within an irrelevant framework? The Balasubramonian, Seelye and DeWeffer study (1973) is another example of misapplied research. The original intent of the three authors is reflected in the title of their study, "Do Bilingual Education Programs Inhibit English Language Achievement? A Report on an Illinois Experiment." Here again, a distinction is made between the purposes of Balasubramonian, Seelye and DeWeffer's study and Baker and deKanter's loose interpretation. The target programs were bilingual/bicultural, corresponding in many instances to a maintenance philosophy. Transitional Bilingual Education as a State mandate in Illinois, was not implemented until 1976. Baker and deKanter claim, as a shortcoming of the study, a failure to present data "on progress in nonlanguage subjects." This presentation of data, however, was not the intention of Balasubramonian, Seelye and DeWeffer since their study focused on a language concern. One could question the motives of Baker and deKanter in setting up a possible "straw dog" to knock down. Of greater concern, however, is the careless methodology employed by the authors in presenting presumably important research issues.

Baker and deKanter continue this trend in their selection process, posing serious questions as to the soundness of their methodology (for example, McConnell, 1980a, 1980b; Ramos, Aguilar, and Sabayan, 1967).[1] A more detailed analysis of Baker and deKanter's selection processes may provide an inquisitive graduate student with a research topic for a Master's thesis.

4. One of my strongest concerns is Baker and deKanter's apparent selective application of their criteria. It would appear that Baker and deKanter are prophetic in their observation that an argument might be made that they "applied arbitrary criteria and personal judgment" (ch. 1, p. 16). The authors at first state what appears on the surface to be rigorous criteria for selection, and then mention in almost the same breath that, "the criteria for methodological soundness were applied in a way that recognized that a design weakness in one area can be compensated by a strength in another area" (Introduction, p. 4). I fail to see how the authors apply scientific canon in their methodology of equating one criterion as equal in weight to another.[2] Take, for example, the application of their criterion of grade-equivalent scores. In essence, the grade equivalent of a particular score could be considered the grade level of the norm group, its median the same as the raw score. (For a discussion, see Beggs and Lewis, 1975.) Baker and deKanter state that the use of grade-equivalent scores was a criterion for rejecting studies (ch. 1, p. 15). The issue is not one of why this criterion was selected (although arguments can be made for and against), but how it was applied. Rejected studies, conducted by Leyba (1978), Trevino (1970) and Olesini (1971), employed grade-equivalent scores. Yet Baker and deKanter waive the application of this criterion for Kaufman's study (1968) and seem to feel that its questionable experimental design sufficed to overcome the shortcoming. This selective application of their criteria appears to be a flaw in Baker and deKanter's methodology and can be cited by other examples.

   The authors supposedly also rejected studies such as Cohen (1974) and South San Francisco, which examined gains without any control groups. The selective application of criteria becomes evident with Baker and deKanter's acceptance of Stern's study (1975) which also failed to include any control group. Here the authors seem to feel that Stern's employment of "some longitudinal analysis and adjustment for the effects of nonrandom selection by analysis of covariance" is sufficient compensation (Baker and deKanter, ch. 2, p. 5). Their arguments are unconvincing, at least to this researcher. The trend continues with the criterion of rejecting studies which utilized norm-referenced designs. Baker and deKanter argue that the nature of the learning curve for limited-English proficient students is unknown, and that a potential phenomenon with achievement tests yields higher gains than are actual true measures of performance (ch. 1 p. 13). We find that a number of rejected studies, particularly St. John Valley (1980), Corpus Christi (1980a, 1980b.) and Arce (1979) employed a norm-referenced design. Nonetheless, the authors waive this criterion for Stern (1975), Ames and Bicks (1978). This kind of selective application of criteria can only serve to destroy the credibility of Baker and deKanter's methodology in their professed scientific review of the literature.

5. Another of my strong concerns in observing the authors' methodology is the way Baker and deKanter apply a selective secondary analysis of data to substantiate their arguments. Along these lines of thought, the authors: a) present incomplete and selective data (errors of omission) and b) present a biased interpretation of studies which do appear. In their review of Covey's study (1973), for example, the authors fail to acknowledge his conclusion that Mexican-American students who were enrolled in a bilingual education program achieved significantly higher in the academic disciplines of English and in reading than those enrolled in regular school programs. (Covey also attributed a favorable attitude toward self and others to those Mexican-American students enrolled in the bilingual program.) Baker and deKanter seem to go out of their way to accentuate problems in interpreting their selective results of the study (see their discussion, ch. 2, p. 5). If the study passed their criteria for selection, why would they not want to present the data as is? They offer "three competing explanations as to why the program worked," which are theoretical and do not detract from the actual results of the study. They also seemed to take the time to have contacted Covey and conclude from the communication that a less than sufficient number in their estimation "achieved a sufficiently higher level of English proficiency to be mainstreamed (p 5.)" While this communication is utilized to clarify their point, why hadn't the authors exercised foresight to clarify their own expressed puzzlement that Covey's study "is very uninformative as to the nature of the program" (see ch. 2, p.4)? Furthermore, how could Baker and deKanter presume to classify the study in support of one of their theses without knowing the nature of the program? A similar pattern emerges with the authors' analysis of Skoczylas' study (1972). Again, Baker and deKanter declare that

[1] It is interesting to note Baker and DeKanter's reliance upon Engle's review (1975) of Ramos, et al., (1967). This reliance strongly suggests that they failed to conduct a first-hand review of the study by Ramos, et al. (1967). They may have also attempted to consult Davis, (1967). Judgements by Baker and deKanter are therefore based upon secondary commentaries. One should also note that Ramos, et al. (1967) do not cite TBE in their study, according to the concept advised by Baker and deKanter in the United States.

[2] The amount of time devoted to students' primary language varied greatly from selected study to study in numerous instances, failing to warrant Baker and deKanter's generalizability regarding the effectiveness of TBE (for example, McSpadden (1979) at 37%; Balasubramonian et al. (1973) at 25%. Kaufman (1968) at 2 1/4 - 3 hours weekly; and Zirkel (1972) at 60%, among others).

bilingual instruction "did not lead to lesser English performance (p. 31). They overlook Skoczylas' conclusion that students who were instructed bilingually were learning Spanish and English simultaneously without apparent difficulty and gained a beneficial "transfer effect" from language to language (1972). Instead of continuing with their own analysis, Baker and deKanter offer a lengthy extract from an evaluation of Skoczylas' study conducted by the National Institute of Education. If the authors accept conclusions regarding insufficient data to determine continued significant differences between control and experimental groups, they are not justified in utilizing this data to justify their arguments.

The subjective interpretation of the authors seems to be emphasized by their treatments of Danoff, Coles, McLaughlin, and Reynolds' study (1977-7B; referred to as the AIR study). Baker and deKanter devote almost the entire part of their discussion of the study in defending it (ch. 2, pp. 57-60) against critical reviews by Cardenas (1977) Gray (1977), and O'Malley (1978) instead of presenting an objective overview and analysis. The authors fail in one representative instance to substantially dispel the argument made by Gray (1977) that the AIR study failed to recognize differences among programs and approached bilingual education as undifferentiated and uniform. Baker and deKanter drop the role of researchers and answer that the criticism does not "acknowledge the needs of policy makers to make informed decisions based on representative data" (ch. 2, p. 58). Another implication of the authors is that the data is representative. Still, they fail to pay heed to their own acknowledgement that the subsample of Danoff's "reanalysis" (1978) was not totally representative of the original sample. Also, the lack of comparative classrooms for the grade III cohort of Title VII children limits the generalizability of results (as well as posing a dilemma for Baker and deKanter in their application of criteria; Danoff, 1978). Moreover, by rejecting studies critical of AIR (such as those cited above) and employing authors in support of the study (Rossi, 1979 and National Institute of Education, 1979), the authors seem to have a political agenda. The reader may find it more than coincidental that Baker and deKanter employed a similar review by the National Institute of Education (cited above) in criticism of Skoczylas' study. Baker and deKanter neglect to mention that a number of individuals involved in the AIR study withdrew from the project and later wrote critical reviews (Arias, Delgado, DePorcel, and Irizarry 1977, also Gray, 1978).

In still another instance, the authors categorically conclude in Cottrell's study (1971) that "the program effect was probably underestimated by the analysis of covariance," since "students from the comparison schools were historically known to outperform students from the project schools..." (Baker and deKanter, 1981, ch. 2, p. 44). First, there is no statistical evidence offered to support their contention. Second, the authors appear to selectively diverge from their absolute interpretation of covariance as a redeeming factor and which they apply in other instances (see subsection on statistical applications below). Another illustration of Baker and deKanter's biased methodology occurs with their examination of Cohen's study (1975). The authors conclude that they found "the programs' effect on English development to have been neutral to a little negative with mixed results in arithmetic" (Baker and deKanter, 1981, ch. 2 p. 52). This is highly questionable, considering that bilingual students, for example, showed greater gains than did the comparison groups regarding communicative ability in English. The bilingual group began, "substantially below the comparison group, so that their gains brought them even with the comparison students" (Cohen 1975, p. 161). The authors also neglect to mention that the follow-up Level I group seemed overall to be roughly equivalent to the comparison group in their development of language skills in English. Here and in other instances, Baker and deKanter failed to present a balanced presentation and to reconcile the data in support or rejection of Cohen's original hypotheses. In terms of academic aptitude, bilingual students scored better than comparison students on one level, and equally on two other levels. In mathematics, bilingual students performed as well as students in the comparison group and better on at least one level. It is interesting that the authors failed to review these findings among others. These examples are but a few which illustrate Baker and deKanter's selective analyses.

6. Another concern is the authors' sense and selective application of statistics. Baker and deKanter, for instance, acknowledge the validity of arguments which criticize analysis of covariance in failing to overcome difficulties posed by nonrandom selection (ch. 1, p. 10; also see for a discussion in the use of the process, Elashoff, 1969, Elashoff and Snow, 1970; Winer, 1971; and Williams, 1979). In studies potentially supportive of the authors' arguments, Baker and deKanter stress analysis of covariance as a strength without further discussion. Note their treatment of the Moore and Paar study (1978). Baker and deKanter cite the use of analysis of covariance as a strength and conclude that the study, therefore, had "better statistical control than many studies having nonrandom assignment" (ch. 2, p. 38). However, Moore and Paar (1978) noted the weakness in their use of covariance to adjust for pretest differences (whereby, the process systematically underadjusted for initial differences between groups). The warning then, was that the resulting scores of nonbilingual classes should be interpreted with caution. In contrast, Baker and deKanter approached one study which tended on face value to discredit their arguments, with a less than balanced treatment of the researchers' application of the statistical process (see their treatment of Cottrell, 1971).

Another curious example is the authors' treatment of the Skutnabb-Kangas and Toukomaa study (1976). Baker and deKanter classify this research

within their chapter of rejected studies. Nevertheless, they take pains to apply a statistical analysis, supposedly to examine the extent to which the data support the use of Level 1 instruction with language minority children (Baker and deKanter, 1981, ch. 3, p. 9). At first glance, the application may appear to an experienced statistician to be a "cook-book" analysis, orignating out of some program such as SPSS (Statistical Package for the Social Sciences). The very application of statistical tests by the authors is questionable, given the potential incompleteness and uncertainty of available raw data. Therefore, secondary presentation of data in the form of a research report portends possible bias in the statistical approaches of the reviewer. In reference to Baker and deKanter's statistical applications, I am at a loss to see some of the significance and interpretation of their approaches. It is unclear as to how (and why), for example, the authors applied both Spearman's rho (r_s) and Pearson's (r). Spearman's rank-order correlation coefficient (non-parametric) is applicable to ordinal-level variables, while Pearson's product-moment correlation coefficient is oriented toward pairs of interval-level variables (for a discussion, see Siegel, 1956, Guilford and Fruchter, 1973). Baker and deKanter provide little or no indication as to what was correlated and why, let alone the appropriateness of their application and interpretation.

In interpreting their data, the authors appear to selectively diminish the values of their X^2, Somer's d, and gamma formulations. Instead of presenting a balanced view of the data, they diminish the importance of these tests, which produce results possibly contesting their arguments. Why, then, would Baker and deKanter want to employ potentially inferior statistical tests in their analysis of the Skutnabb-Kangas and Toukomaa study (1976)?

7. My final concern is with Baker and deKanter's apparent political agenda in the guise of rigorous educational research. Perhaps the authors are unaware of their own dilemma in playing the simultaneous roles of policymakers and rigorous researchers. The dangers of such a synthesis have been voiced by Patton (1978) in his statement:

> The traditional academic values of many social scientists lead them to want to be nonpolitical in their research. Yet, they always want to affect government decisions. The evidence is that they cannot have it both ways (p. 46; also Atkins and House, 1981).

Very powerful evidence suggests that Baker and deKanter indeed provided a political agenda, apparently aimed at discrediting the bilingual education movement in the United States. The premise is simple and logical. If you can discredit the research-base of a particular discipline, you then destroy the substatiation for its *raison d'etre*. What is not known is the level of independent activity or pressured compliance under which Baker and deKanter labored. Let us start with the title page, where Baker and deKanter claim their review "does not represent the official position of the U.S. Department of Education." Yet, on page 1 of their introduction, they state that "the investigation was begun at the request of the White House Regulatory Analysis and Review group for an assessment of the effectiveness of bilingual education." Both Baker and deKanter work in the Office of Planning, Budget and Evaluation. The Deputy Undersecretary of Education for Planning and the Budget, Gary Jones, appeared to injudiciously "leak" information from the report (Baker and deKanter's) to the public media. A former member of the Fairfax County School Board, Jones, was quoted by a leading newspaper as describing ESL as preferable to bilingual education in the schools (Fairfax County). "We were able to prove," he reportedly stated, "our approach was every bit as effective as theirs" (*The Washington Post*, Sept. 29, 1981, p. 16). It should be remembered that, during the past year, President Reagan cited the Fairfax County program as a model, while criticizing bilingual education (*NABE*, 1981). These developments challenge the credibility of Baker and deKanter's attempt to portray their assessment as unbiased or independent of political influence. Additional evidence from an analysis of their review lends greater credence to this assessment. Baker and deKanter seem unable to manipulate their own methodology and must place a flawed Fairfax County Public Schools study conducted in 1980 in the category of rejected studies (ch. 3, pp. 28-30). Yet, Baker and deKanter make sure that the reader notices that the Fairfax County program is "very successful when looked at in terms of gain over the school year (p. 30). Further analysis brings us to the logical second tier of a political agenda.

If one destroys the substantiation for a movement, one has to provide another research base in support of arguments. My analysis suggests that Baker and deKanter have "dusted off" the aforementioned AIR study in an attempted resurrection. At least one self-proclaimed "neo-conservative" researcher (Miller, 1981) used the AIR study to substantiate the evils of maintenance bilingual education (p. 88; which conflicts with Baker and deKanter's attempts to classify it under TBE). Why, for example, would they take the time to place Troike's review (1978) within their category of rejected studies? Troike's work is not an empirical study, such as the ones Baker and deKanter claim to analyze, but a critical review of research. However, Baker and deKanter overlooked this consideration, devoting time to question Troike's analysis while lauding the methodology employed by the American Institutes for Research in their other reports (here Baker and deKanter conclude, "we have more confidence in the validity of the AIR study prepared for the JDRP" (ch. 3, p. 24). As mentioned above, the authors approached their analysis of the 1977-78 AIR study with an examination and refutation of criticisms. This attempted rehabilitation was reinforced by rejecting such studies as Leyba's (1978), which were critical of AIR (Also see Nickel, 1979). The political agenda is reflected as well in additional comments which are cited above in preceding subsections.

Some Observations on Conclusions

Baker and deKanter's faulty methodology, beginning with ill-conceived definitions and criteria, inevitably results in the lack of reconciliation between their conclusions and stated research questions. The best that the authors can offer are qualified general observations which fail to address their originally postulated research questions. In actuality, the two research questions raised by Baker and deKanter have remained unanswered. Again, this is not a surprising outcome, in light of the authors' methodology. They list four categories of conclusions which are excerpted as follows, in the form of the authors' subtitles:

1. Special programs can improve achievement in language minority students.
2. The Federal Government should not place exclusive reliance on transitional bilingual education.
3. Federal policy should be flexible.
4. Improved bilingual research and program evaluations are needed (Baker deKanter, 1981, ch. 4, pp. 1-6).

Under category 1, studies are cited to indicate "that special programs designed to overcome language difficulties in schools can improve the achievement of language minority students." The authors add that "this conclusion says nothing about the effects of any particular instructional approach" (p. 4). The statement is consistent with results of individual studies reviewed by the authors in their selective processes (McConnel, 1980a, 1980b; Zirkel, 1972; Covey, 1973; Plante, 1976; Lambert and Tucker, 1972; Malherbe, 1946; and Cohen, 1975, among others). It would then follow that there is insufficient evidence to warrant the allegation of ineffectiveness of transitional bilingual education, which poses a dilemma for the authors in their second category (see 2, above). A logical outcome would be that the federal government should not place exclusive reliance on any program. The original so-called *Lau* guidelines (which still appear to be in effect) delineate programs as options, according to particular conditions. Title VII legislation addresses another set of conditions for funding. Baker and deKanter are unconvincing in their arguments and analysis that other options are suitable substitutes for the framework of TBE, along the lines of the experimental design they themselves purport to have followed. They are less so in stating that their analysis has "implications beyond the Federal level," because a number of states have followed the lead in developing TBE programs for language minority students. The authors also conclude within this category that their selected findings "do not add up to a very impressive case for the effectiveness of transitional bilingual education" (Baker and deKanter, 1981, ch. 4, p. 5). However, I have found Baker and deKanter's analysis to be one of the poorest and most biased research approaches to a review of literature for the stated observations in preceding subsections.

Category 3 of the authors' conclusions assumes inflexibility on the part of the federal government regarding an emphasis upon "transitional bilingual education to the virtual exclusion of alternative methods of instruction." I fail to see how this premise is substantiated by the authors' approach, or, for that matter, reconciled to their research questions. No criteria have been presented and analyzed relevant to a definition of the methods of measuring flexibility or inflexibility. Since their study does not contend to be an analysis of federal policy, but supposedly an empirical approach, Baker and deKanter's conclusions are highly unsubstantiated and irrelevant. A similar statement by the authors that TBE has been found "ineffective and harmful in other places" also fails to find justification. (Their analysis brings to mind the reasoning of studies over a half century ago which attempted to link bilingualism with retardation.)

The authors' "key to successful teaching in the second language," namely that subject content should not outdistance language, would appear at first logical. However, since the rate and quality of individual language acquisition is variable, the learning-rate of subject content area would be restricted to that particular medium of communication. The authors fail to contend with the reality that second language acquisition will not occur through osmosis and that a necessary time element, which varies from individual to individual, will impede the rate of subject content acquisition in quantity and quality. To coin Baker and deKanter's term, "commonsense observation" tells us that it is pedagogically sound to teach students subject matter in the known language where the symbols convey information in an understandable fashion. It therefore would appear that Baker and deKanter have promoted their political expedient, with little concern as to what is educationally proper. I have no argument against the premise of experimenting in as many promising areas as possible. However, Baker and deKanter's category 4 (see above) goes against their inference that TBE has been tried and proven to be one way or another. Their call for "improved bilingual research and program evaluations" indicates the need for further investments of Federal dollars in this area before any kind of empirical-based conclusions can be drawn. I will agree with Baker and deKanter's observation that bilingual education "involves many complex, difficult issues that have been little or insufficiently studied." It would then follow that the authors' suggested federal policy, recommending the elimination of TBE, is premature and unwarranted.

Bibliography

Ames, J. S. and Bicks, P. *An Evaluation of Title VII Bilingual/Bicultural Program, 1977-1978 School Year: Final Report. Community School District 22.* New York: Community School District 22, 1978.

Arce, A. *Project Rainbow: Pajaro Valley Unified School District.* Wassonville: Arce and Associates, Inc., 1979.

Arias, B.; Delgado, T.; De Porcel, A.; Irizarry, R. *Response to AIR Study: Evaluation of the Impact of ESEA Title VII Spanish/English Bilingual Education Program.* MSS., 1977.

Atkins, J. M. and House, E. R. "The Federal Role in Curriculum Development, 1950-1980." *Educational Evaluation and Policy Analysis,* no. 5 (3), September - October, 1981, pp. 5-36.

Babbe, Earl. *The Practice of Social Research.* Belmont: Wadsworth, 1979.

Babcock, C. "Studies Disavow U. S. Focus on Bilingual Education." *The Washington Post,* September 29, 1981, p. A6.

Baker, K. and deKanter, A. *Effectiveness of Bilingual Education: A Review of the Literature - Final Draft Report,* September 25, 1981. Unpublished, MSS.

Balasubramonian, K.; Seelye, H. N. and DeWeffer, R. E. "Do Bilingual Education Programs Inhibit English Language Achievement? A Report on an Illinois Experiment," Paper presented at TESOL convention, San Juan, Puerto Rico: May 9-12, 1973.

Barik, H. and Swain, M. "Three Years Evaluation of a Large-Scale Early Grade French Immersion Program: The Ottawa Study," *Language Learning,* no. 15 (1), 1975, pp. 1-30.

_____, and Nwanunobe, E. A. "English - French Bilingual Education: The Elgin Study through Grade Five," *The Canadian Modern Language Review,* no. 33 (4), 1977, pp. 459-475.

Beggs, D. and Lewis, E. L. *Measurement and Evaluation in the Schools.* Boston: Houghton - Mifflin, 1975.

Best, J. W. *Research in Education.* Englewood CA: Prentice - Hall, 1977.

Bruck, M., Lambert, W. E. and Tucker, R. "Bilingual Schooling through the Elementary Grades" *Language Learning,* no. 24, 1974, pp. 183-204.

Campbell, D. and Stanley, J. C. *Experimental and Quasi-Experimental Research Design.* Chicago: Rand McNally & Co., 1966.

Caplan, N. "Social Research and National Policy: What Gets Used, by Whom, for What Purposes, and with What Effects," Guttentay, M., ed., *Evaluation Studies Review Annual,* vie 2. Beverly Hills: Sage Publications, 1977.

Cardenas, J. *Preliminary Report: IDRA Response to the AIR Evaluation of the Impact of ESEA Title VII Spanish/English Bilingual Education Program.* San Antonio: IDRA, 1977.

_____. *AIR Evaluation of Bilingual Education.* San Antonio: IDRA, 1978.

Carsrud, K. and Curtis, J. *ESEA Title VII Bilingual Program: Final Report.* Austin: Austin Independent School District, 1979.

Cohen, A. *A Sociolinguistic Approach to Bilingual Education.* Rowley: Newbury House Publishers, Inc. 1975.

Cohen, A. and Swain, M. "Bilingual Education: The Immersion Model in the North American Context," *TESOL Quarterly,* no. 10, 1976, pp. 45-54.

Cohen, Bernard. *Final Evaluation Report of the 1970-71 New Haven Bilingual Program.* Darien: Dunlap & Associates, Inc., 1974.

Corpus Christi Independent School District. *Evaluation Report for the Title VII Aprendemos en Dos Idiomas Bilingual Program 1979-1980.* Corpus Christi: Corpus Christi Independent School District, 1980a.

_____. *Follow-Through Program Evaluation Report: 1979-1980.* Corpus Christi: Corpus Christi Independent School District, 1980b.

Cottrell, M. C. "Bilingual Education in San Juan County, Utah: A Cross Cultural Emphasis," Presented at AERA, New York City, February, 1971.

Covey, D. D. "An Analytical Study of Secondary Freshman Bilingual Education and Its Effect on Academic Achievement and Attitude of Mexican American Students," Ph.D. dissertation, Arizona State University, 1973.

Danoff, M. N.; Coles, G.; McLaughlin, D. H. and Reynolds, J. *Evaluation of the Impact of ESEA Title VII Spanish/English Bilingual Education Program.* Vols. I-IV. Palo Alto: American Institutes for Research, 1977-1978.

Davis, F. B. *Philippine Language Teaching Experiments.* Quezon City: Alemar - Phoenix, 1967.

Ebel, R. "Some Limitations of Basic Research in Education." *Phi Delta Kappan*, no. 49, October 1967, pp. 81-84.

Elashoff, J. D. "Analysis of Covariance: A Delicate Instrument." *American Educational Research Journal*, no. 6 (3), 1969, pp. 383-401.

_____ and Snow, R. E. *A Case Study in Statistical Inference: Reconsideration of the Rosenthal-Jacobson Data on Teacher Expectancy*. Stanford: Stanford University, 1970.

Engle, P. L. *The Use of Vernacular Languages in Education: Language Medium in Early School Years for Minority Language Groups*. Arlington: Center for Applied Linguistics, 1975.

Ennis, R. H. "On Causality." *Educational Researcher*, no. 2, June, 1973, pp. 4-16.

Ferguson, G. A. *Statistical Analysis on Psychology and Education*. New York: McGraw - Hill, 1976.

Gray, T. *Response to AIR Study*. Arlington: Center for Applied Linguistics, 1977.

_____ *Challenge to USOE Final Evaluation of the Impact of ESEA Title VII Spanish/English Bilingual Education Programs*. Arlington: Center for Applied Linguistics, 1978.

Guilford, J. and Fruchter, B. *Fundamental Statistics in Psychology and Education*. New York: McGraw-Hill, 1973.

Hempel, C. G. "Fundamentals of Concept Formation in Empirical Science", *International Encyclopedia of Unified Science II*, no. 7., 1952, p. 6.

Kaufman, M. "Will Instruction in Reading Spanish Affect Ability in Reading English?" *Journal of Reading*, no. 17, 1968, pp. 521-527.

Lambert, W. E. "A Canadian Experiment in the Development of Bilingual Competence." *The Canadian Modern Language Review*, no. 31, 1974, pp. 108-116.

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

Leyba, C. F. *Longitudinal Study: Title VII Bilingual Program, Santa Fe Public Schools*. Los Angeles: National Dissemination and Assessment Center, 1978.

Malherbe, E. L. *The Bilingual School*. London: Longmans Green, 1946.

McConnel, B. *Individualized Bilingual Instruction: Final Evaluation, 1978-1979 Program*. Pullman: author, 1980.

Miller, H. L. "Pollyanna in the Policy Patch: A Response to Frederick Wirt," *Education Evaluation and Policy Analysis*, no. 5 (3), September-October, 1981, pp. 83-93.

Moore, F. B. and Paar, G. D. "Models of Bilingual Education: Comparisons of Effectiveness." *The Elementary School Journal*, no. 79 (2), 1978, pp. 93-97.

NABE, *News Release*, March 2, 1981.

National Institute of Education. *A Reexamination of the Evaluation of the Impact of ESEA Title VII Spanish/English Bilingual Education Programs*. Washington, D. C.: NIC, 1979.

Nickel, K. N. "Experimentation, Extrapolation, Exaggeration: Their Name is Research." *Phi Delta Kappan*, no. 59, 1979, pp. 260-261.

Olesini, J. "The Effect of Bilingual Instruction on the Achievement of Elementary Pupils," Ph.D. dissertation, East Texas State University, 1971.

O'Malley, J. M. "A Review of the Evaluation of the Impact of ESEA Title VII Spanish/English Bilingual Education Program," *Bilingual Resources*, no. 1 (2), pp. 6-10.

Patton, M. Q. *Utilization - Focused Evaluation*. Beverly-Hills, CA: Sage, 1978.

Pena-Hughes E. and Solis, J. *"ABC's"* McAllen Texas: McAllen Independent School District, 1980.

Plante, A. J. *A Study of the Effectiveness of the Connecticut "Pairing" Model of Bilingual-Bicultural Education*. Hamden: Connecticut Staff Development Cooperative, 1976.

Popham, W. J. *Educational Evaluation*. Englewood CA: Prentice Hall, 1975.

Ramos, M.; Aguilar, J. and Sabayan, B. F. *The Determination and Implementation of Language Policy*. Queson City: Alemar - Phoenix, 1967.

Rossi, P. H. "Comments on Title VII Evaluation" (University of Mass. at Amherst.) MSS., 1979.

Skoczylas, R. "An Evaluation of Some Cognitive and Affective Aspect of a Spanish-English Bilingual Education Program," Ph.D. dissertation, University of New Mexico, 1972.

Skutnabb-Kangas, T. and Toukomaa, P. *Teaching Migrant Children's Mother Tongue and Learning the Language of the Host Country in the Context of the Sociocultural Situation of the Migrant Family*. Helsinki: Finnish National Commission for UNESCO, 1976.

Skutnabb-Kangas, T. *Language in the Process of Cultural Assimilation and Structural Incorporation of Linguistic Minorities.* Washington, D. C.: National Clearinghouse for Bilingual Education, 1979.

Siegel, S. *Non-Parametric Statistics for the Behavioral Sciences.* New York: McGraw-Hill, 1956.

Stanley, J., ed. *Improving Experimental Design and Statistical Analysis: Proceedings of the Seventh Annual Phi Delta Kappa Symposium of Educational Research.* Chicago: Rand McNally & Co., 1967.

St. John Valley Bilingual Project. *Five-Year Evaluation Report 1975-1980: Frenchville, St. Agatha, Van Buren, Madawaska.* Madawaska: St. John Valley Bilingual August, 1980.

Stern, C. *Final Report of the Compton Unified School District's Title VII Bilingual-Bicultural Project: September 1969-June 1975.* Compton City: Compton City Public Schools, 1975.

Swain, M. *Bilingual Schooling: Some Experiences in Canada and the United States.* Toronto: O.I.S.E., 1972.

_____ and Barik, H. "French Immersion Classes: A Promising Route to Bilingualism," *Orbit,* no. 31, 1974, pp. 102-107.

Thorndike, Robert L. "Dilemmas in Diagnosis," MacGinitie, W., ed *Assessment Problems in Reading.* Newark: International Reading Association, 1973, pp.57-67.

Thorndike, R. L. "Pygmalion in the Classroom by Robert Rosenthal and Lenore Jacobson, Book Review," *American Educational Research Journal,* no. 5, 1968, pp. 708-711.

_____, ed. *Educational Measurement:* Washington, D.C.: American Council on Education, 1971.

_____ and Hagen, E. *Measurement and Evaluation in Psychology and Education.* New York: Wiley, 1969.

Trevino, B. G. "Bilingual Instruction in the Primary Grades," *Modern Language Journal,* no. 54, 1970, pp. 255-256.

_____ "An Analysis of the Effectiveness of a Bilingual Program in the Teaching of Mathematics in the Primary Grades," Ph.D. dissertation, University of Texas at Austin, 1968.

Troike, R. C. *Research Evidence for the Effectiveness of Bilingual Education.* Arlington: National Clearinghouse for Bilingual Education, 1978.

Tucker R., Lambert, W. and D'Anglejan, A. "French Immersion Programs: A Pilot Investigation," *Language Sciences,* no. 25, 1973, pp. 19-26.

Weiss, C. "Improving the Linkage between Social Research and Public Policy," Lynn, Jr., L. E., ed., *Knowledge and Policy: The Uncertain Connection.* Washington D.C.: National Academy of Sciences, 1978, pp. 33-42.

_____, ed., *Using Social Research in Policy Making.* Lexington: Lexington Books, 1977.

Weiss, C. H. "Where Politics and Evaluations Meet," *Evaluation* no. 1 (3), 1973, pp. 37-45.

Williams, F. *Reasoning with Statistics.* New York: Holt, Rinehart and Winston, 1979.

Winer, J. B. *Statistical Principles in Experimental Design.* New York: McGraw-Hill, 1971.

Young, C. J. and Comtois, J. "Increasing Congressional Utilization of Evaluation," Zweig, F. M., ed., *Evaluation in Legislation.* Beverly Hills: Sage Publications, 1979, pp. 57-79.

Zirkel, P. A. "An Evaluation of the Effectiveness of Selected Experimental Bilingual Education Programs in Connecticut," Ph.D. Dissertation, University of Connecticut, 1972.

BEEHIVES AND ICEBERGS

by
Joseph W. Beard and Marjorie Powell

If teachers teach as they were taught, what do we do now?

These comments are reflections on presentations made during the fourth day of the 1981 Language Assessment Institute held at National College of Education on June 17-20, 1981. Although our references are specific to the four presentations made on that day, responses of participants indicate that our comments reflected points made in the three prior days of presentations.

In a real sense, the four presentations made on the final day brought into focus many of the concepts and ideas that had been presented during earlier discussions at the Institute. Gary Keller's presentation on "Integrating Language Assessment with Teaching Performance in Subject Areas" gave perspective to the relationship between language and the various content areas covered in our schools. The presentation delivered by Inez Bosworth on "Foreign Language and Bilingual Assessment: Issues and Approaches" demonstrated bridges between the bilingual education movement as it has been experienced over the past decade and the current affirmation of needs in the foreign language area. The presentation on "Language Proficiency Assessment for Vocational Occupational Training" delivered by Ron Perlman gave breadth to the understanding of work and career aspirations of students who come from language backgrounds other than English.* Stan Seidner's presentation on "Establishing Entry/Exit Criteria for Language Assessment" gave significant evidence of the state and status of bilingual education in colleges and universities which prepare professional staff for bilingual education programs.

The Keller presentation was particularly gratifying; a background in science education has influenced the author's constant search for people who are dealing with the issue of language usage in the various subject matter areas of the curriculum, particularly mathematics and science. It has seemed at times as though we expect our work using general communication language in bilingual education to automatically result in an understanding of content-specific areas, in the same way that alchemists worked on changing lead into gold. The focus over the past years on general language proficiency, coupled with a concern for competence in language arts, reading and culture, has resulted in our not addressing the use of language in the various subject areas of the curriculum.

Dean Keller's exploration of the language needed for teaching in the various subject matter areas in a bilingual format was both well-developed and consistent in its description of: (a) language variety or register as appropriate for each content area, (b) a selected review of types of tests which might be relevant, (c) recommendations on types of instruments or procedures which might be used to resolve issues raised and (d) the fundamental problem of a lack of a defined language variety or register resulting in a lack of a clear or well-defined language corpus of each of the specific content areas. The presentation raised a number of questions which, though specific to this one presentation, have applicability to the general set of presentations and discussions developed during the course of the whole institute.

1. *What is the primary goal of bilingual education?*

   If the mandates of federal and state funding define the primary purpose of all bilingual education as acquisition of a single language, English, then the issue of the range of language needed to learn in many subject areas becomes a moot argument. In this instance, the development of communication skills in English and the learning of basic tranfer skills in a home language may be all that is necessary to achieve transition to English. However, if a program, school district or learning environment has as its primary purpose or function the development of individuals who have full skills in speaking, reading and writing in at least two different languages, then the issue of how these skills are developed in each of the subject matter fields, as well as in the breadth of human discourse, becomes a significant factor. In the former case, the issues raised by Dean Keller are academic in nature and have no real significance in the development of program focus for students achieving sufficient skill in the English language to continue their learning experiences in only that language. In the latter case, that of expanding one's language experience, exposure and learning opportunity in more than one language (English, the home language, a third or fourth language), the issue of content-specific language as described and developed by Dean Keller will have to be dealt with.

2. *What is the applicability of issues which Dean Keller discussed in terms of Spanish-English bilingual programs to other languages included among our bilingual education programs as well as to the political, economic and social interactions of the late 20th century?*

   Those language groups most similar to English with almost the same alphabet, such as the Romance and the German languages, pose a much smaller problem in the content-specific language areas (especially with the penchant for borrowing words between languages such as French, Spanish, English and German) than do those languages with different

*Unfortunately, the paper was not submitted in time for the publication deadline.

sound systems and/or different alphabets. Even though the words which describe the engineering feats of the 20th century transcend language groups, the basic concepts of science and mathematics are expressed significantly differently within each language community. This difference necessitates a concern for how those concepts can be transmitted to learners who come from those different language groups.

3. *Would the level of sophistication of the learner in a given subject matter field influence his or her need for the development of content-specific language in the home language?*

It may be that the basic skills in mathematics such as adding, subtracting, dividing and multiplying or the basic processes of science, such as inquiry and decision making, contain a language corpus in each language which is fairly consistent and applicable. This basic language may be available within the general language knowledge of both the teacher and the learner. It may be that the development of advanced skills in areas such as calculus, physics, chemistry, topography, etc., results in a more specific language base that would create the kind of problems and difficulties which Dean Keller addressed. Thus, there may only be a need to define the language corpus for advanced studies within a subject-matter area.

4. *Would consideration of how other countries respond to language-different people have value for us in the United States of America?*

For example, if we consider Japan and the development of personnel who speak two languages to serve that country's commercial efforts or the response to linguistically different learners coming to school, would we find Japan facing a similar issue or have they found a way to insure greater success in the development, implementation and measurement of language curriculum?

5. *How much can we separate the issue of the development of content-specific language in a home language from the issue of the development of concepts in a home language which might then be applied in another language?*

The question is whether the development of concepts requires the use of the technical language. It may depend upon the teaching approach used. If the teacher is presenting the concepts, then technical language may be necessary to discuss the concepts. If the teaching approach is process-oriented, designed to assist learners to develop an understanding of how mathematical concepts or scientific knowledge is derived, the same issue of the body of technical language may not be as significant.

6. *Is there a gap between the reality of human interaction using language and the ideal perspective of how language might or could be used?*

The realities of the classroom and of the contexts surrounding the classroom, such as district and union hiring policies, the supply of qualified teachers, and the procedures for teacher certification, may all impose constraints on the achievement of ideal goals in terms of teacher knowledge of and skill in using two languages for instruction. As certification processes have developed over the past twenty years and faculties at all school levels have reaffirmed the principle of academic freedom, where are the forces for change which can result in effective classroom behavior and linguistic skills of teachers? Will the current political perspective which results in an English language chauvinism significantly alter the concern for development of language in subject-matter areas?

7. *How might the development of bilingual education approaches and concepts be truly institutionalized in schools?*

Implicit in the notion of developing content area language in multiple subject areas and in multiple languages is the anticipation that learners will indeed become bilingual or biliterate. The development of bilingual instruction in multiple content areas could indeed make bilingual education an enrichment process fully institutionalized in the schools.

8. *How does bilingual education differ in its issues and problems from learning and teaching experiences in monolingual English programs?*

We might suggest that the issues are the same, that they only differ in the degree of difficulty and of significance. The issue of appropriate language used with consistency and reliability in monolingual English science and mathematics classrooms is similar to the issue of utilizing the appropriate language for instruction in Spanish, Chinese, or any other language. Bilingual education highlights such problems and issues, but they then need to be addressed outside of the political and economic factors of the existence of bilingual education.

9. *Do differences exist in the potential of developing language understanding among experienced teachers and people involved in teacher education programs?*

The development of a teacher pool during the teacher shortage of the 1950's and the early 1960's resulted from and led to conditions very different from those which exist in 1981. The current conditions, a shortage of positions for teachers, coupled with growing shortages of teachers trained in specific subject-matter areas and willing to teach in specific geographic areas, such as rural areas, will impact the development of our schools through the 1980's. The opportunity now exists to utilize our resources much more effectively in developing teachers whose language (and other) skills can, in fact, be responsive to the needs of the students they serve.

10. *Are basic language skills necessary or appropriate in order for students to develop advanced learning?*

    Can the language necessary to master advanced concepts in a language be separated from basic communication skills in a language? The existence of scientists who can comprehend scientific texts in a second language but who cannot carry on simple conversations in that language may point to the separability of basic and technical language learning. As students move from basic communication skills to the development of language which can be used to describe content-based experiences and concepts, students may need to learn the difference between (1) their own inherent abilities to communicate in whatever language is appropriate for the environment in which they find themselves and (2) their need for a more scientific discourse associated with the content area.

11. *Should bilingual education and the issues associated with language assessment respond only to the existence of federal or state or extra local funding specifically for bilingual programs or should it become an inherent part of a school program?*

    What approaches may be taken to institutionalize further the basic concept of responding to linguistically different learners and the development of multiple language skills?

Additional questions and issues were raised by the other three presentations of the final day of the institute. In a delivery of a paper by Woody Woodford, Inez Bosworth presented the history of language assessment testing over the past twenty yers and gave an indication of current and future directions and issues in the field of language proficiency testing. A number of the problems encountered in assessing the extent to which English-speaking students have acquired skills in a second language in secondary or higher education programs remain unresolved as we struggle to assess the language skills of students and teachers in bilingual education programs. This historical review provides a valuable perspective on the current problems and proposed solutions.

In a paper co-authored with Else Hamayan, Ron Perlman discussed test construction and development in response to specific needs associated with vocational education. In some ways, these problems echo those identified by Gary Keller, in that they are related to the use of technical, rather than basic, language. However, basic communication skills are also essential to success in vocational settings, as these authors point out. This presentation also addressed the language and vocational interests of multiple ethnic groups, reminding us of the wide range of language and cultural groups for whom bilingual education is an important issue.

Stan Seidner brought closure to the series of concepts associated with the issue of language assessment by describing the approaches which colleges and universities use to determine the language proficiency of students in bilingual education programs and the types of people involved in making decisions about levels of language proficiency. His survey revealed the lack of consistency in assessment methods, the multiple people involved in assessment, and the focus on general communication skills. The survey also indicated the limited extent to which bilingual teacher training programs have been institutionalized in colleges and universities. There remains a strong reliance on federal dollars and regulations.

These presentations sparked several questions which also are applicable to previous presentations at the Institute.

1. *What is the nature of the testing situation?*

   Although it may be a truism which we tend to forget; the circumstances in which the client is being tested (environment, number of people, noise, light, etc.) can have a significant effect on the likelihood that the person being tested will demonstrate the particular skill or knowledge which the test is intended to measure, if the person, in fact, possesses the skill at a level which the test is designed to measure.

2. *What effect does the age of the people being tested have on their ability to demonstrate the skill which the test is designed to measure?*

   Tests developed and used with young children are not appropriate in testing language proficiency skills among high school students or adults and vice versa. One needs to designate the age and experience of individuals being tested in order to begin to determine the appropriateness of tests and the interpretation of results.

3. *How much influence do the content and the format of a test have on results of that test?*

   Again, the need to specify the impact of the content and format of the test on expected and identified results is a need that we recognize but one that we sometimes overlook in the context of discussions such as those presented at the Institute.

4. *What are the technical issues surrounding the development, use and interpretation of tests?*

   Which of these technical issues can be anticipated and controlled for; which ones are unique to a particular testing situation?

5. *What is the purpose for giving a particular test?*

   What decisions might be made from an interpretation of the results of a given test? In language assessment in particular, the purpose for testing could well influence the kind of test used and steps necessary to develop that test.

6. *What are the local, state and national level political issues which influence the decision-making process prior to, during and after the use of a test for a*

particular purpose?

> The development of language assessment processes for bilingual or foreign language education will not take place in a vacuum, but rather, in an environment in which political and economic issues become contributing factors to the decision-making process.

Each of the papers illustrated, in varied ways, the effects of political issues. These issues will continue to interact with the technical issues to influence and shape future directions in language assessment. While we worry about the technical and political issues, we must also grapple with the real effects on the individuals whose language proficiencies are being assessed, the students for whom all this effort is being undertaken.