DOCUMENT RESUME

ED 226 597 FL 013 504

TITLE Children's English and Services Study: Estimates of Limited English Proficient Children in the United States. Methodological Reviews.

INSTITUTION InterAmerica Research Associates, Rosslyn, Va.
SPONS AGENCY National Inst. of Education (ED), Washington, DC. Teaching and Learning Program.
REPORT NO ISBN-089763-072-6
PUB DATE 82
CONTRACT NIE-400-80-0040
NOTE 61p.; For related document, see FL 013 505.
AVAILABLE FROM InterAmerica Research Associates, Inc., National Clearinghouse for Bilingual Education, 1555 Wilson Blvd., Suite 600, Rosslyn, VA 22209 ($11.50).
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS Adolescents; Children; Data Analysis; Educational Needs; Elementary Education; *Language Proficiency; Language Tests; *Limited English Speaking; Research Design; *Research Methodology

ABSTRACT

Two methodological reviews of the Children's English and Services Study (CESS), a congressionally mandated estimate of the number and needs of limited English proficient children, are presented. Such reviews are needed because the CESS is the only national study of its kind which administered a language proficiency test to determine the language proficiency of language minority groups. The reviews were conducted by the National Center for Education Statistics and the National Institute of Education. They include discussions of design and methodological issues and statistical tables. The major methodological issues reviewed include selection of items for the Language Measurement and Assessment Inventory, establishment of test cut-off scores for classifying children as limited English proficient, and the effects of non-response bias on estimates of the number of limited English proficient children. (RW)

Children's English and Services Study:

Estimates of Limited English Proficient Children in the United States—

Methodological Reviews

National Institute of Education
Teaching and Learning Program
Reading and Languages Studies Division

This document is published by InterAmerica Research Associates, Inc., pursuant to contract NIE 400-80-0040 to operate the National Clearinghouse for Bilingual Education. The National Clearinghouse for Bilingual Education is jointly funded by the National Institute of Education and the Office of Bilingual Education and Minority Languages Affairs, U.S. Department of Education. Contractors undertaking such projects under government sponsorship are encouraged to express their judgment freely in professional and technical matters; the views expressed in this publication do not necessarily reflect the views of the sponsoring agencies.

InterAmerica Research Associates, Inc. d/b/a
National Clearinghouse for Bilingual Education
1555 Wilson Boulevard, Suite 600
Rosslyn, Virginia 22209

ISBN: 089763-072-6
First printing 1982
Printed in USA

10 9 8 7 6 5 4 3 2 1

Contents

List of Tables

Foreword

The methodological review of the Children's English and Services Study (CESS) and a response to the review's estimation procedures discuss issues concerning the state of the art in language proficiency assessment in general and analytical issues regarding the Language Measurement and Assessment Inventory, developed for the CESS, in particular. Providing a single source for these documents, which are no longer available from the National Center for Education Statistics and the National Institute of Education, this volume is published in conjunction with two other reports that present results of the CESS: *Language Minority Children with Limited English Proficiency in the United States* and *Educational Needs Assessment for Language Minority Children with Limited English Proficiency*. Both of these reports, written by J. Michael O'Malley, are also available from the National Clearinghouse for Bilingual Education.

A national investigation of the number and educational needs of language minority students with limited English proficiency and of the instructional services provided to them, the CESS represents the first time that a number of important methodological refinements over previous investigations converged. In addition to the methodological reviews, the present volume contains a brief discussion of the study design issues, sample representation, defining limited English proficiency, and determining a criterion cut-off score, followed by a summary of the major methodological issues.

One of the activities of the National Clearinghouse for Bilingual Education is to publish documents addressing the specific information needs of the bilingual education community. We are pleased to make this title available through our growing list of publications. Subsequent Clearinghouse products will similarly seek to contribute information that can assist in the education of minority language and culture groups in the United States.

National Clearinghouse
for Bilingual Education

Children's English and Services Study:

Estimates of Limited English Proficient Children in the United States—

Methodological Reviews

I. Introduction

As part of the 1974 (ESEA Title VII) bilingual education legislation, Congress mandated a count or estimate of the number of children who were limited in English proficiency (LEP) and a determination of their basic educational needs. The Children's English and Services Study (CESS) of 1978 was the response to this mandate. The CESS was conducted for the National Center for Education Statistics (NCES) and the National Institute of Education (NIE) by L. Miranda and Associates of Washington, D.C. as primary contractor.[1]

Information concerning the number of limited English proficient children is needed in order to establish projections concerning the appropriate levels of services and to determine the probable impact of bilingual education. Because estimates of the number of LEP children are so important and because the CESS is the only national study of its kind that attempts to provide such estimates, information concerning the methodology of the study must be made available. Additionally, the CESS is the only national study of the language proficiency of language minority groups in which a language proficiency test was administered, and its results have been used in subsequent studies, including one estimating the number of LEP children in projections to the year 2000.[2]

The purpose of the present document is to give information on the methodological issues involved in the CESS. The volume consists of a review of the CESS by the National Center for Education Statistics (NCES) and a reanalysis by the National Institute of Education (NIE). They are, respectively, "The Children's English and Services Study: A Methodological Review" and "Reanalysis of the Number of Limited English Proficient Students Estimated in the Children's English and Services Study." In order to place the two reviews in the appropriate context and to aid in understanding the methodological issues of the CESS project, the methodological reviews are preceded by a brief background of the CESS and a summary of the main methodological considerations.

Daniel M. Ulibarri
Research Associate
National Institute of Education

Notes

1. J. Michael O'Malley, *Children's English and Services Study, Spring 1978* (Rosslyn, Va.: National Clearinghouse for Bilingual Education, 1981).

2. Rebecca Oxford et al., *Projections of Non-English Language Background and Limited English Proficient Persons in the United States to the Year 2000,* report prepared for the National Center for Education Statistics (NCES) by InterAmerica Research Associates, Inc., Contract OE 300-79-0737 (Rosslyn, Va.: InterAmerica Research Associates, 1980).

II. The Children's English and Services Study (CESS)

The Children's English and Services Study (CESS) was a direct result of a Congressional mandate as stated in the Bilingual Education Act, as amended in 1974 (Title VII of the Elementary and Secondary Act). The purpose of the CESS was to determine the number of children of limited English proficiency (LEP) aged 5 to 14 living in households where a language other than English was spoken. In addition, the CESS was designed to provide estimates of the number of limited English speaking children from Spanish language backgrounds and the aggregate of all other language minorities combined in major geographic regions of the United States. The results indicate that in Spring 1978 there were 2.4 million LEP children in the 5 to 14 age range and an estimated 3.6 million in the U.S. school-age population (4 to 18 years old). For the age-range 5 to 14 years there were 1.7 million Hispanic LEP children and .7 million LEP children in all other language minority backgrounds combined.

Study Design Issues

Since the Congressional mandate for a count of limited English speaking children called for information that no previous investigation had attempted to produce at the national level, a new study design had to be developed. The major problems this design had to deal with were:

- *Identifying a representative sample.* A procedure was required for developing a probability sample of the population of language minority persons.
- *Defining limited English speaking ability.* Acceptable criteria or test instruments for identifying limited English speaking children as defined in ESEA Title VII were required that could be used in a household survey.
- *Determining a criterion score.* Acceptable techniques were needed for identifying a criterion score below which all students could be accurately classified as limited English speaking.

Sample Representation

The CESS was based on a sample of households in the United States in which a language other than English was spoken usually or often. Adults were interviewed in the Spring of 1978 in a randomly selected, nationally representative probability sample of 35,000 households. Approximately 2,200 households among these 35,000 were identified where a language other than English was spoken and where children between the ages of 5 and 14 were living. Within these 2,200 households, selected children (up to 2 children per household) were individually administered a test in English that determined whether or not they were limited in English language proficiency. The sample was designed to be representative of the number of children in California, Texas, New York, and the remainder of the country.

Defining Limited English Proficiency

The external criterion that served as the operational definition of limited English proficiency and that was used to measure English proficiency was a specially constructed language assessment instrument. Specific needs and constraints that guided the development of an acceptable criterion to be used for defining limited English proficiency were the following:

1. The definition of limited English speaking ability had to be *consistent with the Title VII legislative definition* stated in the Bilingual Education Act, as amended in 1974. The legislation pointed to two requirements for defining limited English proficiency: non-English language background, defined in terms of native language, country of origin, or language environment, and limited ability in speaking and understanding English to the extent of not being able to profit from instruction in English.

2. The language proficiency criterion had to *produce a dichotomous classification* of limited versus not limited English speaking ability.

3. A criterion had to be adopted that would be *usable* to define limited English speaking ability *in the context of a household survey* for children aged 5-14 years.

Since no existing instrument met the needs and constraints of the study or satisfactorily identified limited English proficiency consistent with the ESEA Title VII definition, a new instrument was developed, the Language Measurement and Assessment Inventory (LM&AI). (A summary is provided as Appendix A to the second methodological review in this volume.) The test measures age-specific speaking, listening, reading, and writing skills in English, and was designed to meet the definition of limited English proficiency in the Bilingual Education Act. Representatives from 30 State Education Agencies developed the test specifications and served on a review team for the study.

The LM&AI was developed to differentiate language minorities who were limited in English from those who could profit from instruction in English. In developing the instruments, limited English proficient children and fluent English speaking children who were benefiting from instruction in English were identified by local school personnel as the criterion groups.

The test criterion for limited English proficiency is a cut-off point on the total score that most accurately classifies children as limited in English for their age level. Children in the national sample whose score fell below the critical value were classified as limited English proficient and those whose score fell above the critical value were classified as fluent English proficient. The criterion for determining a cut-off score was derived from a field study described briefly in the following section.

Determining a Criterion Cut-off Score

In selecting an approach to identify a cut-off score appropriate for each age group, an attempt was made to maximize the likelihood of correct classifications of children who were limited and those who were fluent in English proficiency. The approach selected was **discriminant function analysis.** Discriminant function analysis is similar methodologically to regression analysis when the criterion variable is dichotomous. Thus, in discriminant function analysis one or more variables are used to predict group membership in one of two groups (in this case LEP and non-LEP). In the CESS, the predictor was the continuous score on the English language instrument and the criterion was the school's designation as limited or fluent in English speaking ability. The discriminant function maximizes the overall number of accurate classifications such that correctly classifying fluent English speaking children was given as much weight as correctly classifying children who were limited in English proficiency.

Cut-off scores were determined on an independent sample of students obtained in a field test of the instrument. Accurate classifications in predicting language ability occur when a child is classified by both the predictor and criterion as limited English proficient, or as fluent in English. The ratio of the number of correct classifications to the total number of cases yields an accuracy rate. For the field test, the overall accuracy rate ranged from 82.9% to 97.2%, depending on the age level of the children tested.

III. Major Methodological Issues

In January 1980, the National Center for Education Statistics, Office of Research and Analysis (NCES/ORA), U.S. Department of Education issued a report entitled "Analytical Issues Regarding the Children's English and Services Study (AI/CESS)" (see Appendix A of NCES/ORA Methodological Review, p. 23 of this volume). The AI/CESS contained a discussion of three analytical issues that were identified upon review of the CESS Draft Report of September 1979, and a later revision dated November 1979. In the AI/CESS report a call was made for responses to the issues (Appendix B to the NIE reanalysis contains L. Miranda and Associates' response to the AI/CESS report, page 47 of this volume). As a result of the responses, NCES/ORA prepared a methodological review of the CESS in August 1980. In the review, NCES/ORA discussed three analytical issues:

1. Were the items selected for inclusion in the Language Measurement and Assessment Inventory (LM&AI) selected properly?

2. Were the cut-off scores for the LM&AI, which were determined and used to classify children as either English proficient or of limited English proficiency (LEP), set properly?

3. What were the effects of non-response bias on the counts and estimates of the number of LEP children?

In response to the NCES/ORA review, NIE prepared a report comparing the results of the CESS methodology for determining cut-off scores with the alternative procedure recommended in the NCES/ORA review. While the NCES comments focused on three issues, the NIE response addressed only the second issue.

The other two issues were adequately discussed in the NCES document and further comment by NIE was not necessary. The following section contains the NCES/ORA Methodological Review of the CESS report and NIE's response to the issue of determining cut-off scores.

Methodological Reviews

1. The Children's English and Services Study: A Methodological Review

by David D. Dubois
Education Policy Fellow
National Center for Education Statistics
August 1980

Introduction

On January 21, 1980, the Office of Research and Analysis (ORA) of the National Center for Education Statistics (NCES), U.S. Department of Education issued a report entitled "Analytical Issues Regarding the Children's English and Services Study" (AI/CESS). The purpose of the AI/CESS report was:

> . . . to present and discuss three analytical issues which have been identified as a result of a post hoc assessment of the research design, data analyses, and other information which are described in the 1978 Children's English and Services Study (CESS) Draft Report of September 6, 1979 (and a later revision dated November 1979).[1]

A copy of the AI/CESS report is found in Appendix A, page 23 of this volume.

The objective of the present NCES/ORA inquiry is best summarized by the following passage from the January 1980 report:

> Since the results of the 1978 CESS are of tremendous importance to present and future research studies, bilingual [education] program and policy development, and funding for bilingual education, unresolved analytical issues which could adversely affect the validity of the results are being stated with the hope of their resolution.[2]

Secondary analyses of data and research designs frequently reveal analysis errors or areas of skepticism in the design. Sterling and Weinkam (1979), who discovered misclassification in a study of mortality among U.S. veterans, describe the potential response of managers to this discovery as either "cooperative" or "adversary."

In the former case, management tries to determine the source concern and to restructure the procedures or analyses. In the latter case, the response may be to eliminate the discovery of errors rather than their source. Regarding this problem, Sterling and Weinkam further observed that:

> . . . there may be underlying sociological and psychological forces operating which make it more acceptable for management to adopt an adversary rather than a cooperative stance even in scientific instances. From a simpleminded perspective, to acknowledge the existence of errors may require considerable efforts and expenditures to correct them, not to say anything of extracting accountability from some individuals who insist on bringing these errors to the attention of management as troublemakers.[3]

They continued by indicating that "as the value of using available data files for secondary analysis becomes increasingly clear (and there is a great deal of value in the use of properly collected and suitable data), we may expect that other discoveries similar to ours will be made."[4] They recommended that a mechanism be established to encourage secondary analysis for validating the appropriateness of preceding work and to distribute the results of these analyses.

NCES/ORA recognizes that many research reports are incomplete or inaccurate in one or more ways, and the CESS Draft Report mentioned earlier is no exception. Accordingly, NCES/ORA's purpose in this report is to clarify the issues and make specific recommendations for modifications to the NIE Draft Report.

The present report includes a discussion of the three issues presented in the AI/CESS report and is based, in part, upon the responses received from it. In addition, pertinent literature on language development and assessment, the relationship of language acquisition to cognition, and additional data analyses completed by NCES provide the bases for this paper.

Statement of the Issues

The following analytical issues are the subject of this position paper:

1. Were the items that were selected for inclusion in the Language Measurement and Assessment Inventory (LM&AI) selected properly?
2. Were the cut-off scores for the LM&AI, which were determined and used to classify children as either English proficient or of limited English proficiency (LEP), set properly?
3. What were the effects of non-response bias on the counts and estimates of the number of LEP children?

Discussion of the Issues and Recommendations

Issue: Were the items that were selected for inclusion in the Language Measurement and Assessment Inventory (LM&AI) selected properly?

This issue was restated as:

> Is English language proficiency the dimension on which the scores vary, or are other dimensions associated with variations in the scores?

Two subissues were posed, namely:

- Are the test scores related to language dominance?
- Are the test scores related to general language development?

Discussion

Lourdes Miranda, President of L. Miranda and Associates, the prime contractor for the CESS, responded to the AI/CESS report. In discussing the rationale for the test items selected for the LM&AI, Miranda noted that "it was essential for us to measure the ability [of language minority children] to successfully deal with academic classroom tasks that are often as clearly reliant on memory and cognitive abilities as on English language skills."[6] Therefore, "other dimensions [e.g., cognitive] are associated with score variation."[7]

The LM&AI was specifically designed to meet the definition of limited English proficiency found in the Bilingual Education Act, the 1965 Elementary and Secondary Act (ESEA), Section 703 (a) (1) (B), as amended. The 1978 amendment of the act expanded the language skill domains from "limited English speaking ability" to include speaking, reading, writing, and understanding English. By virtue of their limited English proficiency, Congress concluded, language minority children were denied the opportunity to attain levels comparable to others at their appropriate age and grade levels.

J. Michael O'Malley, the NIE project officer for CESS, responded to the first issue that:

> Because functioning in the classroom often requires conceptual skills as well as oral language and literacy, the inclusion of cognitive demands in the test items was seen as an acceptable approach for increasing the content and perhaps the predictive validity of the test.[8]

O'Malley also stated that:

> A "pure" measure of English proficiency could not have possessed the content validity required to identify language minority children who have difficulty profiting from instruction in English.[9]

Earlier in his response, O'Malley said, "simply stated, the test scores are predictors

of the ability to profit from English language instruction, which determines eligibility for ESEA Title VII."[10] He observed, "School decisions about eligibility for ESEA Title VII are often based on a child's general level of functioning in the classroom rather than on English language proficiency alone[11] . . . . The LM&AI used tested skills in English . . . to simulate the decisions schools would make in determining that language minority students could not profit from instruction in English."[12]

A review of recent literature in the areas of language assessment, linguistic and intellectual development, and bilingual education programs revealed that analytical questions in these areas have, for some time, presented serious intellectual challenges to researchers and educators. The issues raised in the AI/CESS report were presented within the framework of the CESS development process and with the knowledge that there are many unanswered basic research questions in the three areas mentioned above. NCES's purpose in this report is to clarify current thought on these issues.

The first subissue raised in the AI/CESS report was stated as: "Are the scores related to language dominance?"[13] O'Malley takes the following position:

> By exclusion in the [legislative] definition of eligibility, language dominance has no role in policy determination for ESEA Title VII eligibility. Thus, the statements in the NIE report on the CESS that language dominance was considered irrelevant [are] understandable.[14]

NCES/ORA believes there is justification for excluding a language dominance concept in the development of the CESS. Specifically, there does not *appear* to be agreement among linguists about an operational definition and, therefore, about the impact of "language dominance" upon the ability of language minority children to profit from instruction in English.

DeAvila and Duncan (1976) argue against using a "language dominance" concept when discussing school achievement. They ask, "How does the concept of [language] dominance clarify the relation between the child's linguistic development and school achievement in such a way that we can do something about it?"[15] They continue, "Another way of asking this question is by asking whether or not 'dominance' in or of itself determines what *is* learned or what *can* be learned."[16] Language dominance "does not address the real issue that the child might have language development problems in *both* languages—the native language and English."[17]

Some experts have argued that a language dominance concept is meaningful only when the use of a language is considered within a social or cultural context, such as home and family relationships, social interactions, an academic domain, a business environment, or within a religious context. The degree of fluency or level of language dominance is meaningful only when the purpose for which language is being used is also stated. In this sense, several "dominance" levels might be defined.

Regarding a child's possible difficulty with both languages, Dubois (1980) states: "Whether it is appropriate to assess English language proficiency, [while] ignoring the child's proficiency in another language, remains a policy question to be addressed."[18] More specifically, this is an empirical question. A recent article by Cummins (1979) addresses this question. In the following excerpt L_1 refers to a child's first language and L_2 refers to the second language. Cummins says:

> The lack of concern for the developmental interrelationships between language and thought in the bilingual child is one of the major reasons why evaluations and research have provided so little data on the dynamics of the bilingual child's interaction with [his or her] educational environment. A direct determinant of the quality of this interaction is clearly the level of L_1 and L_2 competence which the bilingual child develops over the course of [his or her] school career. . . . What level of L_2 competence must the child possess at various grade levels in order to benefit optimally from instruction in that language? . . . To what extent are L_1 and L_2 skills interdependent and what are the implications of possible interdependencies for cognitive and academic progress? In other words, do children who maintain and develop their L_1 in school develop higher or lower L_2 levels of skills than those whose L_1 is replaced by their L_2?[19]

Cummins provides research evidence for a developmental interdependence hypothesis which says that the level of L_2 competence that a bilingual child attains is partly a function of the type of competence the child has developed in L_1 at the time when intensive exposure to L_2 begins. In this sense, a measure of L_1 proficiency is important for policy decisions.

Lambert (1975) suggests that children exhibit either "additive bilingualism" or "subtractive bilingualism." A child's bilingualism is most likely to be additive when L_1 is prestigious or the dominant language and is, therefore, not in danger of being replaced by L_2. In this case, a bilingual child adds L_2 skills without losing L_1 skills. Subtractive bilingualism refers to the form children experience when their L_1 is eventually replaced by L_2, and generally occurs when the child's L_1 is a nonprestigious or minority language. Socioeconomic status also seems to be a factor that is related to whether a child's bilingualism is subtractive or additive. Children from upper or middle class socioeconomic strata, when given instruction in L_2, tend to experience additive bilingualism, while children from lower socioeconomic status tend to experience subtractive bilingualism. Troike (1980) hypothesizes that, for children from lower socioeconomic groups, cognitive development can become disrupted when children begin learning L_2 between the ages of 6-10. Socioeconomic status and sociopolitical status are, therefore, related to language and cognitive development.

The second subissue in the AI/CESS report was: "Are the test scores related to general language development?"[20] The concern was for the inclusion of test items with cognitive components on the LM&AI. Miranda noted earlier that the purpose of the LM&AI was to measure the ability of language minority children to deal successfully with academic classroom language skills. In reply to this subissue, Miranda stated that "it is difficult to imagine how a test of 'pure' linguistic competence could have been developed should we have been asked to do so."[21] DeAvila et al. (1979) observe that "much confusion abounds with respect to both the meaning and the measurement of English language proficiency."[22] Moreover, they note that "the role of language and cognition in general is itself not clearly agreed upon."[23] For the purpose of this discussion, cognition means the act or process of perceiving or knowing.

Cazden (1972) addresses two controversial items of interest in *Child Language and Education*. The first concerns whether a person's thought is affected by the particular language forms or speech patterns with which he or she is familiar. The second item concerns the question of which develops first, the nonverbal idea or the words to express it.[24] Essentially, this poses the central issue: Which develops first, language or cognition? Language experts, educational psychologists, and professionals in related fields apparently do not agree upon the proposed answers to this question.

Cazden's first item is based upon the Whorfian (1956) hypothesis that says that "language influences our perceptions of and responses to the world."[25] This leads us to believe that no learning can take place until language proficiency is attained; therefore, language determines cognition. Regarding Cazden's second item, Piaget indicates a child's cognitive development is the primary factor in language acquisition and development, with a later emphasis on a more balanced interaction between the two.

According to Piaget, cognition develops as a result of experience. He believes that although language contributes to further development, it is the use of language that is determined by development and not the converse.[26]

Cummins (1979), in a summary of research evidence on the role of language and cognitive development, concludes that:

> the level of competence bilingual children achieve in their two languages acts as an intervening variable in mediating the effects of their bilingual learning experiences on cognition. Specifically, there may be threshold levels of linguistic competence which bilingual children must attain both in order to avoid cognitive deficits and to allow the potentially beneficial aspects of

becoming bilingual to influence their cognitive growth."

DeAvila et al. (1979) state that:

> "Edmonds (1976) has recently argued that a full understanding of language acquisition will not emerge until the process is viewed within a larger developmental framework."[28] And, related to this, "Tremaine (1975) has examined 'syntax as an instance of operational intelligence' defined in the Piagetian sense. The results indicated that children at the operational level performed significantly better in terms of syntax comprehension than children classified as nonoperational."[29] Later, DeAvila interprets Tremaine's findings as follows: "What this means is that solutions which focus on English language deficits will be of limited success as long as developmental factors are not taken into account."[30]

Studies have focused on several of these complex relationships. DeAvila et al. (1979) examined the relationship between the degree of bilingualism (relative linguistic proficiency in English and Spanish), level of intellectual development (cognition), and performance on two tests of cognitive-perceptual functioning or field dependence independence.

DeAvila concludes that:

> In terms of educational implications, the most accurate and least value-laden interpretation of the findings would be to conclude that there seems to be a positive interaction between relative linguistic proficiency and cognitive perceptual functioning.[31]

In summary, Tucker (1979) of the Center for Applied Linguistics makes the following comments, with which the NCES/ORA agrees:

> Nor, in my opinion, have we managed to devise appropriate and valid instruments to assess language proficiency. What does it mean to know and to be able to communicate effectively and acceptably in a language? Does there exist some necessary (measurable) threshold of target language proficiency which must be attained before one is able to profit from instruction in that language? Obviously a great deal of additional interdisciplinary research is needed to examine the effects of factors such as intellectual potential, social status, physical or emotional development, age of entry, presence of native speakers, community stereotypes, teacher characteristics, classroom techniques, sequencing of languages, and social setting on the desirability and efficacy of bilingual education programs. I remain optimistic that the proposed Center for Bilingual Research may begin to move us in the right direction.[32]

Troike (1980) suggests that the effect of the density of a specific language minority group upon language proficiency in L_1 or L_2 is another factor that deserves additional research attention.

NCES/ORA cannot determine the effect of the cognitive components in the LM&AI on the test scores based upon the currently available information from discussions with experts in language development and assessment, and a review of pertinent literature. A post hoc study of the cognitive component could be completed using a sample of subjects from the population that was used for calibrating the LM&AI, but at additional cost to the Government. However, the quality of the results of such a study would probably not warrant the cost since tests of language proficiency are generally confounded with language and other factors.

Recommendation

NCES/ORA recommends that NIE state in the final CESS report the caveats found in the discussion of this issue. There are clearly limitations to the CESS results which are a function of the current state of the art in language proficiency assessment.

Issue: Were the cut-off (critical) scores for the LM&AI, which were determined and used to classify children as either English proficient or of limited English proficiency, set properly?

Discussion

The purpose of the LM&AI was to provide a mechanism for categorizing a child as being either English proficient or limited English proficient. Therefore, the critical score determined for each age-level test of the

LM&AI is essential for determining valid LEP counts. The critical score was that which best differentiated LEP children from fluent English speaking (FES) children who were clearly profiting from instruction in English. As an example, if the critical score on each age-level test is lowered by *two items,* the estimated count of LEP children decreases from 2.41 million to 2.13 million children, or a decrease of 280,000. Similarly, if the critical score for each age-level test is raised by two items, the estimated count of LEP children is increased from 2.41 to 2.62 million children, or an increase of 210,000. Thus, a score difference of four items has the effect of altering the count by nearly one-half million.

The NCES/ORA requested the raw data on the student scores from Field Test III—which were used to determine the critical scores from the LM&AI—from the prime contractor, L. Miranda and Associates, Inc. Based upon an examination of these raw data and a comparison of these findings with Table A-4 of the NIE Draft Report on the CESS, a discrepancy in the data of Table A-4 was called to the attention of the prime contractor. Miranda replied that the procedure used for determining the critical scores, based on a discriminant function analysis, was a modification (Grand Mean-Constant = Cut-off) of the more conventional approach and resulted in a more conservative estimate of the number of limited English proficient children (see Miranda, 1980, p. 6). However, the data in Table A-4 did not reflect this conservative approach. To remedy this situation, Miranda submitted a Revised Table A-4 for inclusion in the final NIE report on CESS. A copy of the table is in Appendix C, page 36 of this volume.

In developing the LM&AI, five techniques were proposed as alternatives for determining the critical scores. The five techniques[33] are summarized below.

1. For each age-level test determine the score on Field Test II data that was one standard deviation below the mean score for the fluent English speakers (FES) group of that age.
2. Similarly, use the score that was one standard deviation above the mean score for the limited English speaking ability (LESA—later revised to LEP) group of each age.
3. Use the highest Field Test III LESA (LEP) score made by any individual on each age-level test.
4. Plot the score of LESA and FES separately and select the score equivalent to the point of intersection of the two distributions.
5. Use discriminant function analysis (DFA) that considers subscores to determine a centroid, which can act as the critical point.

After examining the accuracy of the various alternatives, DFA was chosen as the method for determining the critical scores.

While NCES/ORA fully endorses the use of DFA, an issue remains regarding its use: *the application of DFA without concern for the differential "costs" of misclassification.* DFA is a powerful tool in that it minimizes that proportion of the sample that is misclassified. However, if the resulting classification criteria (critical scores) consistently misclassify one subgroup (e.g., LEP) at the expense of the other, a serious bias may result. More explicitly, if there are actually N_1 LEP children and N_2 English proficient (fluent) children among the $N = N_1 + N_2$ children of non-English language background households, then the cut-off score will lead to an unbiased classification procedure if and only if $N_2 \cdot Pr$ (Classified LEP | Actually fluent) = $N_1 \cdot Pr$ (Classified fluent | Actually LEP). *That is, the expected number of fluent children misclassified as LEP must equal the expected number of LEP children misclassified as fluent.*

In defense of the procedures used, since N_1 and N_2 were not known in advance, minimizing the overall misclassification error makes reasonable sense. However, as can be seen in Table A, the discrimination procedure actually used was much more likely to misclassify LEP children than fluent children. This explains why the critical scores for DFA seemed low.[34]

Table A

Estimated Conditional Probabilities of Correct and Incorrect Classifications by the LM&AI (See Appendix C, Revised Table A-4)

Age	P_{11}[a]	P_{12}[b]	P_{21}[c]	P_{22}[d]
5	0.811	0.000	0.189	1.000
6	0.795	0.000	0.205	1.000
7	0.806	0.000	0.194	1.000
8	0.893	0.000	0.107	1.000
9	0.813	0.000	0.188	1.000
10	0.833	0.000	0.167	1.000
11	0.682	0.000	0.318	1.000
12	0.864	0.182	0.136	0.818
13	0.800	0.000	0.200	1.000
14	0.879	0.204	0.121	0.796

a. P_{11} = Pr(Classified LEP | Actually LEP).
b. P_{12} = Pr(Classified LEP | Actually Fluent).
c. P_{21} = Pr(Classified Fluent | Actually LEP).
d. P_{22} = Pr(Classified Fluent | Actually Fluent).

Table A presents the estimated conditional probabilities of correct and incorrect classifications by the LM&AI for the critical scores found in the Revised Table A-4 (Appendix C).

The bias evident in Table A led NCES/ORA to conclude that the critical scores for each age-level test of the LM&AI should be revised in order to remove the estimated bias, once estimates of N_1 and N_2 have been computed. The mechanism by which this can be done follows:[35]

Let P_{11}, P_{12}, P_{21}, and P_{22} be defined as in Table A. Let N_1 and N_2 be the actual number of LEP and fluent children, respectively. Finally let L and F be the expected number of LEP and fluent children, respectively, as estimated by the LM&AI. Then,

$$L = N_1P_{11} + N_2P_{12}$$

$$F = N_1P_{21} + N_2P_{22}$$

Solving for N_1 and N_2,

$$N_1 = (LP_{22} - FP_{12})/(P_{11}P_{22} - P_{12}P_{21}) \text{ and}$$

$$N_2 = (FP_{11} + LP_{21})/(P_{11}P_{22} - P_{12}P_{21})$$

Of course, the values of P_{11}, P_{12}, P_{21}, and P_{22} are functions of the actual critical scores that are used for distinguishing between LEP and fluent children. This means that an iterative procedure must be used to determine the unbiased estimates of N_1 and N_2 based on critical scores associated with "balanced" misclassification errors. To accomplish this, the estimated "misclassification balance," defined by $|L \cdot P_{12} - F \cdot P_{12}|$, must be calculated for each possible critical score. For each age group, the critical score is selected that minimizes the estimated misclassification imbalance. Using the expected number of LEP (L) and fluent (F) children and the revised probabilities (Ps) once the expected misclassification imbalance has been *minimized,* we can approximate the "unbiased" values of N_1 and N_2 for each age group. The values of L, F, P_{11}, P_{21}, and P_{22} which were used to compute N_1 and N_2 are found in Appendix D.

The results of these computations (shown in Table B) clearly demonstrate the consistent bias in the LM&AI classification procedure. The CESS/LM&AI LEP counts underestimate the "true" values at every age, except for ages 12 and 14.

Table B

Effect on 1978 CESS LEP Counts of Removing the Estimated Bias

Age	1978 CESS LEP Count	Unbiased LEP Count
5	192,297	249,734
6	291,622	306,970
7	275,924	320,774
8	257,807	277,422
9	167,304	189,277
10	294,156	329,047
11	190,064	266,706
12	251,680	207,388
13	196,577	227,732
14	291,444	246,282
Total	2,408,875	2,621,332

In actual practice, the LEP counts determined by the critical score will almost always differ from the "unbiased" estimate, since all children with a given score must fall on one side of the critical score or the other. Therefore, some bias must be accepted in the counts, but NCES/ORA has minimized the expected bias by using the procedure just described. Table C contains the CESS Draft Report critical scores (with the slight modification mentioned earlier), the revised critical scores, and the resulting LEP count for each age level. Note that the national LEP figure of 2,631,075 children (Table C) compares to an "unbiased" estimate of 2,621,332 (Table B).

By minimizing the estimated bias, a less conservative, yet more analytically sound, LEP count results with a change in the national CESS estimate from 2,408,875 LEP children to 2,631,075. This change represents a national increase of 9.22% in the number of LEP children estimated in the CESS Draft Report.

Table C

Revised Critical Scores and Resulting LEP Counts

Age	CESS Draft Report Critical Score	Revised Critical Score	Revised LEP Count
5	18.5	25.5	254,657
6	26.5	29.5	303,584
7	39.5	44.5	318,470
8	38.5	40.5	280,256
9	43.5	46.5	188,187
10	49.5	52.5	330,979
11	41.5	51.5	271,485
12	46.5	44.5	208,426
13	48.5	52.5	229,986
14	52.5	49.5	245,045
Total	—	—	2,631,075

Recommendation

NCES/ORA recommends that the NIE final report on CESS reflect this analysis and the revised LEP counts found in Table C.

Issue: What were the effects of non-response bias on the counts and estimate of the number of LEP children?

Discussion

The question is whether non-respondents are similar to or different from respondents to the study. There is no evidence in the NIE Draft Report of November 1979 to indicate that non-response bias was empirically investigated.

Donald Rogers, Vice President for Operations with Resource Development Institute (one of the subcontractors), completed a brief paper in response to AI/CESS. In his paper, Rogers presents "the results of a very, very simple analysis of the effects of non-response during the CESS study."[36] (See Appendix B, page 32 of this volume.)

Rogers states in a letter that accompanied his paper that:

> My assumptions [Appendix B] generated a weighted LESA [LEP] total that fell within the 95% confidence interval for the total weighted U.S. LESA [LEP] count reported by the CESS study. I do not believe that a study of non-respondents will greatly increase or decrease the total, weighted U.S. LESA [LEP] count.[37]

Recommendation

NCES/ORA concurs with Rogers' position that further investigations of non-response bias associated with the 1978 CESS are not warranted.

Notes

1. David D. Dubois, "Analytical Issues Regarding the *Children's English and Services Study*," photocopied material (Washington, D.C.: U.S. Department of Education, National Center for Education Statistics, 1980), p. 1; also see this volume.
2. Ibid., pp. 2-3.
3. T.D. Sterling and J.J. Weinkam, "What Happens When Major Errors are Discovered Long After an Important Report Has Been Published?" (paper presented to the American Statistical Association Annual Meeting in Washington, D.C. on August 16, 1979), photocopied material (Burnaby, B.C.: Simon Fraser University, 1979), p. 2.
4. Ibid., p. 12.
5. Rudolph C. Troike (in a personal communication with David D. Dubois, May 30, 1980) suggests that a restatement of this issue should not detract attention from the fact that there is a dearth of basic research on the question of which types of items are appropriate for language assessment and measurement at each age level. For example, it is necessary to examine the range of grammatical or semantic variations that are tolerable for each test item at each age level. Only after examining this question and others, says Troike, can we hope to be confident of obtaining reliable and valid measures of language proficiency.
6. L. Miranda and Associates, Inc., "Response to the Office of Research and Analysis of the National Center for Education Statistics' Inquiry on Three Analytical Issues Associated with the 1978 Children's English and Services Study," (Bethesda, Md.: L. Miranda and Associates, 1980), p. 2; also see this volume.
7. Ibid., p. 2.
8. J. Michael O'Malley, response to AI/CESS in the form of a letter dated January 31, 1980 at Alexandria, Virginia, p. 2.
9. Ibid.
10. Ibid., p. 1.
11. Ibid., p. 2.

12. Ibid.
13. Dubois, 1980, p. 6.
14. O'Malley, 1980, p. 3.
15. Edward DeAvila and Sharon E. Duncan, "A Few Thoughts about Language Assessment: The *Lau* Decision Reconsidered," photocopied material (Larkspur, Calif.: DeAvila, Duncan, and Associates, 1979), p. 9.
16. Ibid.,
17. Ibid.
18. Dubois, 1980, p. 7.
19. James Cummins, "Linguistic Interdependence and the Educational Development of Bilingual Children," *Review of Educational Research* 49, no. 2 (Spring 1979):227.
20. Dubois, 1980, p. 6.
21. Miranda, 1980, p. 2.
22. Edward A. DeAvila et al., *Predicting the Success of Language Minority Students from Developmental, Cognitive Style, Linguistic, and Teacher Perception Measures* (Austin: Southwest Educational Development Laboratory, 1979), p. 53.
23. Ibid., p. 50.
24. Courtney B. Cazden, *Child Language and Education* (New York: Holt, Rinehart and Winston, 1972), p. 226.
25. Ibid.
26. Cazden, 1972, pp. 230-32; DeAvila et al., 1979, p. 51.
27. Cummins, 1979, p. 229.
28. DeAvila et al., 1979, p. 53.
29. Ibid.
30. Ibid.
31. Ibid., p. 38.
32. G. Richard Tucker, "Bilingual Education: Some Perplexing Observations," *Educational Evaluation and Policy Analysis* 1, no. 5 (September-October 1979):75.
33. Miranda, 1979, p. 38.
34. Ibid., p. 43.
35. The analyses presented here were developed by Dr. Rolf M. Wulfsberg, Assistant Administrator for Research and Analysis, National Center for Education Statistics.
36. Donald D. Rogers, "Non-response Analysis," this volume, page 32.
37. Ibid.

References

Cazden, Courtney B. *Child Language and Education.* New York: Holt, Rinehart and Winston, 1972.

Cummins, James. "Linguistic Interdependence and the Educational Development of Bilingual Children." *Review of Educational Research* 49, no. 2 (Spring 1979): 222-251.

DeAvila, Edward A., and Duncan, Sharon E. "A Few Thoughts about Language Assessment: The *Lau* Decision Reconsidered." Photocopied material. Larkspur, Calif.: DeAvila, Duncan and Associates, 1976.

——. *Relative Linguistic Proficiency and Field Dependence/Independence: Some Findings on Linguistic Heterogeneity and Cognitive Style of Bilingual Children.* Austin, Tex.: Southwest Educational Development Laboratory, 1979.

DeAvila, Edward A.; Duncan, Sharon E.; Ulibarri, Daniel M.; and Fleming, James S. *Predicting the Success of Language Minority Students from Developmental, Cognitive Style, Linguistic, and Teacher Perception Measures.* Austin, Tex.: Southwest Educational Development Laboratory, 1979.

Dubois, David D. "Analytical Issues Regarding the *Children's English and Services Study.*" Photocopied material. Washington, D.C.: U.S. Department of Education, National Center for Education Statistics, 1980; also this volume.

Edmonds, Marilyn H. "New Directions in Theories of Language Acquisition." *Harvard Educational Review,* 46, no. 2 (May 1976):175.

Lambert, Wallace E. "Culture and Language as Factors in Learning and Education." In *Education of Immigrant Students,* edited by A. Wolfgang. Toronto: Ontario Institute for Studies in Education, 1975.

Martínez, José. "Response to Issues on LM&AI." Response to AI/CESS with cover letter dated February 22, 1980 at Sacramento, California.

Miranda, L. and Associates, Inc. *Children's English and Services Study: Technical Report on the LM&AI.* Bethesda, Md.: L. Miranda and Associates, Inc., 1979.

——. "Response to the Office of Research and Analysis of the National Center for Education Statistics' Inquiry on Three Analytical Issues Associated with the 1978 Children's English and Services Study." Bethesda, Md.: L. Miranda and Associates, Inc., 1980; also this volume.

O'Malley, J. Michael. *Language Minority Children with Limited English Proficiency in the United States, Spring, 1978* (Draft Report). Washington, D.C.: National Institute of Education, 1979.

——. Response to AI/CESS in the form of a letter dated January 31, 1980 at Alexandria, Virginia.

Peng, Samuel S. Response to AI/CESS in the form of a letter dated February 12, 1980 at Rockville, Maryland.

Rogers, Donald D. Responses to AI/CESS in the form of a letter and attachment dated February 3, 1980 at Austin, Texas; also this volume.

Rowlett, Karen. Response to AI/CESS in the form of a letter dated February 8, 1980 at Austin, Texas.

Silverman, Leslie J. Response to AI/CESS in the form of an Education Division (U.S. Department of Health, Education, and Welfare) memorandum dated February 7, 1980 at Washington, D.C.

Sterling, T.D., and Weinkam, J.J. "What Happens When Major Errors Are Discovered Long After an Important Report Has Been Published?" Paper presented to

the American Statistical Association Annual Meeting in Washington, D.C. on August 16, 1979. Photocopied material. Burnaby, B.C.: Simon Fraser University, 1979.

Tremaine, R.V. *Syntax and Piagetian Operational Thought.* Washington, D.C.: Georgetown University Press, 1975.

Troike, Rudolph C. Personal communications with David D. Dubois, May 30, 1980.

Truex, Kathy. Response to AI/CESS in the form of an Office of the Secretary (U.S. Department of Health, Education, and Welfare) memorandum dated March 18, 1980 at Washington, D.C.

Tucker, G. Richard. "Bilingual Education: Some Perplexing Observations." *Educational Evaluation and Policy Analysis* 1, no. 5 (September-October 1979): 274-75.

Whorf, B.L. *Language, Thought and Reality: Selected Writings of Benjamin Lee Whorf.* Edited by J.B. Carroll. Cambridge, Mass.: Massachusetts Institute of Technology Press, 1956.

Appendices
to
1. "Children's English and Services Study:
A Methodological Review"
by David D. Dubois
National Center for Education Statistics
August 1980

Appendix A

Analytical Issues Regarding the Children's English and Services Study

David D. Dubois, Policy Analyst
Office of Research and Analysis
National Center for Education Statistics
January 21, 1980

Introduction

The 1978 Children's English and Services Study (CESS) was recently completed under contract from the National Institute of Education (NIE), with shared support from the National Center for Education Statistics (NCES) and the U.S. Office of Education (USOE). The final project report is published by NIE, through the National Clearinghouse for Bilingual Education. The principal objective of the CESS was to determine an estimate of the number of limited English proficient (LEP) children between the ages of 5 and 14 in the United States.

This paper discusses three analytical issues that have been identified as a result of an assessment of the research design, data analyses, and other information described in the 1978 CESS Draft Report of September 6, 1979 (and a later revision dated November 1979) entitled "Language Minority Children with Limited English Proficiency in the United States: Spring 1978."

This inquiry is sponsored by the Office of Research and Analysis (ORA) of the National Center for Education Statistics (NCES). To date, reviewers have included the NCES Assistant Administrator for Research and Analysis, the ORA Policy Analyst, and an external consultant from the American Institute for Research in the Behavioral Sciences (AIR) whose services were obtained under contract with the NCES/AIR Statistical Analysis Group in Education. This paper is based entirely upon these reviews.

Recipients of this paper are invited to respond to the analytical issues. Based upon their responses and other information, the Office of Research and Analysis will publish a position paper on the resolution of the identified analytical issues.

Objective of the Inquiry

From an analytical point of view, the 1978 CESS could become a landmark in the determination of estimates of the number of LEP children in the United States. The CESS estimate of the number of LEP children was accomplished directly by developing and administering a domain-referenced content test to a sample of children from language minority households in order to assess language skills in speaking, reading, writing, and understanding English. Before 1978, estimates of this type were derived by using surrogate or indirect measures.

It is anticipated that the results of the 1978 CESS will be used extensively and frequently cited by U.S. Government officials, members of the U.S. Congress, and others. At NCES, for example, it is anticipated that the CESS data base will be used, with other surrogate measures, to calibrate the 1980 U.S. Census data in order to determine recent and accurate LEP person counts. Additionally, the CESS data base will be a component of the NCES study to determine projections of the numbers of LEP persons in the United States for the next 5, 10, 15, and 20 years.

Since the results of the 1978 CESS are of tremendous importance to present and future research studies, bilingual program and policy development, and funding for bilingual education, unresolved analytical issues that could adversely affect the validity of the results are being stated with the hope of their resolution. As mentioned earlier, the ORA will publish a position paper as a result of

this inquiry, to provide a technical reply to each issue. The reply is expected to include recommendations or suggestions for additional research tasks and/or caveats to current CESS reports which could, in the opinion of NCES/ORA, improve the quality of the existing products.

Invitation to Respond

Recipients of this paper are encouraged to respond to the issues. Respondents are assured that their contributions will be carefully considered prior to the development and issuance of the ORA position paper. The position paper will be released only after each recipient (or his or her designate) has responded or has indicated that he or she will not respond to the issues.

Written replies must be received no later than the close of business, Friday February 8, 1980. Replies to the issues must be written and should be addressed to:

Dr. David D. Dubois, Policy Analyst
National Center for Education Statistics
400 Maryland Avenue, SW, Room 3153
Washington, DC 20202
(Telephone: 202-245-8233)

The persons listed below were designated to receive a copy of this paper:

Name	Agency
Edward Bryant	Westat, Inc.
Lois-ellin Datta	National Institute of Education
Karen Dietz/Don Rogers	University of Texas-Austin/Resource Development Institute
Josué M. González	Office of Bilingual Education, U.S. Office of Education
Ron Hall	Office of the Assistant Secretary for Education (Policy Development)
Ty Hartwell	Research Triangle Institute
Reynaldo Macias	National Institute of Education
José Martinez	California State Department of Education
Lourdes Miranda-King	L. Miranda and Associates, Inc.
J. Michael O'Malley	National Institute of Education
Samuel Peng	Westat, Inc.
Leslie Silverman	National Center for Education Statistics
Kathy Truex	Office of the Assistant Secretary for Planning and Evaluation
James Vanecko	Office of the Assistant Secretary for Education (Policy Development)
Carl Wisler	Office of Evaluation and Dissemination, U.S. Office of Education

History of the 1978 CESS

The 1978 CESS was developed by NIE through a contract with L. Miranda and Associates, Inc. Lourdes Miranda-King was the project director, J. Michael O'Malley was the NIE project officer, and Leslie J. Silverman was the NCES coordinator. Subcontractors included Westat, Resource Development Institute, and Research Triangle Institute.

The primary mission of the 1978 CESS was to determine an objective estimate of the number of LEP children, aged 5-14, inclusive, in the United States. The study surveyed a nationally representative sample of households during Spring 1978, identifying those where a language other than English was spoken and where children between the ages of 5 and 15 were living. The Language Measurement and Assessment Inventory (LM&AI), a test in English that determines whether or not a child is limited in English language proficiency, was developed and

administered individually to selected children from the identified households. Specifications for the survey design and the LM&AI were provided by an advisory group composed of State Education Agency representatives in bilingual education, assessment, and data collection.

The LM&AI was designed to measure skills in speaking, understanding, reading, and writing in English. The test is domain referenced for objectives that children aged 5-14 would be expected to perform in order to profit from instruction in an all-English language educational environment.

Ten separate tests, one for each age, were developed and used in the survey. Reliabilities of the test for the separate forms range from .86 to .92. As a result of preliminary field tests of the LM&AI, a critical score *for each age-level test* was determined which could be used to classify each child as proficient in English or as limited English proficient.

The contractor provided three caveats regarding the LM&AI. First, the LM&AI was not designed to determine placement or diagnosis with individual children in educational settings. Second, the instrument was designed in a manner that resulted in an unknown level of cultural bias. Third, the LM&AI items are not "pure" measures of English language proficiency; some of the items assess English language proficiency, memory, and cognitive ability.

Statement and Discussion of the Issues

Three analytical issues are presented and discussed:

1. Were the items which were selected for inclusion in the Language Measurement and Assessment Inventory (LM&AI) selected properly?
2. Were the cut-off scores for the LM&AI, which were determined and used to classify children as either English proficient or of limited English proficiency, set properly?
3. What were the effects of non-response bias on the counts and estimates of the number of LEP children?

If the first question is answered negatively, then the value of the entire 1978 CESS is brought into question. In the event that it is answered affirmatively, then a negative answer to the second question would imply the need for further analysis of the CESS data—and possibly for the collection of additional data—in order to recompute the cut-off scores. The issue raised by the third question could be empirically investigated in the event that it was decided to collect the additional data described earlier.

A detailed discussion of each issue follows.

Issue: Were the items which were selected for inclusion in the Language Measurement and Assessment Inventory (LM&AI) selected properly?

Discussion

Each age-level instrument of the LM&AI consisted of a set of items that could be scored so that a high score would indicate that the child was proficient in English while a low score would indicate that the child was limited English proficient. Therefore, the issue can be rephrased in the following manner. Is English language proficiency the dimension on which the scores vary, or are other dimensions associated with variation in the scores? More specifically:

- Are the test scores related to language dominance?
- Are the test scores related to general language development?

The question of language dominance is addressed in the project Draft Report (November 1979):

> English should be the exclusive criterion irrespective of the child's proficiency in the non-English language. Thus, language dominance was considered irrelevant to the discussion. (page 11-3)

The objective of the study is subject to question on the basis that, for bilingual education policy development, a child's domi-

nant language might affect the potential benefits he or she could derive from participating in a bilingual education program. The reader is cautioned that this review intentionally does not attempt to define operationally the phrase "bilingual education program." Whether it is appropriate to assess English language proficiency while ignoring the child's proficiency in another language remains a policy question to be addressed.

Are the scores on the test related to general language development? The project Draft Report (November 1979) states that:

> Items on the test are not "pure" measures of English language proficiency. In some cases, the items assess English language proficiency, memory, and cognitive ability. The intermingling of the potentially disparate constructs was intentional to give the items as much validity for representing important school tasks as possible (page A-10)

Any test so developed could also differentiate between two children with *equal* English language proficiency, giving a higher score to the child with greater memory or cognitive abilities. It could be argued, therefore, that the test development procedures should have excluded items not primarily associated with English language proficiency. The types of items selected for the test (Draft Report, November 1979, Table A-1) appear generally to assess relevant content. There is, however, a component of general cognitive development, not merely English language development.

The choice of items for the LM&AI was a function of a field test. Items were selected that best differentiated between two criterion groups. The project Draft Report (November 1979) states:

> The test was being developed to differentiate language minorities who were limited in English proficiency from those who could profit from instruction in English Items under development were to be field tested with two clearly defined *criterion groups*. (a) limited English proficient children, and (b) fluent English speaking children who were clearly profiting from instruction in English. (page II-6)

The test was clearly being prepared for administration to language minority children. The dimension being tested was essentially that on which those two groups differed most. It could be argued that the two groups differed on native language as well as English language proficiency and, therefore, the test scores could be expected to have a partial language dominance loading. A solution to this problem might be to equate the two criterion groups on proficiency in a non-English language, making the test independent of language dominance.

Issue: Were the cut-off scores for the LM&AI, which were determined and used to classify children as either English proficient or of limited English proficiency, set properly?

Discussion

The purpose of the LM&AI was to provide a mechanism for determining whether a child was either English proficient or limited English proficient. Therefore, the cut-off score chosen for each age-level test of the LM&AI is critical for determining valid counts. As an example, if the cut-off score on each age-level test is lowered by *two items,* the estimated count of LEP children drops from 2.41 million children to 2.13 million, a decrease of 280,000. Similarly, if the cut-off score for each age-level test is raised by two items, the estimated count of LEP children rises from 2.41 million children to 2.62 million, or an increase of 210,000. Thus, a score difference of four items has the effect of altering the count by nearly one-half million.

Recall that the cut-off score was that score which best differentiated LEP children from fluent English speaking (FES) children who were clearly profiting from instruction in English.

In developing the LM&AI, five techniques were proposed as alternatives for determining the cutoff scores. The five techniques are summarized on page 38 of the *Children's English and Services Study: Technical Report on the LM&AI* (L. Miranda and Associates, Inc., September 10, 1979):

1. For each age-level test determine the score on Field Test II data which was one standard deviation below the mean score for the FES group of that age.
2. Similarly, use that score which was one standard deviation above the mean score for the LESA (later revised to LEP) group of each age.
3. Use the highest Field Test III LESA score made by any individual on each age group test.
4. Plot the scores of LESA and FES separately and select the score equivalent to the point of intersection of the two distributions.
5. Use discriminant function analysis (DFA), which considers subscores to determine a centroid point that can act as the critical point.

After examining the accuracy of the various alternatives, DFA was chosen as the method for determining the cut-off scores. While this Office fully endorses this choice, three subissues remain of concern.*

First, the preceding excerpt from the Technical Report implies that the subscores were used in the DFA. If this is so, several events must have happened.

1. The subscores would be transformed into a new total score representing a linear combination of the subscores. This new score would be real valued (as opposed to integer valued) and it would be conceivable—in fact, highly likely—that relative scores between two individuals could be reversed. That is, if individual A had a higher original score than individual B, the revised DFA score for A could easily be *lower* than that of B due to differential weighting of the subscores. Since no scores on the final CESS tape are non-integer valued, and since no reversal of the kind discussed above occurred, one can only assume that subscores were, in fact, not used in the DFA.
2. The relative weighting of the items, which was carefully designed, would be totally revised by the differential weighting of the DFA procedure. This is another reason that this Office doubts that subscores were used.

The second subissue concerns the application of DFA without concern for the differential "cost" of misclassification. DFA is a powerful tool in that it minimizes the total proportion of the sample that is misclassified. However, if the resulting classification criteria (cut-off scores) consistently misclassify one subgroup (e.g., LEP) at the expense of the other, a serious bias may result. More explicitly, if there are actually N_1 LEP children and N_2 English proficient (fluent) children among the $N = N_1 + N_2$ children of non-English language background households, then the cut-off score will lead to an unbiased classification procedure if and only if $N_2 \cdot \Pr(\text{Classified LEP} \mid \text{Actually fluent}) = N_1 \cdot \Pr(\text{Classified fluent} \mid \text{Actually LEP})$. That is, the expected number of fluent children misclassified as LEP must equal the expected number of LEP children misclassified as fluent.

In defense of the procedure used, since N_1 and N_2 were not known *a priori,* minimizing the overall misclassification error makes reasonable sense. However, as can be seen in Table A, the actual discrimination procedure used was much more likely to misclassify LEP children than fluent children. This explains why the cut-off scores for DFA seemed low (see page 43 of the aforementioned Technical Report).

*The analyses found here were developed by Dr. Rolf M. Wulfsberg, the Assistant Administrator for Research and Analysis at NCES.

Table A

Estimated Conditional Probabilities of Correct and Incorrect Classifications by the LM&AI

Age	P_{11} [a]	P_{12} [b]	P_{21} [c]	P_{22} [d]
5	0.892	0.000	0.108	1.000
6	0.955	0.037	0.045	0.963
7	0.889	0.000	0.111	1.000
8	0.929	0.000	0.071	1.000
9	0.906	0.000	0.094	1.000
10	0.944	0.000	0.056	1.000
11	0.795	0.000	0.205	1.000
12	0.864	0.182	0.136	0.818
13	0.880	0.000	0.120	1.000
14	0.879	0.204	0.121	0.796

a. P_{11} = Pr (Classified LEP | Actually LEP).
b. P_{12} = Pr (Classified LEP | Actually Fluent).
c. P_{21} = Pr (Classified Fluent | Actually LEP).
d. P_{22} = Pr (Classified Fluent | Actually Fluent).

The evident bias described above raises the third subissue: Should the cut-off scores be revised to remove the estimated bias after the fact (when we have estimates of N_1 and N_2)? This Office tends to feel that this should be done, perhaps by the mechanism described below.

Let P_{11}, P_{12}, P_{21}, and P_{22} be defined as in Table A. Let N_1 and N_2 be the actual number of LEP and fluent children, respectively. Finally let L and F be the expected number of LEP and fluent children, respectively, as estimated by the LM&AI. Then,

$$L = N_1P_{11} + N_2P_{12}$$

$$F = N_1P_{21} + N_2P_{22}$$

Solving for N_1 and N_2,

$$N_1 = (LP_{22} - FP_{12})/(P_{11}P_{22} - P_{12}P_{21}) \text{ and}$$

$$N_2 = (FP_{11} + LP_{21})/(P_{11}P_{22} - P_{12}P_{21})$$

By using the actual CESS estimates for L and F, we can then approximate the unbiased values of N_1 and N_2 for each age group. The results, which are shown in Table B, clearly demonstrate the consistent bias in the LM&AI classification procedure. **The CESS/LM&AI LEP counts underestimate the "true" values at every age, except for age 14.**

If we accept the new LEP count as more realistic estimates of the true values, then we can adjust the cut-off scores to reflect these new counts by raising (except for age 14) the cut-off scores until the proper number of children have been classified as LEP. In reality, this point will usually fall in the middle of a cell (score), so one can choose the cut-off score that will yield the closest estimate to N_1.

In the case presented in Table C, a different rule was used. Since the data show an abnormal "roller-coaster" effect in the relationship between age and percent LEP, the cut-off score leading to the percentage closest to the overall mean percentage was chosen for each age group. That is, the lower cut-off score was generally used for even ages and the higher cut-off score was generally used for odd ages.

Table B

Effect on LEP Counts of Removing Estimated Bias

Age	CESS LEP Count	Unbiased LEP Count
5	192,297	215,580
6	291,622	301,767
7	275,924	310,375
8	257,807	277,510
9	167,304	184,662
10	294,156	311,606
11	190,064	239,074
12	251,680	262,412
13	196,577	223,383
14	291,444	284,766
Total	2,408,875	2,611,135

Table C

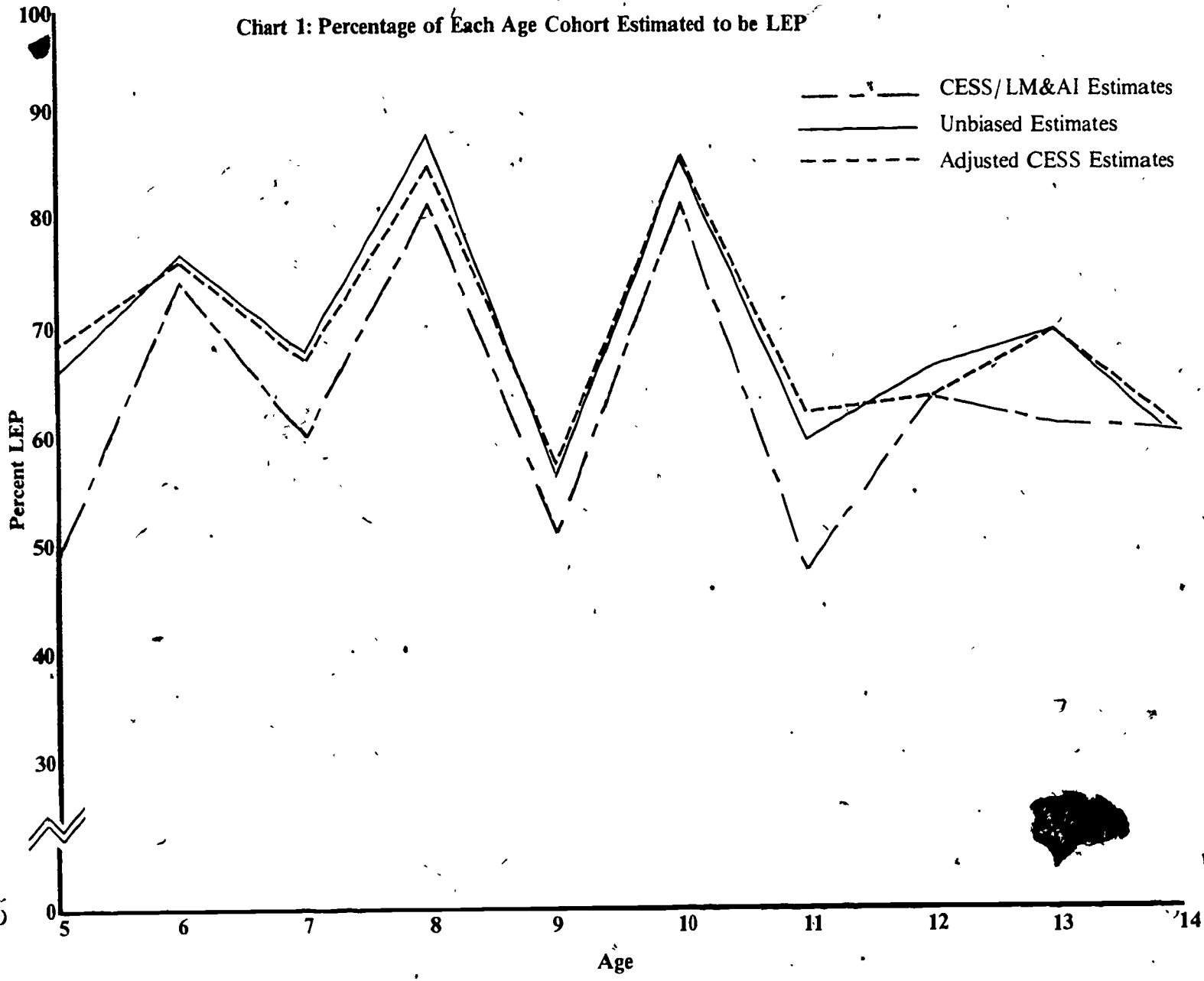
Modified Cut-off Scores and Resulting LEP Counts

Age	Old Cut-off Scores	New Cut-off Scores	New LEP Count
5	18.5	21.5	223,327
6	26.5	28.5	298,929
7	39.5	43.5	307,759
8	38.5	39.5	268,830
9	43.5	46.5	188,187
10	49.5	50.5	310,860
11	41.5	45.5	246,921
12	46.5	46.5	251,680
13	48.5	51.5	223,785
14	52.5	52.5	291,444
Total	—	—	2,611,722

The relationships among the CESS/LM&AI estimates, the unbiased estimates, and the adjusted CESS estimates are evident in Chart 1, following. Chart 1 shows the percentage of each age cohort for each of the estimates of the number of limited English proficient children.

Applying the procedure described above should go a long way toward removing the bias in the LM&AI. Of course, the values of P_{11} used in the derivation are conditioned on the original cut-off scores used by the LM&AI. With the modified cut-off scores, the P_{11}s would change (as do the new L and F

Chart 1: Percentage of Each Age Cohort Estimated to be LEP



counts shown in Table C), so that the results could still change slightly. (This is because the LM&AI sample of 354 fluent children and 337 LEP children are not necessarily representative of their respective populations.) The Office of Research and Analysis is attempting to obtain the original data used to determine the cut-off scores on the LM&AI from L. Miranda and Associates, Inc. in order to explore this issue.

Issue: What were the effects of non-response bias on the counts and estimate of the number of LEP children?

Discussion

In survey research of this type, the potential effects of non-response bias are a reality. The question to be addressed is whether non-respondents are similar to or different from respondents to the study.

Response rates by regional subpopulations (New York, Texas, California, and remainder of the United States) for the household screener, household questionnaire, and administration of the LM&AI are presented in Table III-1 of the Draft Report (November 1979). From the table it can be determined that the response rates, totaled over all subpopulations, were: household screener, 76.2%; household questionnaire, 93.8%; LM&AI administration, 84.6%. Response rates were derived by using the formula

$$\text{Response Rate} = \frac{\text{Total Number Completed}}{\text{Total Number Eligible}} \times 100$$

There is no evidence in the Draft Report (November 1979) to indicate that non-response bias was empirically investigated. Although adjusting weights by poststratification is customary, this is not necessarily a satisfactory substitute for empirically investigating differences between respondents and non-respondents.

If it is determined that the items included in the LM&AI were indeed selected properly and, additionally, a decision is made to collect additional data for recalibrating the LM&AI, an empirical investigation of non-response bias can be undertaken concurrently.

In summary, ORA reviewers believe that these issues can be resolved and, accordingly, that the study can be retained by cooperative, responsible action.

Appendix B

Non-response Analysis

Donald Rogers
Resource Development Institute (Austin, Texas)
February 1980

Purpose

The purpose of this paper is to present the results of a very; very simple analysis of the effects of non-response during the CESS.

General Procedure

The general procedure was to assume that non-responding "SRC incomplete; probable ineligible households (Code 8 households)" had characteristics that were significantly different from responding households. The impact of this assumption was then determined by reweighting the data and recomputing Non-English Language Background (NELB) and Limited English Speaking Ability (LESA) counts.

Limitations

The analysis reported here uses average weights. Ideally, each stratum is considered individually. However, the resources required for a stratum-by-stratum analysis were not available. Therefore average weights were used because they were easy to compute. This means that the results of this analysis only indicate or suggest the type of results that would be obtained by a sophisticated analysis.

References

This paper is based on the information presented in Resource Development Institute's (RDI) reports. Data have been taken from Section 8 (Data Analysis Procedures) and Section 9 (Results) of Volume I. Weighting formulae are taken from Appendix 6.6 of Volume II. The reader must have these reports to follow this paper. For example, the definitions of variables are presented in Appendix 6.6 and are not repeated here.

Assumptions

The following assumptions were made to assess the effects of non-response:

1. All Category 8 households complete the SCR.
2. The percentage of Category 8 households that are eligible and complete the HHQ is twice as great as the percentage of Category 1, 2, and 3 households.
3. All of the eligible Category 8 households complete the HHQ.
4. The average number sampled per eligible household is the same for Category 8 households.
5. The average number of completed LM&AI per household is the same for Category 8 households.
6. The average number of LESA children per household is the same for Category 8 households.

The effects of these assumptions on the "raw" data are presented in the following tables.

Tables A, B, and C

Effects of Assumptions on Raw Data in Assessing the Effects of Non-response

Table A

Household Codes	Number Complete SCR	Percent Complete HHQ	Number Complete HHQ
1,2,&3	25,358	6.5	1,652
8	5,790	13.0	753
Totals	31,148		2,405

Table B

Household Codes	Number Complete HHQ	Average Number Sampled Per Household	Number Sampled	Average Number of Completed LM&AI Per Household	Number of Completed LM&AI
1,2,&3	1,652	1.78	2,953	1.16	1,909
8	753	1.78	1,340	1.16	873
Totals	2,405		4,293		2,782

Table C

Household Codes	Number of Completed LM&AI	Percent LESA	Number LESA
1,2,&3	1,909	71.24	1,360
8	873	71.24	622
Totals	2,782		1,982

Changes in Values

The raw data were used to compute average values for the variables in Appendix 6.6 of Volume II in RDI's final CESS reports. The computed values are presented in the table following; the formulae have been omitted because they appear in Appendix 6.6. Although the use of the symbols is not entirely appropriate and is not precisely consistent with the definitions presented in Appendix 6.6, the results are presented in this manner to make it easy for the reader to follow the calculations.

Table D

Changes in Average Values Computed for Variables

Variable	Average or Estimated Value Before Assumptions	Average or Estimated Value After Assumptions
r_{hij}	33,283	33,283
r'_{hij}	25,358	31,148
W'_{hij}	64.5	64.5
L_{hs}	2,146,753	2,146,753
L'_{hs}	1,635,591	2,009,046
$W^{(1)}_{hij}$	83.9	68.9
S_{hij}	1,762	2,515
S'_{hij}	1,652	2,405
$\tilde{M}_{hs}$	147,832	173,284
$\tilde{M}_{hs}$	138,603	165,704
$W^{(2)}_{hij}$	89.5	72.1
$\Sigma Q'_{hijm}$	3,084,452	3,048,452
C'_{hijm}	2,953	4,293
C_{hijm}	34,061	49,573
Q'_{hijm}	1,032	833
C''_{hijm}	1,909	2,782
$\hat{N}_{hs}$	3,047,496	3,571,776
$\hat{N}'_{hs}$	1,970,088	2,314,624
Q_{hijm}	1,597	1,283
Q (adjusted for SIE)	1,997	1,370
Total NELB	3,811,850	3,811,850

Analysis

The assumptions about the Category 8 households increased the sampled number of NELB persons from 1,909 to 2,782. This is approximately a 46% increase. However, because of the weighting procedures, this increase has *no* meaningful effect on the total U.S. estimates.

The assumption about the Category 8 households increased the sampled number of LESAs from 1,360 to 1,982. This is approximately a 46% increase. The effects on the total U.S. estimate depend upon assumptions about how these cases are weighted. Table E reports the average weights that have been used to this point in the analysis.

Table E

Effects of Assumptions on Average Weights and U.S. Estimate

Assumption	Type	Number	Average Weight	U.S. Estimate
Before	NELB	1,909	1,997	3,811,850
After	NELB	2,782	1,370	3,811,850
Before	LESA	1,360	1,771	2,408,908
After	LESA	1,982	Unknown	Unknown

Table E indicates that before the assumptions, the average LESA weight is less than the average NELB weight. The assumptions that have been made should not affect this relationship, and the average LESA weight should continue to be less than the average NELB weight. However, to test response bias, assume that the NELB and LESA average weights are the same after the assumptions and are equal to 1,370. This yields a total U.S. estimate of 2,715,340 LESAs. This estimate is 306,432 LESAs greater than the LESA estimate reported by the CESS. However, an estimate of 2,715,340 LESAs is within the 95% confidence interval of the total U.S. LESA estimate reported by the CESS.

Conclusion

The analysis that has been reported here is rather simple and superficial. Some of the assumptions that have been made border on being outrageous. Nevertheless, the results of the analysis indicate that these assumptions do not create meaningful differences in the final estimates.

Appendix C

Revised

Table A-4

Classification Errors in Selection of a Criterion Score on the Language Measurement and Assessment Inventory

Age	Proficiency in English on the Predictor	Proficiency in English on the Criterion[a] Fluent	Limited	Critical Score	Percent Accuracy[b]
5	Fluent	32	0	19	90.0
	Limited	7	30		
6	Fluent	27	0	26	87.3
	Limited	9	35		
7	Fluent	31	0	39	89.6
	Limited	7	35		
8	Fluent	36	0	39	95.3
	Limited	3	25		
9	Fluent	35	0	43	91.0
	Limited	6	26		
10	Fluent	35	0	49	91.5
	Limited	6	30		
11	Fluent	34	0	41	82.1
	Limited	14	30		
12	Fluent	27	6	47	83.6
	Limited	3	19		
13	Fluent	42	0	48	92.5
	Limited	5	20		
14	Fluent	39	10	52	82.9
	Limited	4	29		

a. Entries are number of cases in field test three.

b. For example, percent correct at age 5 equals 100 (32 + 30)/69 = 90.0.

Appendix D

Values of L, F, P_{11}, P_{12}, P_{21}, P_{22}, N_1, and N_2 for the Minimized Misclassifications of LEP Children by Age Cohort

Age Cohort	L	F	P_{11}	P_{12}	P_{21}	P_{22}	N_1	N_2
5	254,657	73,212	1.000	0.063	0.000	0.937	249,734	78,136
6	303,584	90,989	0.963	0.091	0.037	0.909	306,970	87,603
7	318,470	144,466	0.968	0.056	0.032	0.944	302,774	142,162
8	280,256	37,083	1.000	0.071	0.000	0.929	277,422	39,917
9	188,187	143,003	0.971	0.031	0.029	0.969	189,277	141,913
10	330,979	32,565	1.000	0.056	0.000	0.944	329,047	34,497
11	271,485	132,731	0.971	0.091	0.029	0.909	266,706	137,510
12	208,426	191,107	0.879	0.136	0.121	0.864	207,388	192,145
13	229,986	94,240	0.976	0.080	0.024	0.920	227,732	96,494
14	245,045	241,365	0.877	0.121	0.123	0.879	246,282	240,128

Appendix E

Contributors

This report was completed while the author was assigned to the National Center for Education Statistics (NCES) as an Education Policy Fellow sponsored by the Institute for Educational Leadership, The George Washington University, and Dr. Rolf M. Wulfsberg, the Assistant Administrator for Research and Analysis at NCES. The author gratefully acknowledges Dr. Wulfsberg's encouragement and support during the development and completion of this project.

Dr. Roslyn A. Korb (Technical Planning Officer, NCES) and Dr. Donald McLaughlin (Statistical Analysis Group in Education, NCES, and the American Institutes for Research) provided technical opinions on the Children's English and Services Study (CESS) in the formative stage of this methodological review.

Dr. Edward A. DeAvila (Linguametrics Group) and Dr. Rudolph C. Troike (then Director, National Clearinghouse for Bilingual Education) provided psychometric and linguistic perceptions on language development, measurement, and assessment.

Dr. Rebecca Oxford (InterAmerica Research Associates, Inc.) shared several references on bilingual education with the author.

Those individuals who responded to the issues paper (Appendix A) provided historical information on the development of the CESS which was essential for obtaining an in-depth understanding of the study.

Finally, Mr. Richard Haber (Division of Multilevel Education Statistics, NCES) completed several programming tasks for this project.

2. Reanalysis of the Number of Limited English Proficient Students Estimated in the Children's English and Services Study

by
Lawrence M. Rudner, Rosedith Sitgreaves Bowker,
and John W. Chambers
National Institute of Education
March 1981

Establishing LM&AI Critical Values

The National Center for Educational Statistics (NCES), Office for Research Analysis (ORA) released, on August 27, 1980, a methodological review of the 1978 Children's English and Services Study (CESS) conducted by NCES and the National Institute of Education (NIE) in conjunction with L. Miranda and Associates of Bethesda, Maryland. The primary objective of the CESS was to derive an estimate of the number of children of limited English proficiency, aged 5-14, from language minority households in the United States. In their report, NCES developed and applied an alternative procedure that resulted in a 9.22% higher estimate. This paper briefly examines the CESS and NCES ORA methods and the differences associated with the final estimates.

A measure of English language proficiency, the Language Measurement and Assessment Inventory (LM&AI), was developed and administered to a probability sample of 1,900 students from language minority households, and total scores for each student were computed. Students within each age level scoring above a predetermined critical value for the age level were classified as fluent (F), students scoring below the value were classified as limited English proficient (LEP). The proportion of sampled students at each level classified LEP was projected to the estimated population of language minority students to estimate the prevalence of limited English proficiency in the population.

In order to set the critical value for each age level, the LM&AI was administered to a validation sample of approximately 60 students per age group whose English abilities were determined by school personnel. The students within each group were selected so that approximately half of the students were fluent and half of them were LEP. The CESS then investigated five methods for determining, from this validation sample, the critical values to be applied to the larger probability sample. NCES ORA developed a sixth alternative.

Different methods for determining the critical values were examined to determine which would minimize the overall error of misclassification of students. Because of the overlap of high-scoring LEP and low-scoring fluent students, two types of error had to be considered: the misclassification of fluent students as LEP (Type I error), and misclassification of LEP students as fluent (Type II error). A cut-off score that was too high would result in excessive Type I error; one that was too low would result in excessive Type II error.

Discriminant function analysis (DFA), the technique used in the CESS report, is designed to maximize the difference in the discriminant function means for the two groups and defines the midpoint of the weighted mean scores of LEP and fluent students as the critical value. The other techniques that were considered, but rejected, maximized the probability of a correct classification of LEP students at the expense of a large Type I error, or they maximized the probability of a correct classification of fluent students at the expense of a large Type II error, or they were inapplicable because of sampling limitations. The CESS report's use of discriminant function analysis sought to meet the objectives of minimizing the overall error of misclassification.

NCES ORA correctly pointed out that the procedure resulted in a greater Type II than Type I error. Due to heterogeneity in the variance of LM&AI scores obtained by the sampled LEP and fluent students, the critical values determined by discriminant analysis appeared to be low. An alternative procedure was developed that sought to equalize the two types of error. This equalization would balance the effect of misclassification and result in a more useful overall count of LEP and fluent students, even though the magnitude of one of the two types of error might be greater. NCES ORA empirically examined every possible critical value at each age level to determine which value best met their objectives.

The NCES ORA and CESS critical values differ substantially and result in different population counts at each grade level. Had

the total score variances of the identified LEP and fluent students been equal, as is assumed in discriminant function analysis, the NCES/ORA and the discriminant analysis would have provided identical critical values and population counts. While the new NCES/ORA method overcomes one problem associated with the validation sample—non-equal total score variances—it does not result in a statistically significant difference in the overall estimate of LEP children.

Accuracy of estimation increases as a function of sample size. The larger the sample size, the more accurate the estimate. Using 1,900 students in the probability sample provided accurate estimates of the proportion of students in each age level scoring at or below each total score value. The sample used to determine critical values, however, was considerably smaller. Table I shows the amount of variation that would be expected as a result of using different samples of approximately 60 students. The 95% confidence intervals, derived from the standard error of estimate, provide an upper and lower boundary to the critical value estimates. Scores within this band are not considered to be significantly different from the original CESS values at the .05 probability level. Due to the small samples, these confidence intervals are extremely wide. In all but four age levels, the NCES/ORA revised critical values fall within the confidence interval. The overall NCES/ORA estimate of 2.6 million LEP students falls within the confidence interval placed around the CESS estimate of 2.4 million.

Table I

Language Measurement and Assessment Inventory
Validation Summary Statistics

Age	N[5]	SD[1, 5]	SEE[2]	CESS Critical Value	95% Confidence Interval[3]	NCES/ORA Critical Value[3, 4, 6]
5	69	12.8	1.54	18.5	15.5-21.5	25.5*
6	71	16.1	1.91	26.5	22.5-30.5	29.5
7	67	18.0	2.20	39.5	35.5-43.5	44.5*
8	64	15.4	1.92	38.5	34.5-42.5	40.5
9	67	20.0	2.44	43.5	38.5-58.5	46.5
10	71	20.0	2.37	49.5	44.5-54.5	52.5
11	78	14.6	1.65	41.5	38.5-44.5	51.5*
12	55	17.3	2.33	46.5	41.5-51.5	44.5
13	67	13.9	1.70	48.5	45.5-51.5	52.5*
14	82	24.0	2.65	52.5	47.5-57.5	49.5

1. SD - overall standard deviation for the combined groups.
2. SEE - standard error of the critical value estimate = sd / $\sqrt{N}$
3. 95% confidence interval = CESS critical value ± 1.96 SEE since only whole number scores are possible. Critical values have been rounded to the nearest midpoint.
4. Starred values fall outside the confidence interval.
5. From Miranda (1979)
6. From Dubois (1980b)

The CESS report estimated the number of LEP children to be 63% ± 8% of all children, aged 5-14, living in language minority households.

In summary NCES/ORA achieved, in an alternative, empirical manner, the CESS investigation goal of establishing LM&AI critical values that result in useful estimations of the number of LEP children at each age level. The resultant critical values on the LM&AI and, hence, the estimates of LEP students at each level, differ from the original CESS values because of the heterogeneity of variance in the validation sample. However, in a majority of cases the NCES/ORA revised critical values and revised total count fall within the confidence intervals placed around the CESS values. Since the approach developed and applied by NCES/ORA is more in line with objectives of the CESS investigation and overcomes one limitation resulting from the relatively small sample size, its resultant estimates of the number of LEP children represent a theoretical improvement over the original values. Further research using larger samples to obtain estimates with much smaller confidence intervals is highly recommended.

References

Dubois, David D. "Analytical Issues Regarding the *Children's English and Services Study.*" Washington, D.C.: U.S. Department of Education, National Center for Educational Statistics, 1980a; also this volume.

_____. "*The Children's English and Services Study:* A Methodological Review." Washington, D.C.: U.S. Department of Education, National Center for Education Statistics, 1980b; also this volume.

Miranda, L., and Associates, Inc. *Children's English and Services Study: A Technical Report of the LM&AI.* Bethesda, Md.: L. Miranda and Associates, Inc., 1979.

Appendices
to
2. "Reanalysis of the Number of Limited English Proficient Students Estimated in the Children's English and Services Study"
by Lawrence M. Rudner, Rosedith Sitgreaves Bowker, and John W. Chambers
National Institute of Education,
March 1981

Appendix A

Description of the Language Measurement and Assessment Inventory (LM&AI)

The LM&AI was designed for the Children's English and Services Study (CESS) to estimate the number of limited English proficient children in the United States. The study provides estimates of the number of limited English proficient (LEP) children aged 5-14 living in households where a language other than English was spoken in Spring 1978. Specifications for the survey design and for the LM&AI were provided by an advisory group composed of State Education Agency representatives in bilingual education, assessment, and data collection. Major results of the study were reported in May 1979 at the annual meeting of the National Association for Bilingual Education.

The LM&AI provides a single score that reflects overall English language proficiency for skills that, in the judgment of the advisory group, are typically taught in the schools and are important for being able to profit from instruction in English. The skills include speaking, understanding, reading, and writing in English. The test is individually administered and requires 20-30 minutes to complete. Language groups included during field tests were Arabic, Chinese, Greek, Italian, Japanese, Polish, Spanish (Cuban, Mexican American, and Puerto Rican), and Vietnamese as well as native English speakers. Items retained in the final version of the test met the content specifications provided by the advisory group, and discriminated statistically between children designated by schools as either limited or fluent in English language proficiency. There are 10 separate forms of the test, one for each age from 5 to 14 years.

Reliabilities of the test for the separate forms range from .86 to .92.

A criterion score for each age form was selected that identified LEP children. For this purpose, a statistical technique was chosen that maximized accurate classifications based on the total test scores when attempting to predict school classifications of fluent or limited English proficient groups. The criterion score was used in a national CESS sample to identify LEP children.

The LM&AI is not intended for individual diagnoses or for use beyond the purpose for which it was designed in the CESS survey, nor should the LM&AI be used as a diagnostic instrument for any of the sub-skill areas in oral or reading proficiencies without further development work.

The National Institute of Education has awarded copyright to the test developer, who should be contacted for further information about the instrument: L. Miranda & Associates, 4340 East-West Highway, Suite 906, Bethesda, Maryland 20014, (301) 656-8684.

Appendix B

Response to
The Office of Research and Analysis of the
National Center for Education Statistics'
Inquiry on Three Analytical Issues
Associated with the 1978
Children's English and Services Study

Submitted to:

The National Center for Education Statistics
Office of the Assistant Secretary for Education
Washington, D.C. 20202

Submitted by:

L. Miranda and Associates, Inc.
4340 East-West Highway, Suite 906
Bethesda, Maryland 20014
(301) 656-8684

February 15, 1980

February 15, 1980

David D. Dubois
Rolf M. Wulfsberg
National Center for Education Statistics
400 Maryland Avenue, S.W., Room 3153
Washington, D.C. 20202

Dear Messrs. Dubois and Wulfsberg:

We are pleased to present herein our response to your "Analytical Issues Regarding the Children's English and Services Study" paper. In order to draw on as many perspectives as possible, several former consultants to L. Miranda & Associates, Inc. (LM&A) during the CESS project were queried on their reaction to the three issues raised. Their recommendations were all taken into account in drafting this response, which we trust is both responsible and comprehensive.

Although the Children's English and Services Study contract has been completed and the work accepted, LM&A has willingly accepted this opportunity to clarify concerns associated with that effort. Further, given previous staff commitments, we appreciate your extension of time to February 15, 1980, which has allowed us to prepare a more detailed response.

Sincerely,

Lourdes Miranda
President
L. Miranda and Associates, Inc.

52

Issue Number One: Were the items that were selected for inclusion in the Language Measurement and Assessment Inventory (LM&AI) selected properly?

The first issue is one that we have addressed before and appears to be raised by individuals relatively new to the LM&AI, who may be unclear about the instrument's functional intent as well as unaware of its theoretical underpinnings.

As stated in the Executive Summary (p iv) of the CESS Draft Report of September 6, 1979 (and the later revision of November 1979): "The test in English was designed to meet the definition of limited proficiency in the Bilingual Education Act." According to Title VII (Bilingual Education Programs) of the 1965 Elementary and Secondary Education Act—the legislation still in effect during the development of the LM&AI—individuals of limited English proficiency were those who had ". . . difficulty speaking and understanding instruction in the English language" [Sec. 703(a) (1) (B)]. This is admittedly a narrow definition. It does not refer to the individual's proficiency in a native language nor to the relative balance between languages or dominance of one language over the other, it does not even refer to general communicative competence in the English language. The crux of the definition is *instruction* in the English language.

The reviewers rephrased this first issue of concern to ask: Is English language proficiency the dimension on which the scores vary, or are other dimensions associated with variation in the scores? Our answer then is yes, other dimensions are associated with score variation. However, these dimensions include skills that are also necessary to function successfully at any given level in the U.S. public educational system. The reviewers have suggested that the intermingling of items that measure memory and cognitive abilities in addition to language skills is inappropriate. We maintain that language proficiency is inextricably related to memory and cognitive abilities. It is difficult to imagine how a test of "pure" linguistic competence could have been developed should we have been asked to do so. Inasmuch as the prescribed function of the LM&AI was to measure language proficiency as it relates to classroom instruction, it was essential for us to measure the ability to deal successfully with academic classroom tasks that are often as clearly reliant on memory and cognitive abilities as on English language skills.

We hope that the preceding comments will clarify any misunderstandings regarding the instrument's (1) functional intent, i.e., the development of an instrument for identifying individuals of limited English proficiency who are not able to learn successfully in a classroom where English is the language of instruction, and (2) theoretical underpinnings, e.g., that language proficiency and memory and cognitive skills are inextricably related.

As the reviewers note, the LM&AI is a domain-referenced test. The issue of the appropriateness of item selection then becomes twofold: Are the proper domains represented, and are items within a domain relevant to that domain? Our answer is an emphatic yes. The domains themselves, as well as the items within each domain, were carefully selected and reviewed by dozens of educators, curriculum specialists, psychologists, counselors, administrators, and evaluators. The expert panel members and their respective institutional affiliations are listed in Appendix A, which also indicates the number of CESS Review Panel sessions each member attended. In addition, there were three field tests of the instrument, at which time about 100 teachers and psychologists experienced first-hand contact with the contents of the tests. These professionals provided critical feedback to the test developers which was of great value in developing subsequent revisions of the test. Furthermore, remarkable foresight on the part of the CESS Review Panel is evidenced by the fact that the very four domains selected in 1977 to constitute the instrument include the two additional skills mentioned in the 1978 amendments to the Elementary and Secondary Education Act. In the Title VII (Bilingual Education Programs) amendment . . . the definition of individuals of limited English pro-

ficiency has been expanded to include those who. ". . . have sufficient difficulty speaking, reading, writing, or understanding the English language to deny them the opportunity to attain levels comparable to others at their appropriate age and grade levels" [Sec. 703 (a) (1) (B)]. Item selection within the domains was a two-fold process. First, each potential test item was studied and weighed for its presumed representativeness of a given domain and assessed by competent judges as to its relevance to the skill being measured. Second, selected items were retained in subsequent versions on the basis of their ability to differentiate between two criterion groups. Independent as well as collective judgments were used in the successive versions of the 10 age-level tests.

The final subissue of the first issue brought up by the reviewers deals again with the question of dominance and the distinguishing characteristics of criterion groups. The quotation on page 8 from the November Draft Report (p. II-6) is misleading when taken out of context. It implies that the fluent English speaking children forming criterion group B were also of language minority groups. Only if this were the case would the subsequent reviewers' comments be relevant: "It could be argued that the two groups differed on native language as well as English language proficiency. . . . A potential solution to this problem would be to equate the two criterion groups on proficiency in a non-English language" (p. 8). However, as stated in the following paragraph on p. II-6 of the November Draft Report: "The fluent English speaking group would be composed of students who were *native English speakers*. . . . There is, therefore, no danger that the test scores differentiating the criterion groups have a partial language dominance loading. The influence of language dominance on respective language competence and development in bilingual children is a subject that psycholinguists are just beginning to address, generally through descriptive case study research. There are as yet no empirical research findings in this area upon which to draw in test development. Furthermore, until the legislation governing bilingual education programs expands to include concern about the development of the native language of linguistic minorities, language dominance is a moot point.

Issue Number Two: Were the cut-off scores for the LM&AI, which were determined and used to classify children as either English proficient or of limited English proficiency, set properly?

Subissue One:

Unfortunately, this issue arose because a necessary revision in the Technical Report excerpt cited on page 26 of this volume was not made. Item 5 describes one type of discriminant function analysis considered, but not the procedure ultimately decided upon. Although a discriminant analysis using subscores offered the most powerful use of the test to classify pupils, it was decided against for practical reasons. An easier, more practical discriminant analysis procedure was chosen which uses the total score as a single scale and thus produces a single cut-off score. Item 5 should have been changed to read: "Use discriminant function analysis (DFA) which *utilizes total test scores* to determine a centroid point, which can act as the critical point." Thus, the comments (1) and (2) on page 27 of this volume are no longer applicable as subscores were indeed not used as the reviewers suspected. We apologize for this oversight on our part and the confusion it created.

Subissues Two and Three:

Balancing the misclassifications (as the mechanism suggested for revising cut-off scores would do) was an option considered by LM&A. We chose, instead, to minimize overall misclassifications, i.e., to improve the overall "hit-rate." The "Least Squares" approach was selected in order to minimize the total number of errors. We are aware that the approach chosen results in relatively conservative estimates of LEP children because it classifies more LEP children as FES than vice versa. While this has unfortunate impli-

cations for the individual child who will be deprived of participation in a needed ESL or bilingual program because he or she is misclassified, the CESS concerns were with an overall estimate of numbers. The method proposed by the reviewers does not take into account the potential consequences of misclassifying the FES group while correcting the LESAs. Increasing the number of FES pupils misclassified in a new "unbiased" count would touch upon a sensitive political issue. As you may be aware, since the inception of bilingual education in this country, the issue of providing equal educational opportunities to children of non-English language background has been surrounded by controversy. The issue has centered around Congressional concern with tightening definitions of eligible children to ensure that resources reach those who are most in need. The AIR Evaluation Study of Title VII, by estimating that only approximately one-third of Title VII pupils were LEP, is an example of how the controversy has been fueled.

L. Miranda and Associates, Inc., an Hispanic-owned, minority firm, has been particularly sensitive to this controversy. With that sensitivity in mind, the Contractor, with the concurrence of the NIE Project Director, opted for the more conservative approach which appeared to offer an objective estimate of children of limited English proficiency beyond Congressional reproach.

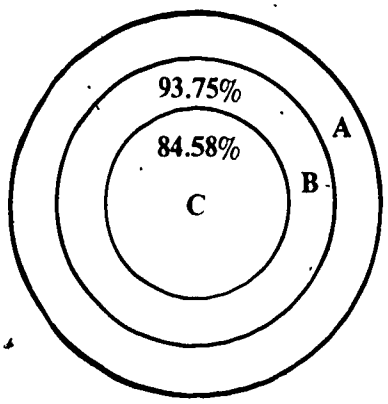
Another reason that an approach balancing the misclassifications was not selected is that it would have failed to account for the unique leptokurtic curve of the FES group. Because of the leptokurtic distribution, the misclassifications were unbalanced. Since the imbalance is an artifact of the distribution, it makes no sense to balance the misclassifications and thereby give an artificial impression of where cut-off scores ought to be.

Issue Number Three: What were the effects of non-response bias on the counts and estimate of the number of LEP children?

We do, of course, realize that the potential effects of non-response bias are a reality. Unfortunately, we were working under time and cost restraints that did not permit us to carry out an analytical inquiry into non-response bias, nor was that activity in the originally proposed study design agreed upon by NIE.

Because of the way the CESS was conducted, it would not be difficult to go back to the computer for an empirical investigation of non-response bias without collecting additional data. The concentric circle diagram following serves to illustrate the kind of data on tape.

Diagram I: Response Rates of Identified NELB Households



Area A. Households responding to Household Screener but not to Household Questionnaire
Area B Households responding to Household Screener and Questionnaire, but not to LM&AI
Area C. Households, responding to Household Screener, Household Questionnaire, and LM&AI

The total area (A, B, and C) represents the non-English language background (NELB) households responding to the Household Screener (76.19% of the total sample). This response rate is not particularly disturbing when closer analysis reveals that only 57 out of the 7,214 households not responding were expected to be language minority (November Draft Report, page III-10). The actual response rate of NELB households within the total sample was very good indeed. Since the relevant population for the Household Questionnaire and the LM&AI test was NELB, the purpose of the Household Screener was to narrow the sample to that population, therefore, there could be only minimal non-response bias at this level.

The response rate for the Household Questionnaire was 93.75% nationally (Areas B and C), indicating that once a language minority household was located, cooperation was very high. Area A, then, represents the households that responded to the screener but did not respond to the Household Questionnaire The national response rate for the language test was 84.58%, represented by the Area C. The largest categories within the non-response group to the LM&AI test (Area B) were refusals by parents and unavailability of children at appointed test time (November Draft Report, page III-12).

Since Areas A and B represent the non-response groups to the Household Questionnaire and the LM&AI test respectively, the two groups could each be compared with the Area C group with regard to the household linguistic environment. The Area A group (those not responding to the Household Questionnaire) and the Area C group (the total response group) could be compared, for example, on their responses to items S2 - S5 on the Household Screener Test (e.g., "What language do the people in this household *usually, often, sometimes* speak here at home?") Since the response rate for the Household Questionnaire was relatively high (93.75%), this comparison would not be as crucial as the one that could be made between the Area B group (non-respondents to the LM&AI test) and the Area C group (total response). Because of the lower response rate on the LM&AI (84.58%), this would appear to be the more relevant comparison for empirically investigating differences between respondents and non-respondents. Whereas the previous comparison would be made with the household as the unit, this comparison would be individual child based, comparing parent-reported linguistic behavior and ability of children who were not tested with those children who were tested. Several items on the Household Questionnaire lend themselves to this end. These items include:

4. How well does (Name) *understand* spoken English?
5. How well does (Name) *speak* English?
6. What language does (Name) usually speak to (Name's) brothers and sisters?
7. What language does (Name) usually speak to (Name's) best friends?

Thus, as we have shown, an empirical investigation of non-response bias could be carried out without further data collection, but it would involve the cost of computer and staff analysis time. Since this activity was not part of the approved CESS study design, we do not feel obligated to carry out the activity at our own expense, nor do we feel that the issue is of such magnitude as to jeopardize the validity of our findings. It is an intriguing issue, however, and one which we would certainly be interested in pursuing with additional funding. This is probably a dimension which should be built into any further CESS study research design.

Appendix C

Master List—Reviewers

Name	Address	Reviewers Meetings Attended
Jerry T Barton	Texas Education Agency 201 East 11th Street Austin, TX 78701 (512) 475-4296	I, II, III, IV
Brian E Bethke	Educational Specialist Illinois Office of Education 188 West Randolph Street Chicago, IL 60601	IV
Ann A Beusch	State Department of Education Box 8717 - BWI Airport Baltimore, MD 21240	IV
Edith T Byrne	Illinois Office of Education 188 West Randolph Street Chicago, IL 60601	IV
Elena Chaves-Muell[illegible]	Bilingual Consultant State Department of Education 126 Langdon Street Madison, WI 53702 (608) 266-2658	
Dan Chávez	Bilingual Consultant Urban Education Section State Department of Public Instruction Grimes State Office Building Des Moines, IA 50319, (515) 281-3805	I, II, III, IV
David Chestnut	Foreign Language Education Advisor State Department of Education Education Building Room 501 Harrisburg, PA 17126 (717) 787-7089	I, II, III
Anne Covill	Assistant Director Joint Data Project Council of Chief State School Officers 1201 16th Street, N.W. Washington, D.C 20036	II,III, IV
David Cox	Assistant Superintendent of Foreign Language State Department of Education Richmond, VA 23216 (804) 786-7757	I, IV

Keith Crosbie	Coordinator of Bilingual Education Office of Superintendent of Public Instruction Old Capitol Bldg Olympia, WA 98504	I, II, III
Robert Esparza	Deputy Superintendent State Department of Education Santa Fe, NM 87503	II
Jerry Fuller	Director Compensatory Education Section State Department of Education 942 Lancaster Drive Salem, OR 97310	II
Al Gage	Consultant in Foreign Language Bilingual Education State Department of Education Oklahoma City, OK 73105 (405) 521-3361	I, II
Carlos González	Bilingual Bicultural Education Section State of California Department of Education State Education Building 721 Capitol Mall Sacramento, CA 95814 (916) 445-2972	III
Dr Renato J. González	Director of Bilingual Education Department of Education 420 Michigan National Tower Lansing, MI 48909 (517) 373-9467	I
Elaine Gordon	Assistant Director Bilingual Education Office Michigan Department of Education Box 30008 Lansing, MI 48906	III
Fannetta N Gordon	Senior Advisor Language Education Pennsylvania Department of Education Bureau of Curriculum Services Box 911 Harrisburg, PA 17126	IV
Dr. Arturo L. Gutiérrez	Deputy Associate Superintendent Dallas ISD 3700 Ross Avenue Dallas, TX 75204 (214) 824-1680	I, II
Michael Hebert	Bilingual Education Specialist State Department of Education P O. Box 44064 Baton Rouge, LA 70804 (504) 389-6486	I, II

Dean Hirt	State Department of Education 201 E Colfax Denver, CO 80203	I, II, III, IV
Dr John Howard	Foreign Language Specialist Department of Education 501 Dexter Avenue, RM 606 Montgomery, AL 36130 (205) 832-3820	I, II
Dr Elliot Howe	Director Bilingual Education State Board of Education 250 East 5th South Street Salt Lake City, UT 84111	II
Ms Robin Johnston	Chairman State Bilingual Department of Education 340 Marion Street Denver, CO 80218	II
J L Kelley	Education Specialist Alabama State Department of Education State Office Building 501 Dexter Avenue Montgomery, AL 36130	II
Dr James A Kemp	Educational Information Services Florida Department of Education Room 275-Knott Building Tallahassee, FL 32304	I, II, III, IV
Anthony Koester	Consultant for Foreign Languages Kentucky State Department of Education Capitol Plaza Tower - 18th Floor Frankfort, KY 40601	II
Sam R Lacher	Assistant Director of Planning and Evaluation Department of Public Instruction Bismarck, ND 58505	I, II, III
Dr Hernan La Fontaine	Executive Administrator Office of Bilingual Education New York City Board of Education 66 Court Street Brooklyn, NY 11201 (212) 596-8038	I, II
Dr José Martinez	Department of Education State of California State Education Building 721 Capitol Mall Sacramento, CA 95814 (916) 322-5012	I, II, III
Ernest J Mazzone	Director of Bureau of Transitional Bilingual Education State Department of Education 182 Tremont Street Boston, MA 02111 (617) 727-8300	I, II, III, IV

Harvey Miller (representing Dr Ramiro Reyes)	Bilingual Bicultural Education Section State Department of Education 721 Capitol Mall Sacramento, CA 95814	II
Jessie Montoya	Director Bilingual Education Section State Department of Education 809 Capitol Square Building St Paul, MN 55101	IV
Omar P Norton	Department of Education and Cultural Services State of Maine Augusta, ME 04333	IV
Agnarda D. Palsha (Mrs.)	Director Bureau of Bilingual Education New Jersey Department of Education 225 West State Street Trenton, NJ 08625	III
Richard Parker	Department of Education and Cultural Affairs Division of Elementary and Secondary Education State Capitol Building Pierre, SD 57501	IV
Henry Pascual	Bilingual Teacher Training Unit State of New Mexico Department of Education Education Building Sante Fe, NM 87503	III, IV
Dr Maria Ramirez	State Coordinator of Bilingual Education State Education Office Room 308 Albany, NY 12234 (518) 474-8775	I
Hubert Richards-	Language and Humanities Florida Department of Education 108 Miles Johnson Building Tallahassee, FL 32304 (904) 489-1636	I
Fred Rodriguez	Nebraska Department of Education 301 Centennial Mall South Sixth Floor Lincoln, NE 68509	II
Dr. Joe Thompson	Office of Program Evaluation and Research State Department of Education 721 Capitol Mall Sacramento, CA 95814 (916) 322-5012	I
Elena Vergara (Mrs.)	Project Coordinator State Department of Public Instruction 120 West Market 10th Floor Indianapolis, IN	I, II, IV

Ms Ethel Ward	Program Specialist for Limited English Speakers 1210 Queen Emma Honolulu, HI (808) 548-4635	IV
Dr Protase Woodford	Educational Testing Service Princeton, NJ 08540	I, II
Rose Yamada (Mrs)	Advisors National Compensatory State Department of Education P O Box 2360 Honolulu, HI 96804	I
Ernesto Zamora	Educational Consultant Texas Education Agency 201 East 11th Street Austin, TX 78701	I
Nancy F Zelasko	Educational Research Assistant D C Public Schools Division of Bilingual Education Room 900 415 12th Street, N W Washington, D C 20004	II, III, IV