DOCUMENT RESUME

ED 144 410 FL 008 916

AUTHOR Brown, Mark E.; Zirkel, Perry A.
TITLE Emerging Instrumentation for Assessing Language Dominance. Occasional Papers on Linguistics, No. 1.
INSTITUTION Southern Illinois Univ., Carbondale. Dept. of Linguistics.
PUB DATE Apr 77
NOTE 32p.; Proceedings of the International Conference on Frontiers in Language Proficiency and Dominance Testing (1st, Southern Illinois University, Carbondale, Illinois, April 21-23, 1977)

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS Bilingual Education; *Bilingualism; Bilingual Students; Communicative Competence (Languages); Diagnostic Tests; Educational Testing; English (Second Language); *Evaluation Criteria; Language Proficiency; Language Skills; *Language Tests; Language Usage; *Measurement Techniques; Morphology (Languages); Phonology; Pronunciation; Psycholinguistics; Screening Tests; Second Language Learning; Sociolinguistics; Spanish Americans; Spanish Speaking; Syntax; *Test Reviews; *Test Selection; Vocabulary
IDENTIFIERS *Language Dominance

ABSTRACT

This paper offers a two-step review to be used in designing dominance assessment plans and in determining appropriate instrumentation. The first step provides a classification system of dominance instruments according to testing specificity and strategy. The second step suggests criteria by which such instruments can be evaluated and selected. Selected dominance assessment instruments are categorized in a three-way descriptive matrix. The global/specific dimension distinguishes instruments which tend toward generic screening of gross language behavior from those which tend toward a refined classification of specific language indicators. Within the global and specific modalities, oral and aural performance subclasses are designated. The third dimension consists of four major strategies: rating, home interview, indirect, and parallel instruments. Specific examples of instruments are given to clarify how the classification matrix operates. Criteria for evaluating and selecting tests include examinee factors relating to developmental and cultural appropriateness, administrative and logistic factors, and psychometric considerations. A sample evaluation of Burt's Bilingual Syntax Measure is provided. (CLK)

Emerging Instrumentation for Assessing Language Dominance

Mark E. Brown, University of Hartford
and
Perry A. Zirkel, Lehigh University

In the early 1950's, UNESCO (1953) sponsored a study of the use of vernacular languages in education. The study's conclusions were in part:

> The mother tongue is a person's natural means of self-expression, and one of his first needs is to develop his power of self-expression to the full.
>
> Every pupil should begin his formal education in his mother tongue.

Nearly three decades have passed since that message, and with time psycholinguistic and sociolinguistic sophistication has expanded in the educational domain. Ironically the fate of non-Anglo linguistic minorities has not, until recently, benefitted from this sophistication. We have treated Spanish-speaking pupils--particularly of Puerto Rican, Mexican-American, and Cuban background--with either benign neglect or benevolent paternalism, quite independent of the growth of research expertise in the domain of language dominance testing and bilingual/bicultural education. The melting pot perspective we have maintained for generations has been the filter for many of our attempts at improving educational opportunities for Spanish-speaking students. We have with myopic insight, implemented ESL programs for students with "exotic" surnames, without ascertaining the student's language dominance and proficiency. We have historically provided the bootstraps for linguistically disadvantaged Horatio Algers, but not realistic programs to meet instructional needs.

But things have changed recently, though not out of rage or indignant posturing by educators. Things have changed because legislative and court precedents have been set in recent years to encourage or mandate bilingual programs (Teitelbaum and Hiller, 1977). The instructional opportunities of Spanish-speaking students, beginning with the opportunity to be assessed accurately in terms of their language dominance, have expanded in proportion to the revision of our legal conceptions of the students' (children's) status and rights. We may be on the road to redress the heretofore unenforceable moral obligation to "special interests" of disadvantaged minorities which may have backlashed into "institutionalized racism" (Baratz 1970).

As Rodham (1973) articulated:

> The needs and interests of a powerless individual must be asserted as rights if they are to be considered and eventually accepted as enforceable claims against other persons or institutions.

Dominance testing occupies a unique linkage role in the educational-legal entanglement of linguistic and cultural pluralism. In serving the child as an initial assessment procedure, it also serviced institutions as a means for acquiring baseline data from which to make program decisions.

The landmark Supreme Court decision in Lau v. Nichols is a clear instance of how the role of dominance testing must be carefully prescribed before program decisions are made. In finding that a school district's failure to provide non-English speaking students with appropriate instructional programs is a violation of Title VI of the Civil Rights Act of 1964, the Supreme Court reasoned:

> There is no equality of treatment merely by providing students with the same facilities, textbooks, teachers, and curriculum, for students who do not understand English are effectively foreclosed from any meaningful education.

The basis of their decision were the guidelines issued by the Office of Civil Rights(OCR) to implement Title VI in the public schools. The guidelines state in part:

> Where inability to speak and understand the English language excludes national origin minority group childrer from effective participation in the educational program offered be a school district, the district must take affirmative steps to recitfy the language deficiency in order to open its instructional program to these students.

Citing Serna v. Portales New Mexico School District (1974), Aspira v. Board of Education of the City of New York (1974), Keyes v. Denver Unified School District (1974), the Lau General Assistance Center (1976) at Columbia University stated that "all court decisions which have applied and interpreted Lau v. Nichols have concluded that Lau required bilingual education to overcome the deprivations suffered by limited English-speaking children." The bootstrap philosophy of ESL has been specifically identified, via OCR's post-Lau guidelines, as an inadequate instructional strategy for secondary school students. Rather, the guidelines require language dominance assessment and bilingual/bicultural education. "Effective participation" cannot be limited to artificial treatment of the linguistic/cultural heritage of students. A former OCR Director Mazon (1976) cites former Director Pothinger:

> Children should not be penalized for cultural and linguistic differences, nor should they bear

a burden to conform to a school-sanctioned culture by abandoning their own.

Language dominance assessment procedures for bilingual program planning converge with the procedures school systems are also required to implement in accordance with federal legislation (P.L. 93-380 and P.L. 94-142) on non-discriminatory assessment in the domains of intelligence and achievement testing. The disproportionate representation of non-Anglo minorities in placements for retarded or learning disabled persons requires detailed scrutiny. Language dominance assessment may be the initial step in re-evaluating students who have become victims to "Anglo-centric" norms and stigmata (Mercer, 1974).

While the original Lau decision pertained to "only" 1800 Chinese-speaking pupils in San Francisco, subsequent legal decrees have required school districts to implement large-scale assessment of student language dominance. The Aspira v. New York City Board of Education consent decree, for example, mandated system-wide dominance testing for over 250,000 Spanish-speaking pupils. Dominance assessment testing requires systematic procedures to ensure the gathering of reliable and precise information. Yet, when OCR and other legal agencies enforce compliance to current standards, the time necessary to develop or determine optimal dominance assessment procedures is often unavailable.

For this reason, we offer a two-step review which may be useful in designing dominance assessment plans and in determining appropriate instrumentation. The first step provides a classification system of dominance instruments according to testing specificity and strategy. The second step suggests criteria by which such instruments can be evaluated and selected.

As peremptorily alluded to in the introduction to this paper, identifying bilingual or monolingual persons on the basis of surname is highly imprecise both for census data and--more importantly--for baseline data for bilingual program placement. With the ultimate intention of bilingual education for children who are identified as dominant in another language, as mandated by OCR compliance remedies, a functional classification of classroom bilingual performance is necessary. The conception of bilingualism as complete mastery, or literacy, in two languages (more appropriately termed by Zirkel (1976) "equilingualism") is rare and imposes unrealistic criteria. More appropriate is a description of bilingualism as a broad continuum of listening and speaking skills in two languages. This continuum not only has range in terms of degree of fluency or proficiency in a language, but must also have depth, in terms of domain or situational variables which influence constructive behavior. Fishman's (1967) sociolinguistic conceptions of domain-specific and diglossic bilingualism are useful perspectives in demonstrating such "depth." Hymes (1972) in his introduction to The Function of Language in the Classroom, also underscores this contextual depth. He noted that while it is scientifically absurd to describe children as coming to school with a linguistic tabula rasa, children may be linguistically deprived and even punished if the language of their natural competence is not that of the school "if the contexts that elicit or permit use of that competence are absent in the school; or if the purposes to which they put language,

and the ways in which they do so, are absent or prohibited in the school."

Bilingual competence is intertwined with the functional as well as structural expectations of the school and the home. In this way, Fishman's concept of diglossia, while seemingly an exotic description of far away language communities is a relevant dimension of bilingualism issues in the classroom. Students' receptive and productive competencies are influenced by the attitudes, values, and behaviors supported by each language that they master.

If dominance assessment is to reflect a complete linguistic profile of a student, it must elicit information which is representative of language structure, function, and context. In this case the dominance-testing procedure is analogous to criterion-referenced testing in areas of instruction. Just as the content of a criterion-referenced test, for example in reading skill mastery, must reflect accurately the content of reading skills taught in the classroom and the technique of instruction used, the dominance assessment must reflect the content of language usage--the structure, functions, and context--in the student's home and school situations.

Although existing individual dominance assessment procedures do not indicate the complete content of the students' language experience it is possible to construct profiles of student language behavior, by selecting from different assessment modes. In the next section we will highlight several modes within a classification matrix.

Modalities of Language Dominance Assessment

Various situational constraints, such as time factors, population variables, staffing, etc. will influence which modalities of dominance testing a school district will select to meet OCR compliance remedies or to self-initiate an identification process in the implementation of a bilingual/bicultural program. Figure 1 presents a framework for classifying dominance assessment modalities.

FIGURE 1

STRATEGY	GLOBAL MODALITY		SPECIFIC MODALITY	
	AURAL PERFORMANCE	ORAL PERFORMANCE	AURAL PERFORMANCE	ORAL PERFORMANCE
RATING SCALE		DAILEY'S LANGUAGE FACILITY TEST		NEW YORK CITY'S RATING SCALE OF PUPILS' ABILITY TO SPEAK ENGLISH - INCLUDES SPANISH RATING
HOME BACKGROUND INTERVIEW SCHEDULES	HOFFMAN'S BILINGUAL BACKGROUND SCHEDULE AND ZIRKEL'S REVISION			FISHMAN'S CENSUS FORMAT LANGUAGE LOYALTY QUESTIONNAIRE BROWN'S DOMAIN ATTITUDE SCALE
INDIRECT MEASURE		SPOLSKY AND MURPHY'S WORD AVAILABILITY SUBTEST		COOPER'S WORD NAMING AND WORD ASSOCIATION TASKS
PARALLEL TESTING	SHUTT'S ORAL COMPREHENSION SUBTEST IN SPANISH/ ENGLISH	JAMES' SPANISH PRODUCTION AND ENGLISH PRODUCTION SUBTESTS	NEW YORK CITY'S LANGUAGE ASSESSMENT BATTERY RECEPTIVE SUBTEST	BURT'S BILINGUAL SYNTAX MEASURE

Figure 1 presents selected dominance assessment instruments categorized in a three-way descriptive matrix. The global-v.-specific dimension distinguishes instruments which tend toward generic screening of gross language behavior from those which tend toward a refined classification of specific language indicators. Note first that "language-behavior" is, in itself, a generic term describing linguistic and attitudinal measures. Secondly, the distinction between "gross" and "refined" should not be misconstrued as an analogy between general achievement and diagnosis. A more appropriate analogy would be to the dichotomy between achievement testing and criterion-referenced testing, where the latter are more customized instruments derived from heirarchies of specific objectives, while the former are derived from more conservative, condensed sets of goal statements. Within the global and specific modalities, oral and aural performance subclasses are designated. On the opposite dimension are the four major strategies: rating, home interview, indirect, and parallel instruments. Specific examples of instruments are given to clarify how this classification matrix operates.

Rating Scales

Second in frequency to surname census measures, rating scales are common means of identifying Hispanic children's language dominance. Ratings are inescapably subjective. They are usually "administered" by teachers of the students under consideration, which is both an advantage and disadvantage. The most obvious disadvantage is conscious or unconscious rater-bias, including attitude toward the student's native language and culture, as well as feelings of job insecurity. Teachers who are monolingual in a language other than the student's will also be at a disadvantage in understanding rating criteria. Advantages range from saving the staffing costs for outside examiners to involving teachers in the planning process.

The use of trained bilingual examiners to do rating seems promising from the criterion of objectivity. However, this alternative must be weighed against the total training time and costs, the difficulty of designing practical observation settings, the need for establishing rapport with the students being rated. In the final analysis, the assessment or predictive "power" gained by objective bilingual raters may not warrent the costs.

Dailey's Language Facility Test

We have classified Dailey's instrument as a global rating scale since it is concerned with language maturity as measured by the cohesiveness and organization of story content elicited by pictorial stimuli (photos and drawings). Dailey's (1969) manual reports two scoring systems: one is a 9-point scale ranking story organization, which measures how well an individual is able to conceptualize and communicate in his chosen language. This score is purportedly independent of standard English vocabulary, enunciation, information, or grammatical exactness. The second scoring system codes errors or deviations from standard English pronunciation or usage and gives a diagnostic profile of ability to speak standard English.

The test purports to measure a person's language fluency from three years of age to maturity. Its "norms" however, are

described in terms of age-levels, somewhat like the diagnostic subtests of the Illinois Test of Psycholinguistic Abilities. As Dailey's manual indicates, "zones" of developmental language ability have been posted. However, the information upon which these zones hinge are only a skeletal or global view of two aspects of language: the cohesion of visually directed verbal expression and a sample of phonological and syntactic structures. The behavioral sample is useful for practicioners making early screening decisions, however, it is doubtful whether one can generalize as does the manual:

> The Language Facility Test had proven to be a useful tool for the early identification of the mentally retarded... The basic theory and rationale for the Dailey...is to extend the range of (the measure of comprehension) downward to assess at early ages the characteristics of the child who has the capacity to develop into one with a higher verbal I.Q. than another...how well an individual uses language the way he has been learning it.

These assumptions are too heavy a predictive load for this one test to bear.

There are advantages to the Dailey as a screening tool. Those who administer it need only a minimum of training (even clerks) according to the manual. The coding procedure for stories is complex and time consuming.

New York City's Rating Scale of Pupils' Ability to Speak English

In contrast to the Dailey, the N.Y.C. Rating Scale does not attempt to relate language proficiency to a general model of symbolic/cognitive functioning. It states specific criteria to assess four areas of oral language:

1) vocabulary usage - use of function and content words;
2) structural pattern - a systematic arrangement of words that are characteristic of the language an individual is speaking and meaningful to all speakers of the language;
3) pronunciation - oral production of vowel and consonant sounds or combinations of sounds;
4) intonation - rise and fall of the speaker's voice (melody of language).

The device is intended as an "efficient screening device for oral proficiency--not intended to replace comprehensive diagnostic information...can be used as an aid in determining grouping or placement of pupils for either English or other language instruction." As with the Dailey, a series of drawings is used to elicit oral language which are then rated on a 7-point scale.

Note these distinctive characteristics of the N.Y.C. Scale: the elicitation procedure (which takes from 3-6 minutes to administer)

is based on research findings (Brown, 1973; Williams and Naremore, 1969; Cazden, Hymes and John, 1972) which show that different students from the same linguistic community require different elicitation demands to perform optimally. Thus, the procedure starts by asking an open-ended question ("What can you tell me about this picture?"), but also offers the examiner optional probing questions for "reluctant" speakers.

The probing questions are designed to elicit three levels of oral performance:

1) identification or labeling
2) functional relationships
3) interpretation of context

In addition to probing questions, five grammatical closure items are included to elicit specific linguistic samples which the examiner may require. There are variable points of entry and exit for the student being assessed. If the examiner can make a judgement without going through the whole picture series, it is acceptable to make a rating. Further, the examiner needs no special training other than inservice with the rating materials.

The rating scale itself was customized to match nominal proficiency judgements to specific behavioral criteria, and is, thus, in contrast to the Dailey model. The rating scale range is from non-English/monolingual Spanish through structural and semantic vascillation between two language systems to English dominant. In addition, an open category was included as "Ø": a child who does not produce English/Spanish for reasons other than monolingual dominance. This category was maintained to screen for generic languages/function. For each criterion specific behavioral examples are cited.

Scoring is weighted so that vocabulary and structure count twice as heavily toward the total proficiency rating as do pronunciation and information.

Home Background Interview Schedules

The collection of language census data adds an important dimension to dominance assessment. Such interviews solicit information on how language usage: the "grammatical analysis... supplemented by ethnographic description, ethnohistory..." (Gumperz & Hernandez-Chavez, 1972).

In other words, as important as frequency of language usage data is to developing a home dominance scale, interviews should also be used to probe (in both languages) the complexities of social interaction and the contrastive uniqueness between cultural influences. Again, as Gumperz and Hernandez-Chavez (1972) related:

> Minority groups in urbanized societies are never completely isolated from the dominant majority. To study their way of life without reference to surrounding populations is to distort the realities of their everyday lives...almost all minority group members...have at least a passive knowledge of the dominant culture. What sets them off from others is not simply

the fact that they are distinct, but the juxtaposition of their own private language and lifestyle with those of the public at large.

Hoffman's Bilingual Background Schedule

Over forty years ago, Hoffman (1934) developed an interview instrument which was first used with 10 and 11 year old Jewish students attending Hebrew parochial schools. It is historically unique format since it requests respondents to provide information about:

1) languages spoken and understood by all members (including self) of immediate family
2) dyadic use of language between members of the family (including self)
3) languages of media used in the home (including newspapers, radio) by family members
4) literary and aesthetic language preference (including books, theatre)
5) written use of language (including letters)
6) cognitive use of language ("What language do you think in?")

Zirkel (1973) condensed, refined and updated the Hoffman scale, including T.V. and films, and replaced Hoffman's "Never to Always" productive/receptive scale with a numerical system. Both the Hoffman and Zirkel forms obtain a total dominance score weighted by sub-score categories (self/other: speaks, understands, reads, writes in X language). Both interview formats are considered global, however, using the Gumperz & Hernandez-Chavez (1972) criteria of ethnographic complexity.

Fishman's Census Format and Language Loyalty Questionnaire

The Fishman language (1971) language census and interview instruments were developed within a sociolingualistic research paradigm. Yet there is great value in the degree of specificity each instrument provides for assessing language dominance, language function, and language contextualization.

While there are basic similarities between this and the Hoffman-Zirkel schedule, an important difference is Fishman's attempt to document both current dominant language of respondent and family as well as "first language" acquired for different uses. This fits into Fishman's (1966) theoretical concern for "language loyalty" and language ethnohistory. Another distinctive attribute is Fishman's breakdown of census information by

1) domain (home, school, work, church)
2) function (conversation, instruction)
3) status of participant (supervisor, customer, priest)

The Fishman census interview treats language performance in speaking, understanding, reading and writing in three ways: 1) developmentally; 2) currently; and 3) relative frequency; and in three contextual domains: home use; religious use, work use. Only two items in the

schedule ask personal language preference and the language of school instruction. Self ratings of competence/adequacy were not included because of purported low validity, however, self ratings were inferred from responses to current performance questions.

Fishman (1971) found that the census/recensus reliability for this interview is high (approximately .80). The validity of his census is supported by the five distinct factors yeilded through inter-correlation techniques. These factors assist the identification of functional language dominance by providing information on Spanish literacy, English oral and written usage, Spanish oral usage in the home, Spanish usage at work, and Spanish usage in religion.

For purposes of assessing dominance in the instructional domain, Fishman's census interview instrument is limited to a profile of adult/parental language patterns, and thus, it only indirectly impacts on students. For example, the degree to which the survey identifies claims for bi-literacy in the home may help to provide a more complete picture of environmental influences to school performance.

Fishman has devised a 64-item questionnaire which does not focus on high school student language usage and preferences, as well as attitudes, towards Puerto Rican culture (with an adapted form for primary school-age kids). Taken collectively, the items in this questionnaire present a complex contextual picture of language and cultural loyalty. It solicits language dominance information in terms of current competencies as well as attitudinal factors (e.g., "Do you usually speak in English when you want your parents or grandparents to do you a favor? or When you have children do you want them to be able to speak Spanish fluently?"). Also, like the Hoffman-Zirkel interview, information on preferences for Spanish movies, etc. is rejected. A separate 12-question scale was developed by Brown (1968) as a supplement to the 64-question scale to probe further on how students perceive the utility of Spanish/English. Questions like "Can you make a better impression on a boy you like by speaking more English than Spanish?" and "Do Puerto Ricans who mainly speak English get ahead faster in the job world of N.Y.C. etc....?" complement the use of a third related scale on committment to Spanish language cultural groups.

Indirect Measures

McNamara (1967) used "indirect measures" as a catch-all category for instruments used by psycholinguists to determine language dominance. For example, in studying the contextualization of language dominance, word-naming and word association techniques have been used to assess the frequency with which languages are spoken.

Cooper's Word Naming/Word Association Tasks

Cooper (1969) worked with Fishman's Bilingualism in the Barrio documentation and developed several techniques for assessing language dominance in specific social domains. One is the "Word Frequency Estimator." Using a seven-point ranking from "more than once a day" to "never", a person estimates how often he/she uses approximately eight different words. The examiner reads lists of words in English and Spanish for the examinee to rate/ranks. A second technique is "Word Naming." The examiner asks the respondents to "tell me as many

different words in English as you can. They don't have to be big words or words they teach you at school..." for each of five domains (home, neighborhood, church, school, jobs). The task is given in English and Spanish.

Spolsky-Murphy Word Availability Subtest

A briefly administered (10 minute) instrument, the Spolsky-Murphy assumes also the dominence is socially contextualized and that the self-report of respondents in an accurate measure of language usage and preferences. The instrument is in six blocks--three in Spanish and three in English. These parallel blocks tap comprehension and production through interview questions and picture stimulus elicitation. In a procedure similar to the Cooper "Word Naming Task: a "Word Availability Task" is administered in five domains of usage.

Parallel Testing

Dominance assessment techniques discussed above have played key roles in ethnographic research in bilingualism. However, in terms of instructional program planning or grouping and placement of students, none of the above provide practical, objective decision point data. Parallel testing with English and Spanish alternate forms in specific receptive and expressive performance areas begin to lay a framework for gathering data for a diagnostic/prescriptive instructional model rather than a classification or census model.

The instrumentation comprising parallel testing include school-district devised "special instruments," university research-purpose assessment designs and commercially published tests. The following examples represent this range. Several comprehensive listings of these instruments are available (e.g., Dissemination Center, 1975; Non-Biased Assessment, 1976; National Consortia, 1971; Zirkel, 1976). The list of such instruments continues to lengthen each hear.

Shutt's Primary Language Indicator Test (SPLIT)

This is a pictorial screening device consisting of sub-tests in aural comprehension in Spanish and English respectively, and oral fluency in English. It is interesting to note that the impetus for the design of SPLIT was to meet language dominance requirements for psychological assessment of students and their placement in special education programs. Resulting from a 1972 U.S. District Court order in Arizona, dominance assessment was required prior to psychological evaluation. While the standardization sample of 130 is small, attempts were made to specify representation by ethnographic criteria (border-rural, border-urban, inland-rural, etc.). The standardization population consisted of bilingual students from 5-13 years with approximately equal number in each grade K-7. Judgements of bilingualism was a cumulative assessment by parents, teachers, and school principals. Norms devised from this sample, however, are not reported as age/grade means but as "collective" means for each subtest. The manual does not encourage local norm development.

We have classified SPLIT as a general assessment instrument for several reasons: (1) the test relies on face validity (based on

"enthusiastic acceptance of the experimental edition by agencies and public school personnel") rather than specific cross validation with external criteria. (2) The item content taps receptive performance in Spanish and English with only a limited sample of "vocabulary." While time considerations are critical in all dominance testing, the SPLIT needs a wider sample of items and needs to reorganize them in terms of degree of semantic frequency level and linguistic structural complexity. Note these examples from the Spanish translation.

"Look at this. This is candy. Look at these. Which has the candy?"

"Look at this. This is a woman in the kitchen. Look at these. Which is a face?"

"Look at this. This is a boy. Look at these. Show me the fat one."

"Look at this. This boy is thirsty. Look at these. Who is drinking the drink?"

"Look at this. Maria is eating. Look at these. Which one do you not eat?"

Examples from English:

"Look at these pictures. Which one would you use for drinking?"

"Look at these pictures. Put your finger on the picture of the boy wearing a cap."

" Look at these pictures. Show me the woman who has finished ironing."

Each structure employed may be important, but as presented, there is no consistency in the format for presenting items nor in the groupings of items by linguistic or semantic criteria. Both of these factors influence the student's responding set. The Spanish items, however, do reflect cultural content as purported and this concept serves as a useful technique for the content validity in Spanish dominance. Although our focus is the "oral" [sic] subtest, it should be noted that the "Verbal Fluency" subtest is only given in English, and is coded with very general scoring sriteria.

Burt's Bilingual Syntax Measure (BSM)

In contrast to SPLIT, the BSM is a refined instrument designed to measure the individual primary grade (K-3) student's acquisition of grammatical structures in English and Spanish respectively. The BSM is an oral instrument which uses illustrations which are within a child's frame of experience and not culture-bound to elicit natural speech samples. Only syntactic rules are assessed in the BSM, unlike other forms of parallel testing. The testing procedure combines open-ended response with structured probe questions by the examiner.

The BSM rationale declares that traditional vocabulary tests make misaligned assumptions of common experience and shared "semantic space" among Hispanic children from distinctly different backgrounds. Therefore, the BSM avoids testing vocabulary. In like manner, the BSM rationale points out dialect and idiolect drawbacks in testing for pronunciation and the awkwardness and artificiality of testing for functional use of language. The resulting BSM dominance test selects syntax. Burt et al. (1976) stated in their technical manual:

> Of the various characteristics of a particular language, syntax shows the least variation among speakers of the language. Most research in child language acquisition focuses on syntax and, as a result, descriptions of syntactic developmental sequences provide invaluable input for the assessment of language development.

Tne BSM recognizes the heirarchical structure of both Spanish and English syntax, and assumes that children acquire a second language "by a process of creative construction." Essentially language is viewed as a cognitive or problem-solving process. A distinctive feature of the BSM, then, is that it attempts to reflect several areas of language proficiency: Language dominance; structural proficiency of English and Spanish each as a second language; and degree of language maintenance for loss of Spanish as a first language. In providing data for placement and individual instructional program decisions, the BSM provides a rational model for assessing bilingualism. Again, the authors asserted:

> Bilingualism can be viewed as a continuum or, perhaps better, as a pair of sliding scales. The development of each of the child's two languages may be at any point along its scale depending on such factors as motivation, amount and type of exposure to the language, opportunities for use of the language and the learning environment.

The BSM does not then assume its measure as the equivalant of the student's actual dominance, but as an approximation in one area. The student is perceived not as being bilingual, but as becoming bilingual: "learning a new language and maintaining proficiency in the native language." This perspective is carried into the BSM scoring system which generates four categories:

1. Proficient
2. Intermediate
3. Survival
4. Non-speaking

0 points - no structure (two child)
1 point - misformed structure (two childs)
2 points - correct structure (two children)

Thus, the misforming or over generalization of rules is tabulated as language proficiency in progress. In addition, "non-speaking" levels of language development in which students understand some vocabulary and syntax in a target (non-native) language but can't produce it, is also documented via the BSM. The SAI ranges include:

95-100	Proficient
85-94	Intermediate
45-84	Survival
below 44	Non-speaking

As mentioned before, each language elicited can also be measured as

a heirarchical sequence of structures. The heirarchical score and SAI correspond with well, and complement each other.

The BSM is construct validity is well documented in psycho-linguistic research and it's heirarchical focue even spans research in speech diagnosis. Criterion validity is approached by attempts to match English and Spanish scores on the BSM with length of time in the United States. Further validity data are required.

James' Language Dominance Test

The James' instrument was developed, as was the BSM, in response to local school district requests for an instrument to evaluate language competence of kindergarten and first-grade Mexican-American children. It was later commercially published. It consists of forty visual stumuli to elicit both oral and aural performances. The administration takes 7-10 minutes, and can be conducted by para-professional as well as professional staff.

The instrument distinguishes between the following five levels of language dominance:

1) Spanish dominant in both comprehension and production
2) Bilinugal with Spanish as the home language
3) Bilingual with both English and Spanish as home languages
4) English dominant but bilingual comprehension
5) English dominant in both comprehension and production.

The distinction between levels 2 and 3 is unique for this mode of dominance testing, and is derived from a selection of vocabulary items which distinguishes specific home "knowledge" of a language from knowledge obtained in the school or community. The validity of this distinction was established through home interviews. This attempt at incorporating "domain" concepts into a reception/production instrument is reminiscent of the Fishman (1971) domain specific research.

The universe of items used here to assess dominance is oriented primarily towards competency in vocabulary comprehension in a limited semantic and structural space:

"Where is the bottle?"
"Which boy is drinking?"
"Where is the lightening?"

and in vocabulary production:

"What is this (car...hose...chair)?"
"What is the girl doing (sleeping...crying)?"
"What's happening here (the girls are playing with dolls.)"

Only two items in English and two in Spanish tap more elaborate syntactic or relational performance. In addition, because Spanish and English forms are not alternate forms in this mode of testing the practice effect may well influence dominance outcomes.

The scoring of the James is not weighted. There is a contingency for dialect differences if more than 50% of the children miss the same comprehension item, or if more than 10% defer an alternate

production form. Data on content and criterion-related validity are presented. Content validity is purportedly based on the author's selection of familiar materials from children's school and home environments. Predictive validity was based on follow-up check of first-grade children previously assessed with instrument. Teachers confirmed that placement decisions were appropriate. Test-retest administration with an alternative version demonstrated 98% agreement with previous classifications.

Norms are available for English production for children aged 3 to 6. Comparisons between monolingual English speaking students' scores with each of the five dominance categories demonstrated clearly defined differences. This serves cross validation for the dominance assessment instrument, but only in relation to English. Spanish monolingual students were not part of the norming procedure for the Spanish items. No norms are presented for comprehension in English or Spanish.

New York City's Language Assessment Battery (LAB)

In response to the Aspira consent decree, the New York City Board of Education commissioned the development of a comprehensive battery of oral/aural dominance assessment instruments for kindergarten through high school. The instruments were developed with equivalent, though not parallel, forms in Spanish and English. Three instrument difficulty levels are identified by grade: Level I, kindergarten through grade two; Level II, grade three through grade six; and Level III, grades seven through twelve. Within each level tests are utilized to elicit dominance in four modalities of language experience: Listening, Reading, Writing and Speaking. Although we are here only considering listening and speaking, it is worth noting that literacy dominance becomes a part of the total assessment, as was found in the Fishman (1971) procedures. We, thus, have twenty-four sub-tests within the battery.

	Level I	Level II	Level III
grade:	K-2	3-6	7-12
English listening			
Spanish listening			
English speaking			
Spanish speaking			
English reading			
Spanish reading			
English writing			
Spanish writing			

(figure 2)

The total battery testing time for each grade level is approximately 40 minutes. All subtests, with the exception of Speaking which is an individualized test, can be administered in large or small groupings by a teacher. The total test can be machine scored.

For grades 3 through 12 (levels II, III), the listening test is in two parts: (1) comprehension or receptive vocabulary using pictorial cues and (2) phonological analysis using sets of contrastive word pairs. Below are examples from each level.

Level II

Comprehension (4) cone
(11) basement
(16) diagonal

Phonology	(21) call-cool law-low saw-sow (23) gam-jam jello-jello jot-yacht (29) sheep-cheap dish-dish chop-shop
Level III	
Comprehension	(4) spider (8) nasal (17) corpulent
Phonology	(21) jest-jest yam-jam yello-jello jot-yacht (26) sing-sin hands-hands run-rung bang-bank (30) doze-those day-they there-dare then-then

Directions for both levels are the same in the comprehensive part, students are to mark the answer box which corresponds to the picture which represents the word spoken by the examiner. In phonology, students must choose the pair of words that sound the same from three or four pairs orally presented, in each item, by the examiner. Sample items are provided in both parts for practice.

The Speaking test for levels II and III is identical. With pictorial stimuli, the examiner elicits specific grammatical structures by employing two strategies: Probing questions and completion statements. The probing strategy is used first to give an opportunity to the student to produce the target grammatical markers spontaneously. If, however, the student replies incorrectly or not at all, the examiner may then use a more formal completion (or closure) technique to elicit linguistic performance. Below are two samples of how the procedure works. Note, however, that over half of the Speaking items have provided no probe questions.

Statement	Probe	Completion
(3) The dog is ripping the coat. (Point to the first picture.)	What did the dog do to this coat? (point to second picture)	This is the coat that the dog _____. (point to the second picture)
(12) Sally has a cold and can't go out. (point to first picture)	NO PROBE	She asks, "Mother _____."

Sample Correct Responses
(3) ripped, ate, etc.
(12) "Do I have to stay home?" or any appropriate question.

The grammatical targets tested range from tense markers to possessives. Level I for primary-age pupils tests both Speaking and Listening in a single integrative task which is set up like a dialogue with picture cues to support questions such as:

I would like to ask you a question?
Are you ready? (point to picture)

Why is the boy running after the dog? Here are two colors. Which is red?

Items elicit personal data, such as age and grade as well as asking the child to identify productively and receptively parts of the body, colors, familiar objects (dish, clock, truck). Three items ask the child to demonstrate comprehension of more complex syntactic-semantic relationships (e.g., touch the picture which shows what I say: the puppy is following the girl).

In comparison to other parallel dominance assessment formats discussed, the LAB Speaking test shows similarity to the BSM in elicitation strategy, while the LAB Listening test show considerably more consistency in receptive vocabulary than Shutt's Primary Language Indicator Test, and more range than the James' receptive language task.

Although the LAB was custom-designed for New York City Puerto Rican students, its comprehensiveness is an interesting model for dominance assessment.

Standardization procedures for the LAB are extensive. Monolingual English (n=12, 523) and Spanish (n=6,721) speaking students, respectively, in kindergarten through the twelfth grade provided data for norming. Students were selectively sampled based on total percent of English or Spanish speaking students in their schools, and according to literacy in their language (based on average N.Y.C. Reading test scores). Both Spring and Fall norms were generated, providing percentile scores for each grade level.

Reliability (using Kuder Richardson formula) is in the .90's for levels II and III English and Spanish, and in the .80's for level I English and Spanish. Validity data are presently being collected, and an equivalency study between Spanish and English forms is also in progress.

Finally, the LAB has been used as a direct placement instrument with specific criteria. Students scoring in the 20th percentile or below in the English version are administered in the Spanish version. If scores on the Spanish test are above the scores in English, students are eligible to participate in consent decree bilingual instruction. While the 20th percentile is an arbitrary cut-off point, the practical usability of the LAB assessment procedure is evident in school district instructional decisions.

Figure 3, Parts A, B, and C, outlines criteria by which each of the instruments discussed above might be described. Figure 3, Part A deals with Examinee Factors in terms of both developmental and cultural appropriateness. Figure 3, Part B deals with administrative and logistic factors. Figure 3, Part C presents psychometric considerations by which one might evaluate a language dominance instrument.

Figure 4, Parts A, B, and C presents a sample evaluation using the BSM. Note that one must review the numerical rating for each set of factors separately in order to construct a meaningful profile of the utility and appropriateness of any instrument. A total score will not give accurate information since one very low rating among a particular factor or component could invalidate the usefulness of an instrument

under a specific set of circumstances. Also, there is no true continuity of criteria from component to component. Thus Figures 3, A-B-C do not present a model for quantitative evaluation of dominance assessment instruments in any summative sense. The model suggested is rather one which provides for categorizing and weighing specific types of criteria in order to match them to the needs of a specific population and situation. Figures 3 and 4 are useful tools only when an educational system has assessed its own dominance testing needs first and then seeks instrumentation.

Conclusion

Benjamin Whorf (1956) conveyed his linguistic barbs into the ethnocentric preservation of English as the motherlode of logic when he stated:

> Language, for all its kingly role, is in some sense a superficial embroidery upon deeper processes of consciousness which are necessary before any communication, signalry, or symbolism whatsoever can occur...it may even be in the cards that there is no such thing as "Language" (with a capital "L") at all. The statement that "thinking is a matter of LANGUAGE" is an incorrect generalization of the more nearly correct idea that "thinking is a matter of different tongues."

If it takes statutory changes or compliance decrees to erode the ethnocentric abuse of minority language groups, so be it. Yet, the activities in the field of language dominance assessment seems to indicate that there are concerned professionals who prefer to cultivate a readiness for the complex and profound changes which must occur to redistribute, equitably, educational opportunity for all our students.

FIGURE 3-A, B, C

Figure 3-A

IDENTIFYING INFORMATION

NAME OF TEST: _____ GRADE/AGE: _____

PUBLISHER: _____ COSTS: _____

LANGUAGES: _____ ADMINISTRATION: _____

EXAMINEE FACTORS

Age/Grade Appropriateness *Cultural Appropriateness*
(2=completely, 1=moderately, 0=inappropriate)

CONTENT

DIRECTIONS

ITEMS

FORMAT

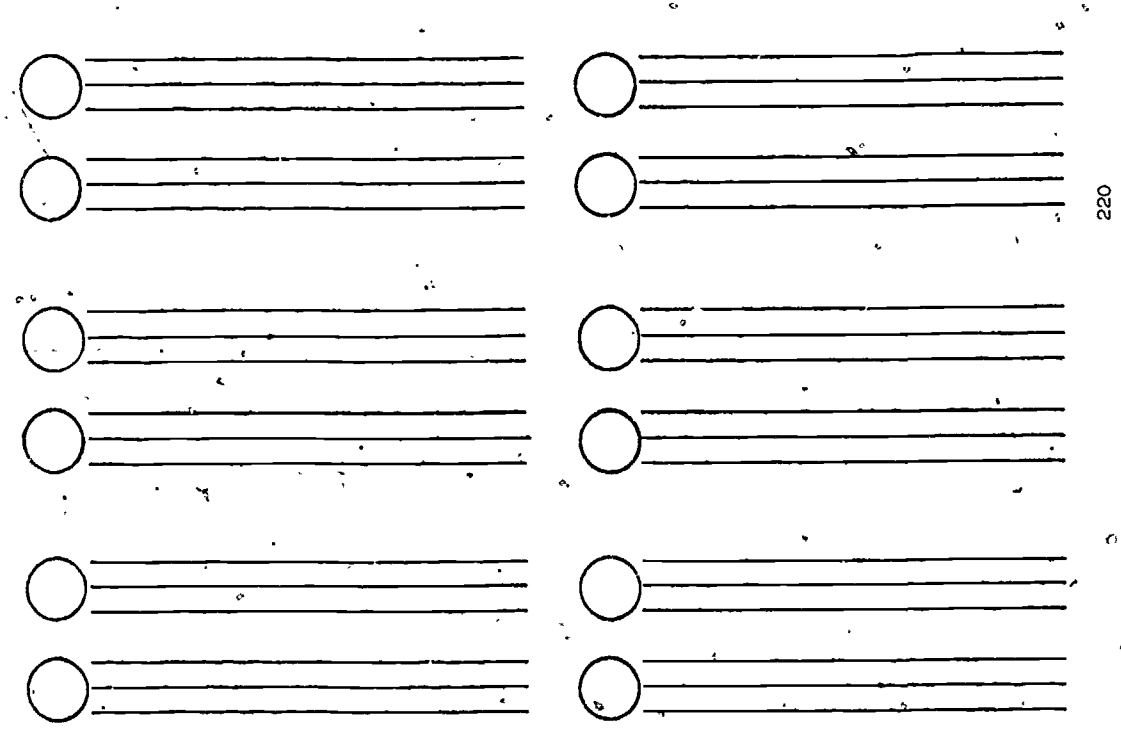
LAYOUT

ILLUSTRATIONS

PROCEDURE

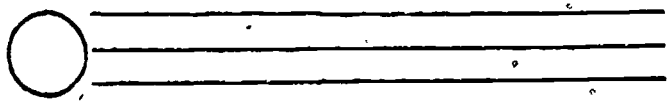
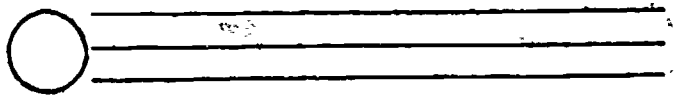
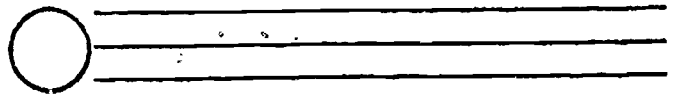
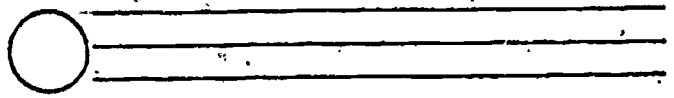
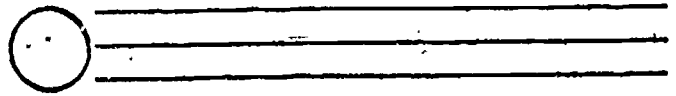
TIMING

RESPONSE MODE



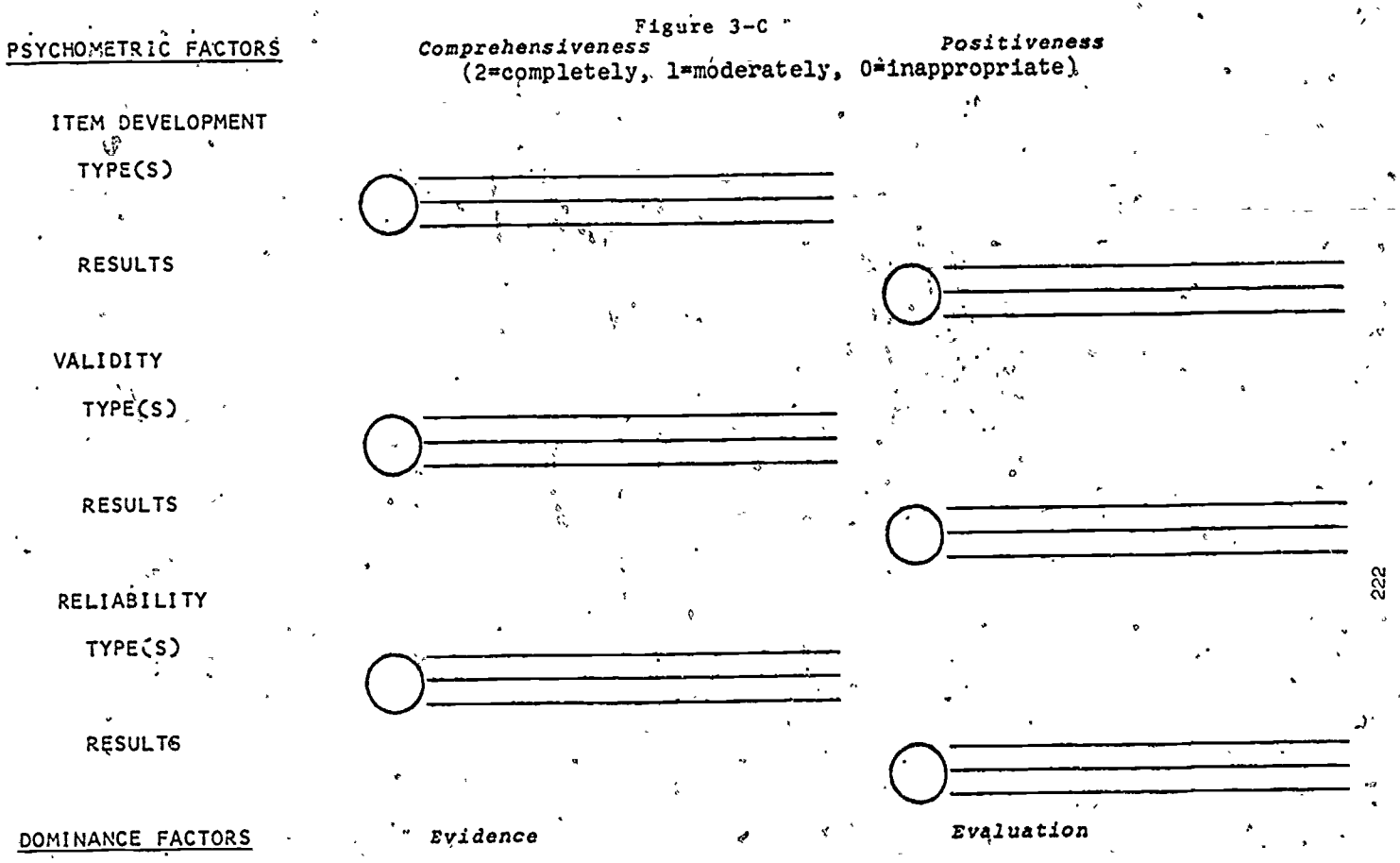
220

Figure 3-B

ADMINISTRATIVE FACTORS	*Practicality* (2=completely, 1=moderately, 0=inappropriate)	*Productiveness*
TEST QUALIFICATIONS		
GENERAL		
TEST		
SETTING		
SIZE OF GROUP		
MATERIALS/EQUIPMENT		
SCORING		
MANNER		
RESULTS		

221

Figure 3-C



222

FIGURE 4-A, B, C

IDENTIFYING INFORMATION

Figure 4A

NAME OF TEST: BILINGUAL SYNTAX MEASURE

PUBLISHER: HARCOURT, BRACE, JOVANOVICH

LANGUAGES: ENGLISH, SPANISH

GRADE/AGE: GR. K-3

COSTS: $1.50 / FIRST 35 STUDENTS
.75 THEREAFTER

ADMINISTRATION: INDIV. - 15 MIN. PER LANG.

EXAMINEE FACTORS	Age/Grade Appropriateness (2=completely, 1=moderately, 0=inappropriate)	Cultural Appropriateness
CONTENT		
DIRECTIONS	(2) SIMPLE, CLEAR, THOUGH SOME LEFT TO EXAMINER'S DISCRETION	(2) SAME
ITEMS	(2) SCREENED VOCABULARY BUT SOME STILTED FORMS (e.g., "SKINNY HOUSE")	(2) SCREENED BUT OCCASIONALLY STRAINED (e.g., "trepeador")
FORMAT		
LAYOUT	(2) SPACIOUS & STIMULATING	(2) SAME
ILLUSTRATIONS	(2) LARGE, COLORFUL CARTOON TYPE	(2) NEUTRAL RATHER THAN CUSTOMIZED
PROCEDURE		
TIMING	(1) UNSPECIFIED PACING, UNDERESTIMATED DURATION	(2) NOT COMPETITIVE NOR RUSHED
RESPONSE MODE	(2) OPEN FOR APPROPRIATE, RATHER THAN JUST CORRECT LANGUAGE	(2) NOT DIALECT-BOUND

223

Figure 4-B

ADMINISTRATIVE FACTORS	*Practicality* (2=completely, 1=moderately, 0=inappropriate)	*Productiveness*
TEST QUALIFICATIONS		
GENERAL	(2) PARA PROFESSIONALS, TEACHERS or SPECIALISTS	
TEST	(1) 1-2 HOURS OF TRAINING plus FLUENCY IN LANG. OF ADMINISTRATION	
SETTING		
SIZE OF GROUP	(0) INDIVIDUALLY ADMINISTERED	
MATERIALS/EQUIPMENT	(2) JUST TEST BOOKLET AND QUIET CORNER	
SCORING		
MANNER	(2) SIMPLE CODING SYSTEM, BUT DETAILING TECHNICAL INFO	
RESULTS		(2) APPLICABLE SCREENING GRID PLUS SOME NORMATIVE DATA

224

Figure 4-C

PSYCHOMETRIC FACTORS	*Comprehensiveness* (2=completely, 1=moderately, 0=inappropriate)	*Positiveness*
ITEM DEVELOPMENT		
TYPE(S)	(1) GENERAL RESEARCH EVIDENCE, FIELD STUDY OF 2,000 STUDENTS IN GRADES K-2	
RESULTS		(1) NO FIELD STUDY IN GR. 3 - ITEMS LARGELY LEFT UNCHANGED
VALIDITY		
TYPE(S)	(1) CONSTRUCT (PSYCHOLINGUISTIC THEORY & RESEARCH) PLUS CRITERION (REL'N TO LENGTH OF TIME IN U.S.)	
RESULTS		(1) SUPPORTIVE, BUT MORE APPLICABLE TO PROFICIENCY THAN DOMINANCE
RELIABILITY		
TYPE(S)	(1) TEST-RETEST (BASED ON CHILDREN - GR. K-2 IN CALIF.)	
RESULTS		(1) KAPPA COEFFICIENTS OF .50 - .68

DOMINANCE FACTORS	*Evidence*	*Evaluation*
LINGUISTIC EQUIVALENCE	TRANSLATION WITH SOME ADAPTATION AND ADDITIONS IN SPANISH.	1
PRACTICE EFFECT	SOME PICTORIAL STIMULI AND SEVERAL IDENTICAL ITEMS	1
PROFICIENCY INFORMATION	PRIMARY PURPOSE OF THE TEST	2