DOCUMENT RESUME

ED 403 739 FL 024 277

AUTHOR Lara-Alecio, Rafael; And Others
TITLE Math Assessment with Manipulatives for Hispanic Limited English Proficient Students.
PUB DATE [97]
NOTE 40p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Difficulty Level; *English (Second Language); Evaluation Methods; Grade 2; Grade 3; *Hispanic Americans; Kindergarten; Language Proficiency; *Limited English Speaking; *Manipulative Materials; *Mathematics Education; Primary Education; Sex Differences; Spanish Speaking; *Student Evaluation; Test Reliability; Test Use

ABSTRACT

A study investigated the use of manipulatives-based mathematics tasks as an alternative method for assessing young Hispanic students with limited English proficiency. Students (n=45) from kindergarten and grades 2 and 3 were administered 14 manipulatives-based tasks. Each task was retested over a period of 2-3 weeks. Task performance was scored on four scales: conceptual understanding; efficient strategy use; accuracy; and fluency. The study addressed research questions of retest reliability, task difficulty, and relationships among subscales, patterns of change in subscales, relationship to general math skills, relationship to language proficiency, and relationship to gender. Results indicate that the manipulatives-based tasks were challenging for this population, despite use of manipulatives and continuous access to them in the classroom. Students improved significantly on six of the 14 tasks over the retesting period, without targeted instruction. Subskills were tightly interrelated at earlier ages, highly differentiated in grade 3, suggesting that manipulative tasks may function differently according to developmental level. Changes from first to second assessment were even across the four criteria subscores. Little relationship was found between math ability and English or Spanish fluency. Small but consistent gender difference was found. (Contains 75 references.) (MSE)

Math Assessment with Manipulatives for Hispanic Limited English Proficient Students

Rafael Lara-Alecio, Ph. D.
&
Richard Parker, Ph. D.
Texas A&M University
College Station, Texas

Beverly J. Irby, Ed. D.
Sam Houston State University
Huntsville, Texas

Samantha Mason, M. Ed.
&
Claudia Avila, M. Ed.
Texas A&M University
College Station, Texas

Abstract

The purpose of this study was to examine the use of manipulatives-based mathematics tasks as alternative assessment methodology with young, Hispanic students with Limited English Proficiency (LEP). Forty-five students in Kindergarten, Grades 2 and 3 were administered a total of 14 manipulatives-based tasks. Each task was retested over a period of 2-3 weeks. Task performance was scored on four scales: (1) conceptual understanding, (2) efficient strategy use, (3) accuracy, and (4) fluency. This study addressed research questions of retest reliability, task difficulty, and relationships among subscales, patterns of change in subscales, relationship to general math skills, relationship to language proficiency, and relationship to gender. Answers to each question are presented, along with the caution that these results are from a small sample population, exploratory study, and need replication.

Mathematics Underachievement by Hispanic Students

Hispanic students in the United States are at risk for underachievement and failure (Laosa & Henderson, 1991; Orum, 1986; Valencia, 1991) with mathematics being especially problematic (Haycock & Navarro, 1988; MacCorquodale, 1988; Policy Analysis for California Education, 1990). Hispanic students' underachievement in mathematics is recognized by third grade -- the time at which the children can be reliably, formally tested (Dossey, Mullis, Lindquist, & Chambers, 1988). Because of the generally, low mathematics achievement of Hispanic students, a disproportionate number are classified as disadvantaged and enrolled in Title I programs (Kennedy, Birman, & Demaline, 1986). The majority are classified as Limited English Proficient (LEP) and placed in Bilingual or English as a second language (ESL) programs, which currently lack sufficient published evidence on their effectiveness (Cziko, 1992; Lam, 1992).

Several possible explanations have been advanced for underachievement of Hispanic students in mathematics. These explanations are not mutually exclusive; more than one is typically held by any particular researcher. One argument is that academic content is ineffectively taught and tested because both occur outside of Hispanic students' ethnographic context. The solution offered is to efficiently develop competence and confidence through "Ethnomathematics" (D'Ambrosio, 1987; Massey, 1989; Stigler & Barnaces, 1988). The Ethnomathematics point of view calls for a thorough reconstruction of the mathematics curriculum to achieve "cultural compatibility" (Moll & Diaz, 1987; Trueba, 1988). The term "curricular contextualization" has also been used to refer to relating mathematics instruction to the home culture (Tharp, 1989). In this reconstruction, students' family experiences and vocabulary are used to frame mathematics problems in the classroom (Henderson & Landesman, 1992).

A second explanation for Hispanic student underachievement in mathematics is that assessment problems are presented in abstract form and are overly reliant on verbal/reading skills and on linear reasoning. Some evidence exists that LEP student achievement improves with use of hands-on teaching and testing, permitting hands-on manipulation of three dimensional props (Garcia, 1991; Tharp, 1989). Through manipulatives, students can create concrete, visible connections of mathematical principles and procedures (Bohan, 1995). Other researchers have lauded the potential of manipulatives in mathematics instruction, but cautioned that manipulatives are commonly misused (Carpenter, Fennema, Fuson, Heibert, Human, Murray, Olivier, & Wearne, 1994).

A third point of view is that LEP students lack "linguistic engagement" of their first language in mathematics (Henderson & Landesman, 1992). Their lack of English fluency prevents engagement in the highly abstract vocabulary and principles of mathematics. Furthermore mathematics word problem's semantic structure is an important contributor to the problem's difficulty level (Carpenter & Moser, 1984, 1983, 1982; Riley & Greeno, 1988, 1983). In addition, recently educators have begun to understand that mathematics, itself, is a discrete language (Dale & Cuevas, 1987; Orr, 1987, Pimm, 1987; Reilly, 1988). Several researchers contend that full understanding of mathematics problems requires facility with their unique language, including logical, semantic, and syntactic features (Briars & Larkin, 1984; Dean & Malik, 1986; Kintsch & Greeno, 1985; Riley & Greeno, 1988; Riley, Greeno, & Heller, 1983).

A fourth point of view on Hispanic students' underachievement in mathematics contends that poor-quality instruction typifies many classrooms with high enrollment of ethnic minorities, especially Title I, Bilingual and ESL programs. This instruction is denigrated as focusing on basic skills, rote learning, rigid problem-solving procedures and lower-order thinking skills (Kennedy, Birman & Demaline, 1986; Nunes, 1992). It is further contended that schools with predominately Anglo enrollments are more likely to attend to higher order thinking skills (Simmons, 1985). Evidence is mounting that an emphasis on higher order thinking activities lead to achievement gains in LEP students (Tikunoff, 1983; Wong Fillmore, Ammon, McLaughlin, and Ammon, 1985), especially those of Hispanic origin (Garcia, 1991, 1988).

Concerted efforts such as those by the National Council of Teachers of Mathematics appear to be influencing math instruction in regular education, but no evidence exists on the impact on such special programs as bilingual, ESL, and Title I (National Council of Teachers of Mathematics, 1995).

Alternative Assessment in Mathematics

Formal assessment in both mathematics and literacy has been under sustained attack for over a decade, and the criticism seems to be accelerating. Wixson, Valencia, and Lipson (1994) recently observed that the criticism is coming from all camps, including psychometricians (e.g., Linn, Baker, & Dunbar, 1991; Shepherd, 1991), policy-makers (e.g., Darling-Hammond, 1991; Smith, 1991), test publishers (e.g., Kean, 1992), researchers (e.g., Calfee & Hiebert, 1991; Farr, 1992), and classroom teachers (e.g., Howard, 1990). Formal mathematics assessment is criticized on several grounds: for (a) over-emphasizing basic skills, (b) for relying on unauthentic multiple choice selection responses, (c) for rigidly considering only a single solution, (d) for not permitting students to earn partial credit, (e) for scoring only the final answer or product in exclusion of the problem-solving process, (f) for failure to embed math problems in a realistic cultural context, and (g) for testing with artificial time constraints and without intermediate feedback (Alexander & James, 1987; Hambleton & Swaminathan, 1985; Gentry-Norton, 1995; LaCelle-Peterson & Rivera, 1994; Murnane & Raizen, 1988; Schmidt, 1983; Freeman, Kuhs, Porter, Floden, Schmidth, & Schwille, 1983; Stone, 1989; Tindal & Marston, 1990; United States Department of Education, 1992).

Thus, the contemporary movement in mathematics education toward more applied problem-solving, higher order thinking skills, multiple problem-solution strategies, and use of hands-on manipulatives can offer little in the way of assessment to accompany these instructional changes. Yet assessment in math and other skill areas is pervasive and strongly impacts what lessons are planned and taught, and what teaching efforts are rewarded (Salmon-Cox, 1980; Stiggins, Conklin, & Bridgeford, 1986). Because math assessment results are influential, one might expect a negative impact from assessment results which poorly reflect changes in curriculum and instruction (Gentry-Norton, 1995).

Work with varied math testing formats, such as process assessment and open-ended questions is still only at an exploratory level (Romberg, Zarinnia, & Collis, 1990). Reports on varied assessment formats typically emphasize their informal use, without establishing psychometric properties of interscorer reliability, as well as criterion-related validity and stability of scores over time. In fact, most informal, alternative assessment formats, although possessing face validity, demonstrate low psychometric quality (Gullickson, 1982; Natriello, 1986; Stiggins & Bridgeford, 1985). Only with reliable assessment can judgments be reproduced and be unambiguously communicated to students, parents, and others (Bird, 1990; Linn, Baker & Dunbar, 1991; Valencia & Pearson, 1987).

Alternative assessment formats are often accompanied by scoring methods which depart from the psychometrically well-behaved and well-understood multiple choice selection. Complex tasks may have multiple scorable dimensions, just as tasks which assess the problem-solving process may be scored in multiple stages. Furthermore, a simple additive scoring of these discrete dimensions or stages may yield deceptive results. In addition, different types of score scales may be required for a single test task, e.g., a holistic rating along with a simple objective "0" or "1" score. The reliable scaling and scoring of complex tasks is a continuing concern of assessment specialists, but is yet to yield satisfactory solutions.

Purpose

The purpose of this study was to investigate the use of math manipulative tasks for assessing Hispanic limited English proficient students. The rationale for selecting manipulatives-based assessment was based on four main reasons suggested by the literature, as well as best practices in mathematics classrooms that include Hispanic second language learners and our experiences working within those contexts. First, a manipulatives-based task relies less on verbal linguistic skills. Second, the manipulatives permit concrete representation of abstractions and hands-on problem-solving. Third, a manipulatives-based task is not culturally embedded in either the dominant white culture or the Hispanic minority culture. Fourth, through a manipulatives-based task a student is more likely to publically exhibit the problem-solving process, permitting it, as well as the product, to be assessed and evaluated.

Recognizing that this study was exploratory in nature, we wished to examine different methods of scoring manipulatives-based math tasks. We also wanted to measure the difficulty of these tasks for young Hispanic second language learners. Finally, we were concerned with the traditional psychometric qualities of retest reliability and criterion-related validity. From this initial study, we hoped to establish guidelines for task construction and scoring which could be applied in a more comprehensive study with more tasks and more students.

Research Questions

We posed the following questions of our manipulatives-based math assessment tasks:

1. Retest Reliability: Will Hispanic second language learners earn similar scores on manipulative-based math tasks when retested over a 3 day period?
2. Task Difficulty: How difficult are manipulative-based math tasks for limited English proficient Hispanic learners in kindergarten and the primary Grades 2 & 3?
3. Relationships among Four Criteria Subtest Scores: What patterns of improvement are seen in the four criteria scores (1) conceptual understanding, (2) valid and efficient strategies, (3) accuracy, and (4) fluency intercorrelated?
4. Patterns of Change in Criteria Subtest Scores: What patterns of improvement are seen in the four criteria scores (1) conceptual understanding, (2) efficient valid strategies, (3) accuracy, and (4) fluency over a 3-4 week interval?

5. Math Manipulatives Tasks and General Math Skills: How does performance in the four criteria scores of the manipulatives-based tasks compare to overall math skills as measured by a standardized math achievement test?

6. Math Manipulatives Tasks and Language Proficiency: How does performance on the four manipulative task scores compare to student proficiency in both first and second languages (Spanish and English)?

7. Math Manipulatives Tasks and Gender: How does performance on the criteria scores relate to student gender?

Method

Context

This study occurred in an urban elementary school within a Houston, Texas, school district experiencing rapid growth in Hispanic enrollment. Hispanic students made up 60% of the enrollment of this school, and 78% of the school's enrollment qualified for free or reduce lunch. In this K-4 school, reading and math achievement among the Hispanic population was at the 18^{th} percentile on state norms for the state-mandated Texas Assessment of Academic Skills (TAAS) achievement tests. Only 9% of these students demonstrated the mastery level expected of students at the end of Grade 3.

The elementary school was the site of a three-year federally funded transitional bilingual grant project which focused on mathematics and English competency. This grant funded computers, including printers, other instructional materials, manipulatives, teacher aides, and staff development and training time for bilingual teachers and bilingual teacher aides. Funds were provided for five classrooms, which served as the source for our student respondents. The federally-funded program employed four main components: (l) active learning, (2) math manipulatives, (3) computer-assisted instruction, and (4) parental involvement. As part of the manipulatives component, classroom teachers were trained in the use of math manipulatives, not only in teaching but also in learning centers, focusing on three systems: unix cubes, cuisennaire rods, and base ten blocks. Math learning centers in all classrooms contained these materials. Fluent bilingual teachers and teacher aides worked with students in small and large group instruction. Although the manipulatives were emphasized daily, this part of the curriculum was supplementary to the standard district-adopted basal math program (Spanish edition).

Respondents

From the five special program classrooms (two kindergarten, two Grade 2, one Grade 3) a total of 45 student respondents were selected. Selection was strategic, based on the classroom math functional level of the students as determined by teacher judgment. Nine students from each classroom were selected, two each with high (70th-87th %iles), medium (39th-58th %iles), and low (12th-30th %iles) math performance for their class. We wished to obtain a range of abilities at each grade level to ascertain how the novel manipulatives tasks would function for most students. We also required that students selected be communicative enough to respond verbally in Spanish in a one to-one test situation. For each student an alternate was also selected in case of frequent absenteeism or moving. Because of student attrition and inclusion of alternates, we completed the study with 45 students (18 males, 27 females). Males and females were evenly distributed across general math ability levels. Table 1 shows the distribution of students by gender, and by general math skill level.

Insert Table 1 about here

Table 1 indicates that we completed the study with double the number of kindergarten students as those from Grade 2, or Grade 3. At each grade level, the distribution of the numbers of high, medium, and low achieving students were similar. Students' Spanish and English language competencies had been evaluated within the past 6 months on the Idea Oral Proficiency Test (IPT). Table 2 indicates that no students were fluent in English (FES level), and less than half of the respondents were fluent in Spanish. This deficiency in both languages can be expected to influence learning in all academic areas. Half the students spoke not even basic survival English (NES level), which is why classroom instruction was conducted completely in Spanish, with conversational English taught as a skill area. For all 45 students, Spanish was the home language.

Insert Table 2 about here

One-third of the students had been in the United States less than 3 years. While in Mexico and/or Central America, most of the younger students did not attend school or did so only occasionally. Immigrant children who were age-ready for Grades 2 or 3 often had only sporadic or no prior schooling in Latin America. Thus, in addition to the burden of lack of linguistic fluency, these students were still undergoing the process of acculturation to U. S. society and schools.

Materials

From the supplementary teachers' manuals and workbooks accompanying the cuisennaire rods and base ten blocks, we modified several student learning tasks to create tasks suitable for assessment. Suitable tasks had to meet the following six standards:

1. Tasks embodied important mathematical concepts and principles,
2. Tasks were appropriate for low-age-to-average math skill levels in kindergarten through Grade 3,
3. Tasks had not been directly taught or practiced by teachers.
4. Both the process and product of task solution was visible and appeared to be scorable,
5. Tasks appeared sufficiently challenging for the designated grade level so we could anticipate multiple practice occasions for each student before mastery, and
6. Tasks could be efficiently completed with a minimum of teacher directions, and with relatively few standard materials.

We began the study by developing 35 tasks; however, after three months of pilot-testing that number was reduced to 14 key test tasks. Through pilot testing, we identified inaccuracies in the initial designation of Grade level tasks. Of the 14 final tasks, 7 were most appropriate for kindergarten, only two for Grade 2, and five for Grade 3.

Had we been able to include Grade 1, several of the Kindergarten tasks would likely have been more appropriate there. However, because we conducted final testing in March/April, several of the Kindergarten students were performing similarly to Grade 1 students. The 14 tasks are named in Table 3, with appropriate grade level and key mathematical concepts, principles, and skills embedded in each task. Illustrated directions and materials for each task are included as Appendix I.

Insert Table 3 about here

Task Scoring

Student task performance was scored according to four criteria, covering both the problem-solving process and the final solution. Each criterion received rating on a 1-5 ordinal scale. The first three criteria used the same scale, but the fourth criterion required a different scale. Following are the four criteria and their scales: (a) conceptual understanding, (b) use of valid, effective strategies, and (c) accuracy of task completion -- [Scale: 1 = none or almost none, 2 = only a little, 3 = some, partially, 4 = most, mainly, 5 = fully, complete]; and (d) fluency and speed of task completion -- [Scale: 1 = very slow & disfluent, 2 = somewhat slow & disfluent, 3 = adequate/average speed & fluency, 4 = good speed & fluency, 5 = very good/excellent speed & fluency.]

Because of the multiple scoring criteria, all students were required to attempt the full task. For this reason, if students became stuck, they were given minimal prompts, in the form of questions, to help them focus on relevant task features. All such examiner prompts were recorded on the score sheet, and served to negatively weight the final ratings.

Procedure

For three months, two fluent bilingual graduate students in a Master of Education Program (M.Ed.), piloted the manipulatives-based tasks with 15 Hispanic second English language learners in a different elementary school in a different school district. This pilot group mirrored the main sample in that it contained similar numbers of high, medium, and low achieving students, as nominated by teachers. During the pilot phase we improved the instructions, task structure, and scoring procedures. Also during this pilot phase, inter-rater reliability was assessed over a three-day period with the 15 students, using the final versions of the tasks and score sheets. For retest reliability, each student was assessed twice with four different tasks which were selected with the assistance of the classroom teacher to be challenging for that student.

Following the pilot-testing and revisions, the test-tasks were presented to the selected 45 respondents. All directions were in Spanish. Three-to-four challenging tasks were chosen for each student, depending upon the speed with which students were able to work. The three or four tasks were presented to each student individually within a single session. Administration of each task required approximately 3-5 minutes. After a 2-to-3 week break, equivalent versions of the same tasks were re-administered to all students.

Results

Retest Reliability

Retest reliability was assayed during the pilot phase of the study with a small group of 15 students. Our expectations for retest reliability were different for these manipulatives-based tasks than they would have been for standard test items. To the extent that the manipulatives-based assessment was "dynamic" ; i.e., the examiner provided prompts to the student, students might be expected to learn from the examination. This learning phenomenon might occur even though the student did not receive feedback on the correctness of his/her final response. In fact, a professed advantage of dynamic assessment is the students' ability to learn while being assessed. (Lidz,1987). To the extent that our students might learn differentially from examiner prompting, we expected lower retest reliability. Retest reliability results are presented in Table 4. This table indicates that reliability coefficients obtained from simple rank order (Contingency Coefficient) were moderately strong (.69 - .77). However, when corrected for scaling irregularities (Cramer's V) the coefficients were lower (.48 - .71). Finally, when the level of agreement was conservatively corrected for all chance agreement through Cohen's Kappa (Cohen,

1960) the resulting coefficients varied from strong to weak (.74 - .32). Kappa values between .40 and .75 represent fair-to-good agreement beyond chance (Fleiss, 1989). Table 4 indicates that the retest reliability varied depending on grade level and/or on which test-tasks were used (these two variables were confounded). The lower two grade levels (Kindergarten, Grade 2) produced weaker retest reliability.

Insert Table 4 about here

Task Difficulty

The second research question asked how difficult these manipulatives-based tasks were for low-achieving LEP students. The published literature fails to offer such basic information, although manipulatives-based tasks are commonly recommended. We examined task difficulty separately for the first and second assessments, spaced 3-4 weeks apart. Here, our index of task difficulty was a simple average of all four criteria (conceptual understanding; use of valid and effective strategies; accuracy of task completion; and fluency & speed of task completion). Results are presented in Table 5.

Insert Table 5 about here

Manipulatives-based math tasks proved challenging for the students. For the first assessment, overall task difficulty was similar for the three grade levels, closer to a middle rating of 3 (some, partially) than a rating of 4 (most, mainly). (Kindergarten = 3.04; Grade 2 = 3.30 and Grade 3 = 3.25). Students found the tasks somewhat less difficult during the second assessment. (Kindergarten = 3.37, Grade 2 = 3.73, and Grade 3 = 3.62).

Of the 14 tasks, students' average performance improved on twelve, and deteriorated on two. Of the twelve improved tasks, the skill improvement was statistically significant for only six. The effect sizes of these six gains ranged from small (one third of a standard deviation) to large (nearly a full standard deviation). An effect size of .40 to .50

or larger is considered large enough for practical significance in most educational achievement (Hedges, & Olkin,1985; Wolf, 1986).

The significantly improved tasks were not from one single Grade level. The two deteriorated tasks were from Kindergarten and Grade 3. Task improvement could not be predicted from the initial performance level.

Relationship among Four Subtest Scores.

More than one judgment scale are required to score both a student's answer or product and the problem-solving process he or she used. Although multi-scale scoring is commonly recommended, published literature on alternative assessment fails to describe relationships among separate scales.

Intercorrelations among the performance ratings of the four criteria were examined (conceptual understanding, use of valid, & effective strategies, accuracy of task completion, and fluency & speed of task completion). Prior to analysis, all scores were standardized within tasks, to control for differences in difficulty across tasks. Hierarchical cluster analyses of intercorrelation matrices were then conducted both within and across tasks. The cluster trees in Figure 1 show similar results for different tasks within a grade level, but quite different cluster patterns from one grade level to the next.

Insert Figure 1 about here

At Kindergarten level, the four criteria skill ratings were all strongly intercorrelated. The two most cohesive or "tightest" clusters were : (a) fluency + efficient strategies, and (b) conceptual understanding + accuracy. At Grade 2, the four skills were moderately to strongly intercorrelated. The Grade 2 clusters differ from those for kindergartners. In Grade 2, fluency + accuracy formed the tightest cluster, followed by Conceptual understanding + valid and efficient strategies. At Grade 3 only one very tight cluster was noted, and it was fluency + accuracy (r = .96). The other two criteria were only moderately correlated. In general, there appears to be greater differentiation of component skills as one proceeds through the primary Grade levels. In addition, as one proceeds up the grade levels, the relationship between Fluency and Accuracy becomes stronger, i. e., students who solved tasks rapidly and fluently are also more likely to obtain the correct answer.

Patterns of Change in Criteria Subtest Scores

We earlier examined the overall improvements of average ratings from first to second assessment. The current question examines patterns of improvements among the four criteria subtest scores. Do ratings of the four criteria scores change together, or do improvements in one criteria tend to lead the others? The obtained patterns are described graphically in Figure 2. Figure 2 depicts for each grade level the criterion score changes between first and second testings. Error bars mark 85% confidence intervals, based on standard error of the mean. Score patterns at the three Grade levels bear similarities and differences. At all Grade levels students were rated highest for conceptual understanding. At Kindergarten, fluency was rated next higher, and efficient strategies lowest. At Grade 2, valid and efficient strategies, accuracy, and fluency were rated similarly. At Grade 3, efficient strategies was rated second highest, followed by accuracy and fluency.

Insert Figure 2 about here

Patterns of improvement from first to second testing appear relatively consistent across skills and grade levels. A minor exception is the relatively greater mean gain in accuracy and fluency by Grade 3 students. However, these mean gains must be interpreted in light of more highly dispersed score distributions (larger standard error of the mean).

Subskill Performance by General Math Skills

For a new assessment tool to demonstrate criterion-related validity, it must relate well to established measures of a similar nature. However, when the target and criterion tests require substantially different types of performance, smaller validity coefficients generally result. To relate general math levels to manipulatives-based task scores, we first standardized the scores within each grade level to control for between-grade differences in task difficulty. Then all grade levels were combined for analysis. We conducted a mixed (repeated measures & factorial analyses) ANOVA, with the criteria scores serving as repeated measures (with four levels) and general math level as the categorical grouping variable with three levels (High = 15, Medium = 16, Low = 14). The ANOVA resulted in a large main effect for general math skill level, [F(2 df) = 6.209, (p = .0043]. The main effect for criteria score type was not significant, [F(3 df) < 1, (p = .99]. No significant interaction was noted between these two variables, [F(6 df) = .294, (p = .93)]. Figure 3 presents the interaction graph for this analysis. Figure 3 also demonstrates that the manipulatives subtest scores neatly separate low general math skill levels from medium and high levels. Interestingly, students at the medium skill level were the top performers in using efficient strategies, although a post-hoc analysis proved this difference between medium and high groups to be too small for statistical significance. Figure 3 also demonstrates relatively flat performance across the four criteria subtest ratings at each Grade level, especially for students in the low achieving group.

Insert Figure 3 about here

Subtest Performance by Language Proficiency

Manipulatives-based tasks are recommended for LEP students in part because the tasks appear to be less verbally loaded, and thus permit demonstration of non-verbal math skills. The students in this study, scored low on both English and Spanish abilities; however, a range of language skills did exist. We, therefore, were able to conduct a mixed ANOVA, with the criteria scores as repeated measures and language skills as the categorical grouping variable (five levels). The five language skill groups were 1. FSS/LES (n = 5), 2. FSS/NES (n = 13), 3. LSS/LES (n = 12), 4. LSS/NES (n = 11), 5. NSS/NES (n = 4).

Interestingly, the analysis showed no main effect for language, [F(4 df) = 1.12, (p = .36)]. Nor did a main effect exist for subtest scores, [F(3 df) = .0001, (p = .99)]. However, a significant interaction did exist; different criteria score patterns were revealed for different language groups, F(12 df) = 3.003, (p = .001)]. Figure 4 presents the interaction graph for this analysis.

Insert Figure 4 about here

All manipulatives-based tasks were presented in students' first language, Spanish. We expected that because manipulatives-based tasks were presented in Spanish, a student's level of Spanish proficiency would predict math-task performance. However, Figure 4 shows clearly that Spanish fluency alone relates only randomly to any one of the four criteria scores. We further expected that because English is the main academic language, even in Texas's highly bilingual districts, fluency in English would also predict math-task performance. However, Figure 4 also shows that a simple English proficiency level fails to predict any criterion score except for Fluency. For the Fluency rating, students with limited English skills (LES) outperformed students with no English skills (NES).

Because of the lack of predictive power of simple English or Spanish proficiency, we next examined patterns of English & Spanish combined proficiency. Here, again, Figure 4 depicts counterintuitive results. On the manipulatives-based tasks, the lowest overall performers were not those students with lowest combined language skills. In fact, the overall lowest performing students were fluent in Spanish, with no English skills (FSS/NES), and the highest performing students (by a slim margin) were those without fluency in either language (NSS/NES). This result was so surprising that we individually examined four individual NSS/NES student cases from our database to help explain this phenomenon. Of these four students, one tested with high math skills, one medium, and two low. All four of these students were females, two in Grade 2 and two in Grade 3. Only one student was a recent arrival to the US. (1.75 years). Two of the four had been in the US approximately 3.5 years, and one student was born here. Figure 4 also shows that these four NSS/NES students exhibited a great range in subscores, with efficient strategies the highest among any group, and fluency only average.

Two groups of students performed nearly identically: FSS/LES and LSS/LES. which seems to indicate the relative unimportance of Spanish fluency on these academic tasks. given at least limited English proficiency.

Subtest Performance by Gender

The field of math education has a considerable body of knowledge on the underperformance of female students (Hanson, 1992; Olszeweski-Kubilius, 1990). Among the Hispanic population, this underachievement is especially pronounced, possibly because the culture has tended to emphasize female submissiveness (Heller, 1966; Madsen. 1961). Furthermore, some evidence exists that whereas females exhibit relatively stronger verbal skills in solving academic tasks, males exhibit stronger problem-solving skills on three-dimensional, engineering-type tasks. (Sadker & Sadker, 1994) Based on these generalizations, we predicted that the manipulative-based tasks would favor the male students.

The student respondents included 27 females and 18 males. Gender was distributed nearly identically across general math skill levels (see Table 1), so a gender-based comparison on manipulatives-based tasks could be made without concern that the general math skill level would confound the comparison. We conducted a mixed ANOVA, with the subtest scores as repeated measures and gender as the categorical grouping variable. The ANOVA resulted in no main effect for gender, [F(1 df) = .80. (p = .37)]. Neither was there a main effect for type of criteria subtest score, [F(3 df) = .05, (p = .98)], nor was the interaction significant, [F(3 df) = .07, (p = .97)]. The interaction graph from this ANOVA is presented in Figure 5. It indicates that average subtest scores for males were consistently higher than for females, but not significantly so. In every case, the 85% confidence interval bars overlap with mean scores. We concluded that there was a consistent, but non-significant gender effect on math manipulatives-based test scores.

Insert Figure 5 about here

Discussion

This study was an exploratory study of the use of manipulatives-based mathematics assessment with 45 Hispanic limited English proficient students in Kindergarten, Second and Third Grades. According to published literature, such manipulatives-based tasks potentially offer several benefits: (a) abstract concepts can be demonstrated in concrete terms, (b) multi-modalities (including tactile and kinesthetic) are applied in problem-solving, (c) the problem-solving process as well as the final answer are scored, and (d) students without solid language proficiency have a better chance at success. Despite these claims, little empirical evidence exists on manipulatives-based assessment with Hispanic LEP students. We were especially interested in the psychometric behavior of these tasks -their reliability and validity, and how Hispanic LEP students performed. Although the student sample was small, it was strategically selected to permit several revealing analyses.

Constructing suitable, scorable tasks required considerable pilot-testing which yielded only 14 usable tasks from an initial set of 35. Thus the conversion from teachable task to testable task is not automatic. In addition, several initial tasks appeared upon closer scrutiny to be superficial; i. e., they did not reflect important, relevant math concepts or processes worthy of assessment. Moderately strong retest reliability was obtained on 12 of these tasks over a 3-day retest period, although somewhat lower reliability was obtained for Kindergarten students than for older students. The obtained reliability was better than we had anticipated, given the different levels and types of cues needed by our examiners for students to complete the tasks. Perhaps our greatest scoring dilemma continues to be the reliable treatment of idiosyncratic cues, which must occur in this type of dynamic assessment.

The manipulatives-based tasks proved challenging for Hispanic LEP students, despite the fact that they had practiced using the manipulatives in the classroom and had continuous access to them. The students improved significantly on six of the 14 tasks over a retesting period of 2-3 weeks, without targeted instruction. This improvement could have been due to the occasional prompts received by students during the first testing or to student-initiated discovery back in the classroom.

The research question on interrelationships among the four criteria scores seemed to favor a developmental hypothesis. The sub skills changed from being tightly interrelated in highly differentiated in Grade 3. In fact, the combination of Fluency and Accuracy became more tightly linked, and approached unity ($r = .96$) in Grade 3. These findings caution against cross-Grade pronouncements about how students react to or use math manipulatives. In Grade 3, Fluency implied Accuracy, to the extent that the separate scores were largely redundant. However, in Kindergarten, Fluency was less likely to imply Accuracy. Thus, manipulatives tasks may function differently at different developmental stages, and future research should be sensitive to that probability.

We were also interested in how sub skills changed -- together or unevenly, with one sub skill leading the others. Two assessment periods did not provide an ideal opportunity for profiling change over time; however, we were able to draw some gross conclusions. If skills change unevenly, perhaps one subskill could serve as diagnostically useful advance notice of later learning. An orderly acquisition of subskills might also provide suggestions for an instructional sequence or changing focus. However, our findings were not supportive. Changes from first to second assessment were relatively even across the four criteria subscores. Only some differentiation appeared in Grade 3, where relatively greater gains were made in Fluency and Accuracy compared to Efficient Strategies and Conceptual Understanding. This phenomenon might be described as students learning how to quickly arrive at the solution without concomitantly showing greater understanding. In that case, students might be treating the manipulatives as artificial tools rather than as conceptually meaningful as intended by the examiners. On the question of criterion-referenced validity of the manipulatives-based math tasks, the evidence was only mildly supportive. The criterion measure was a traditional, mainly multiple-choice, broad assessment of math skills which had been district-normed. However, the manipulatives-based tasks were not able to differentiate medium (39th - 58th %iles) from high (70th - 87th %iles) were clearly differentiated from medium and high scores, may suggest that manipulatives-based tasks are most sensitive to the abilities of low-achieving students and are best used with them.

The lack of relationship between either English or Spanish language fluency and math ability was found to be interesting and confirmatory of those researchers who claim an advantage of manipulatives to be that they are not as verbally dependent as other tasks. On the other hand, these results might be considered disconfirmatory of those who claim lack of academic achievement to be largely due to the lack of language fluency. However interpreted, students without fluency in either English or Spanish scored highest or nearly highest in three of the four criteria subskills. Further investigation showed that all of these atypical students were females, and that all had been in the US for more than 1.75 years. Although tasks were administered in Spanish, it mattered little to scores whether students were fluent in Spanish or possessed only limited Spanish skills.

The examination of the role of gender in manipulatives-task performance revealed small and insignificant, but consistent results. Although boys and girls showed virtually identical average performance on the district's criterion-referenced math test, the manipulatives-based test slightly favored boys, across all subtests. We had hoped not to find a gender bias built into the manipulatives-based tasks, but had noted that their manipulation may reflect the construction play which is more often modeled for boys, and which boys are therefore more likely to practice. Manipulatives also require a modicum of physical forwardness or initiative in reaching out to take and manipulate. Some young girls observed were very hesitant and shy about so doing. Less shyness may have been present had the classroom teachers or aides, themselves, conducted the assessments. On the other hand, the examiners were frequent visitors to these classrooms. Furthermore, confirmatory remarks by teachers indicated that the performances of the students tended to be typical.

An initial investigation into the behavior of novel manipulatives-based math tasks, the current study raised more questions than possible answers, Even in the case of clear-cut results, this study's number of participants (N), was too small to draw firm and valid conclusions. Yet, some commonly held assumptions were not confirmed by this study: (a) students' language skills were largely irrelevant to task performance, (b) the tasks were not well validated (by an external measure) for medium and higher achieving students, (c) criteria subskills of conceptual understanding, efficient use of strategies, accuracy, and fluency all appeared to improve in unison, rather than in a logical sequence, (d) these same criteria subskills did not covary the same way across the three Grade levels. For example, the fluency subscale may be measuring something very different at Grade 3 versus Kindergarten.

Given the fact that we were able to obtain moderate interrater reliability in scoring a number of manipulatives-based tasks, and could efficiently administer these tasks to the sample of Hispanic LEP students, indicates that these types of tasks are amenable to experimental study. We are concerned about the vagaries of small samples, and seek and recommend a replication study with a separate and larger sample. As an exploratory study, it has served its purpose, raising several issues with classroom implications. However, at this point, to discuss implications in detail would be overstating the case; first, replication is needed.

References

Alexander, L., & James, H., T. (1987). The Nation's Report Card. Cambridge, MA: National Academy of Education.

Bird, T. (1990). The school teacher's portfolio: An essay on possibilities. In J. Milan & L. Darling-Hammed (Eds.). The New Handbook of Teacher Evaluation (pp. 241-256). Newborn Park, CA: Sage.

Bohan, H. (1995). Math Manipulatives. Mathematics. Boston, Massachusetts: Houghton Mifflin.

Briars, D. J., & Larkin, J. H. (1984). An integrated model of skills in solving elementary word problems. Cognition and Instruction 1, 245-296.

Calfee, R., & Hiebert, E. (1991). Classroom assessment of reading. In R. Barr, M. L., Kamil, P. Mosenthal, & P. D. Pearson, Handbook of reading research (pp. 281-309). New York: Longman.

Carpenter, T. P., & Moser, J. M. (1982). Young Children are Good Problem Solvers. Arithmetic Teacher. 30, 24-26.

Carpenter, T. P., & Moser, J. M., (1983). The acquisition of addition and subtraction concepts. In R. Lesh & M. Landau (Eds.). Acquisition of Mathematics: Concepts and Processes (pp 7-44). New York: Academic Press.

Carpenter, T. P., & Moser, J. M., (1984). The acquisition of addition and subtraction concepts grades one through three. Journal of Research in Mathematics Education. 15(3) 179-202

Carpenter, T. P., Fennema, E., Fuson, K., Heibert, J., Human, P., Murray, H., Olivier, A. & Wearne, D. (1994). Teaching Mathematics with Understanding in the Primary Grades. Unpublished paper presented at the annual meeting of the American Educational Research Association. New Orleans, LA.

Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20, 37-46.

Cole, M., & Griffin, P. (Eds.). (1987). Contextual factors in education: Improving Science and Mathematics Education for Minorities and Women. Madison, WS: Center for Education Research, University of Wisconsin, Madison.

Cziko, G. A. (1992). The evaluation of bilingual education: From necessity and probability to possibility. Educational Researcher, 21, 10-15.

D'Ambrosio, U. (1987). Etnomatematica: Raizes Socio-Culturales da Arte ou Tecnica de Explicar e Conhece. Campaigns, UNICAMP.

Dale, T. C., & Cuevas, G. A. (1987). Integrating Language & Mathematics Learning ESL through Content-Area Instruction: Mathematics, Sciences, Social Studies, Language. Education Theory and Practice. No 69, Office of Educational Research and Improvement, ERIC Document. Reproduction Service No. ED283387.

Darling-Hammed, L. (1991). Implications of testing policy for quality and equity. Phi Delta Kappan, 73, 220-225.

Dean, A. L., & Malik, M. M. (1986). Representing and Solving arithmetic word problems: A study of developmental interaction. Cognition & Instruction, 3, 221-228.

Dossey, J. A., Mullis, I. V. S. Lindquist, N. M., & Chambers, D. L. (1988). The mathematics report card: Are we measuring up? Princeton, New Jersey: Educational Testing Service, National Assessment of Educational Progress.

Farr, R. (1992). Putting it all together: Solving the reading assessment puzzle. Reading Teacher, 46, 26-37.

Fleiss, J. L. (1981). Statistical methods for rates and proportions New York: Wiley.

Freeman, D., J., Kuhs, T. M., Porter, A., C., Floden, R. F., Schmidt, W., H., & Schwille, J., R.(1983). Do textbooks and tests define a national curriculum in elementary school mathematics? Grouws, D. A. & Good, T. L. (1989). Issues in problem-solving instruction. Arithmetic Teacher, 36(8), 34-35.

García, E. (1988). Attributes of Effective Schools for Language Minority Students. Education and Urban Society, 2, 387-398

Garcia, E. (1991). The education of linguistically and culturally diverse students: Effective instructional practices. (Educational Practice Report No. 1). Santa Cruz, CA and Washington, DC: National Center for Research on Cultural Diversity and Second Language Learning.

Gentry-Norton, B. (1995). Assessment. Mathematics. Boston, Massachusetts: Houghton Mifflin..

Gullickson, A. R., (1982). The Practice of Testing in Elementary and Secondary Schools. ERIC Document. Reproduction Service No. 229391

Hambleton, A. R., & Swaminathan, H. (1985). A Look at Psychometrics in the Netherlands. Paper presented at the joint meeting of the AERA and the National Counsel on measurement in Education. Chicago, Illinois, ERIC Doc. Service No. 273665

Hanson, K. (1992). Teaching mathematics effectively and equitably to females. Trends and Issues 17, ERIC Clearinghouse on Urban Education. Teachers College, Columbia University: New York.

Haycock, K., & Navarro, M. S. (1988). Unfinished Business: Fulfilling our Children Promise. A Report for the Achievement Counsel, ERIC Document. Reproduction Service No. 299025

Hedges, L. V., & Olin, I. (1985). Statistical methods for meta-analyisis. Orlando: Academic Press.

Heller, C.S. (1996). Mexican-American youth: Forgotten youth at the crossroads. New York: Random House.

Henderson R. W., & Landesman, E. M. (1992). Mathematics and middle school students of Mexican descent: The effects of Thematically integrated instruction. University of California, Santa Cruz. National Center for Research on Cultural Diversity and Second Language Learning.

Howard, K. (1990). Making the writing portfolio real. Quarterly of the National Writing Project and the Center for the Study of Writing and Literacy, 12, 4-7, 27.

Kean, M. H. (1992). ESEA Chapter I re-authorization: Testimony before the advisory committee on testing in Chapter I. New York: Association of American Publishers.

Kennedy, M. M., Birman, B. F.,. & Demaline, R. E. (1986). The effectiveness of chapter I services. Washington, DC: Government Printing Office.

Kintsch, N., & Greeno, J. G. (1985). Understanding and Solving Word Arithmetic Problems. Psychological Review, 92, 1, 109-129

LaCelle-Peterson, M, & Rivera, C. (1994). Is it really real for all kids? A framework for equitable assessment policies for English language learners. Harvard Educational Review 64, 1, 55-75.

Lam, T. C. M. (1992). Review of practices and problems in the evaluation of bilingual education. Review of Educational Research, 62, 181-203.

Laosa, L. M., & Henderson, R. W. (1991). Cognitive socialization and competence: The academic development of Chicanos. In R. Valencia (ED.), Chicano school failure and success: Research and policy agendas for the 1990s. London: Falmer.

Lidz, C. S. (1987). Dynamic Assessment. New York: The Guilford Press.

Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Complex performance-based assessments: Expectations and validation criteria. Educational Researcher, 20(8), 15-21.

MacCorquodale, P.(1988). Mexican-American women and mathematics: Participation, aspirations, and achievement. In R. R. Cocking & J. P. Mestre (Eds.). Linguistic and cultural influences on learning mathematics. (137-160).Hillsdale, NJ: Eribaum.

Madsen, W. (1961). Society and health in the lower Rio Grande valley. Austin, Texas: Hogg Foundation for Mental Health.

Massey, W. E. (1989). Science education in the United States: What the scientific community can do. Science, 2, 45, 915-921

Moll, L. D., & Diaz, S. (1987). Change as the goal for educational research. Anthropology and Education Quarterly. 18, 300-311.
Murnane, R. J., & Raizen, S. A. (1988). Improving indicators of the quality of science and mathematics education in grades K-12. Washington, DC: National Academy Press.
Natriello, G. (1986). School Dropouts: Patterns and Policies. New York: Teachers College Press, Columbia University.
National Council of Teachers of Mathematics. (1995). Assessment Standards for School Mathematics. Reston, VA: The Council.
Nunes, T. (1992). Cognitive invariants and cultural variation in mathematical concepts. International Journal of behavioral Development 15, 4, 433-53.
Olszewski-Kubilius, P. (1990). Predictors in achievement in mathematics for gifted males and females. Gifted Child Quarterly. 34 (2), 64-71.
Orr, E. W. (1987). Twice as less: Black English & the Performance of Black students In Mathematics and science. New York: Norton.
Orum, L. S. (1986). The education of Hispanics: Status and implications. Washington, DC: National Council of La Raza.
Pimm, D. (1987). Speaking mathematically: Communication in mathematics classrooms. New York: Routledge & Kegan Pau.
Policy Analysis for California Education, (1990). Conditions of Education in California. Berkeley: Author.
Reilly, T. (1988). ESL through content-area instruction. ERIC Digest. Washington, D C.: Center for Applied Linguistics.
Riley, M. S., & Greeno, J. G. (1983). Development of children's problem solving ability in arithmetic. In H. Ginbsburg (Ed.). The development of mathematical thinking. (pp. 153-200). New York: Academic Press.
Riley, M., & Greeno, J. G. (1988). Developmental analysis of understanding language about quantities of solving problems. Cognition & Instruction, 5,1, 49-101.
Riley, M. S., Greeno, J. G. & Heller, J. I. (1983). Development of children's problem-solving ability in arithmetic. In H. Ginsburg (Ed.). The Development of Mathematical Thinking, 153-196. New York: Academic Press.
Romberg, T. A., Zarinnia, E. A., and Collis, K. F. (1990). A new world view of assessment in mathematics. In G. Kulm (Ed.). Assessing higher order thinking in mathematics. (pp. 21-38). Washington, DC: American Association for the Advancement of Science.
Sadker, M., & Sadker, D. (1994). Failing at Fairness. New York: Macmillan Publishing Company.
Salmon-Cox, L. (1981). Teachers and standardized achievement tests: What's really happening? Phi Delta Kappan, 62, 631-634.
Schmidt, W. H. (1983). Content bias in achievement tests. Journal of Educational Measurement, 20, 165-178.
Shepherd, L. (1991). Psychometricians beliefs about learning. Educational Researcher, 21(4), 2-27.
Simmons, W. (1985). Social class and ethnic differences in cognition: A cultural practice perspective. In S. F. Chipman, J. Segal, & R. Glaser, Thinking and learning skills. (Vol. 2. pp. 519-553). Hillsdale, New Jersey: Eribaum.
Smith, M. L. (1991). Put to the test: The effects of external testing on teachers. Educational Researcher, 20(5), 8-11.
Stiggins, R. J., & Bridgeford, N. J. (1985). The ecology of classroom assessment. Journal of Educational Measurement, 22(4), 271-286.
Stiggins, R. J., Conklin, N. F., & Bridgeford, N.J. (1986). Classroom assessment: A key to effective education. Educational Measurement: Issues and Practice. Summer, 5-17.
Stigler, J. W. & Barnaces, R. (1988). Culture and mathematics learning. In E. Z. Rothzopf (Ed.) Review of research in Education (Vol. 15, pp. 253-306). Washington, DC. American Education Research Association.

Stone, C. A. (1989). Testing software review: Micro CAT Version 3.0. Educational Measurement: Issues and practice, 8(3), 33-38.
Tharp, R. G. (1989). Psychocultural Variables and Constants: Effects on Teaching and Learning in Schools. American Psychologist, 44, 2, 1-11
Tikunoff, W. J. (Ed.). (1983). Teaching in successful bilingual instructional settings. San Francisco, CA: Far West Laboratory for Educational Research & Development.
Tindal, G. & Marston, D. (1990). Classroom-based assessment. Columbus, OH: Merrill.
Trueba, H. (1988). Culturally based explanations of minority students' academic achievement. Anthropology and Education Quartely, 19, 270-287
United States Department of Education. (1992). The condition of bilingual education: A report to the Congress and the President, Washington, DC.
Valencia, R. R. (1991). The plight of Chicano students: An overview of schooling conditions and outcomes. In R. R. Valencia (Ed.), Chicano school failure and success: Research and policy agendas for the 1990s. London: Falmer.
Wixson, K. K., Valencia, S.W., & Lipson, M.Y. (1994). Issues in literacy assessment: Facing the realities of internal and external assessment. Journal of Reading Behavior, 26, 315-337.
Wolf, F. M. (1986). Meta-analysis: Quantitative methods for research synthesis. Beverly Hills, California: Sage.
Wong Fillmore, L., Ammon, P., McLaughlin, B., & M. Ammon, (1985).
Learning English through bilingual instruction. Final report submitted to the National Institute of Education, Berkeley: University of California. ERIC Document Reproduction Service No. ED 259579

Table 1

Students Respondents, with Gender and Achievement Levels on District-standardized Criterion Referenced Math Test.

Gen. Skills*	Kindergarten	Grade 2	Grade 3	Totals & Gender
High	9	3	3	15 (9 F, 5 M)
Medium	8	4	4	16 (9 F, 7 M)
Low	9	3	2	14 (9 F, 6 M)
	26	10	9	45 (27 F, 18 M)

*General math skills from school district CRT standardized math test.
*High= (70th-87th %iles), Medium= (39th-58th %iles), Low= (12th-30th %iles)

Table 2

Spanish and English Proficiencies of Respondents, From Recent IPT Testing

Spanish / English*	Kindergarten	Grade 2	Grade 3	Totals
FSS / LES	4	0	1	5
FSS / NES	6	2	5	13
LSS / LES	8	4	0	12
LSS / NES	8	2	1	11
NSS / NES	0	2	2	4
	26	10	9	45

*NSS = Non Spanish Speaker, LSS = Limited Spanish Speaker, FSS = Fluent Spanish Speaker, NES = Non English Speaker, LES = Limited English Speaker, FES = Fluent English Speaker

Table 3

Fourteen Manipulative Test-Tasks, with Grade Level and Key Mathematical Concepts, Principles, Skills & Computation

TASK	CONCEPTS & PRINCIPLES	SKILLS & COMPUTATION
KINDERGARTEN		
1. Making Stair-steps.	•ordinal relationship •increments of one •fixed volume increments	•visual scanning for approximate rod length •compare & contrast rods •sort by size
2. Equal trains.	•equivalence in rod combinations •addition of two quantities to make a new whole	•addition •selecting rods to fit missing part of whole
3. Find the missing rod.	•missing part to make a whole •increase in quantity and/or difference in quantities	•addition or subtraction •visual scanning for approximate rod length •measuring rods
4. Which is more?	•increase in quantity to make a new whole (equivalence) •"more" related to rod length. •combinations form a whole •number sense - quantity differences	•measuring rod combinations •compare/contrast rods
5. Finding "one-bigger".	•increment of one and/or addition of one •counting related to increment of one	•rod length-name association. •measuring rods •addition
6. Find the length of a train.	•relation between rod combinations & continuous counting •one-to-one relationship •increase in quantity to make a new whole	•selecting rods by name •use of numberline to find total length •addition counting (numeration)
7. Half the size.	•"half the size" •"doubling" to make a whole •equivalent fractions •addition of two quantities to make a new whole	•identifying halves of even numbers 2 - 10 •addition or dividision
GRADE TWO		
8. Missing addends.	•concept of missing addend as missing rod to total •tying concrete model to abstract equation •addition and/or subtraction of quantity needed to form equal whole	•solving missing addend equation •addition and/or subtraction •scanning for quantity needed to complete whole

Table 3 (continued)

Fourteen Manipulative Test-Tasks, with Grade Level and Key Mathematical Concepts, Principles, Skills & Computation

9. Ten more than.	•concept of "ten more than" as increment to tens place •increase base ten values by tens	•ability to read tens and ones rods as two-digit number •addition •counting
GRADE THREE		
10. Writing Three-Digit Numbers.	•relation between numeral place and particular size rods •concrete representation of digital quantity •base ten place value (hundreds, tens, ones)	•writing three-digit number from place value model •making correct value substitution
11. Simple Two-Digit Addition.	•relation between rod size and place value •base ten place value •add to quantities to form new whole (base ten) •tying abstract and concrete: column addition as combining similar rods	•adding two-digit numbers in column addition
12. Two-Digit Addition with Regrouping.	•relation between regrouping in addition and place value •Base ten place value •Add to quantities to form new whole (base ten) •Tying concrete and abstract representation of increase in quantity •regrouping to represent correct value	•two-digit addition in columns, with regrouping •substitutions of units for tens
13. Trading Up.	•base ten place value substitutions: equivalence of ten units to one 10 & ten 10s to one hundred •concrete representation of quantity using base ten values	•making value substitution: regroup from ones to tens and tens to hundreds
14. Two-Digit Subtraction with Regrouping.	•base ten place value •subtract to quantities to form new whole (base ten) •regrouping to represent correct value •relation of backward regrouping in subtraction to place value	•two-digit column subtraction numbers with borrowing •substitutions of tens for units

Table 4

Retest Reliability of Twelve Manipulatives-based Tasks with Young Hispanic ESL Students (N=15).

Grade	N	CC*	Cramer's V	Cohen's Kappa
Kindergarten	6	.69	.48	.40
Grade 2	5	.74	.63	.32
Grade 3	4	.77	.71	.74

*Contingency Coefficient

Table 5

Mean Ratings and Improvement for 14 Manipulatives-Based Tasks Over a 3-4 Week Interval

TASK	1st M	SD	2nd M	SD	p*	ES**
KINDERGARTEN						
1. Making Stair-steps.	3.76	0.94	3.60	1.21	*NS*	-.17
2. Equal trains.	3.69	1.17	3.88	1.10	*NS*	.16
3. Find the missing rod.	3.53	1.22	3.88	0.74	*.07*	.29
4. Which is more?	3.45	1.21	3.84	0.82	*.03*	.33
5. Finding 'One bigger'	3.44	1.20	3.48	1.25	*NS*	.04
6. Find length of a train.	2.14	1.20	3.10	1.16	*.00*	.80
7. Half the size.	1.25	1.01	1.84	1.16	*.06*	.59
GRADE TWO						
8. Missing addends.	3.78	0.46	4.17	0.18	*.01*	.85
9. Ten more than.	2.83	1.46	3.29	1.42	*NS*	.32
GRADE THREE						
10. Write 3-digit numbers.	3.85	1.14	4.26	0.36	*NS*	.36
11. Simple 2-digit addition.	3.41	1.36	4.06	0.35	*NS*	.47
12. Two-digit addition with regrouping.	3.33	0.94	3.63	0.78	*NS*	.31
13. Trading up.	2.93	0.83	3.52	0.65	*.06*	.71
14. Two-digit subtraction with regrouping	2.75	0.93	2.64	1.38	*NS*	-.12

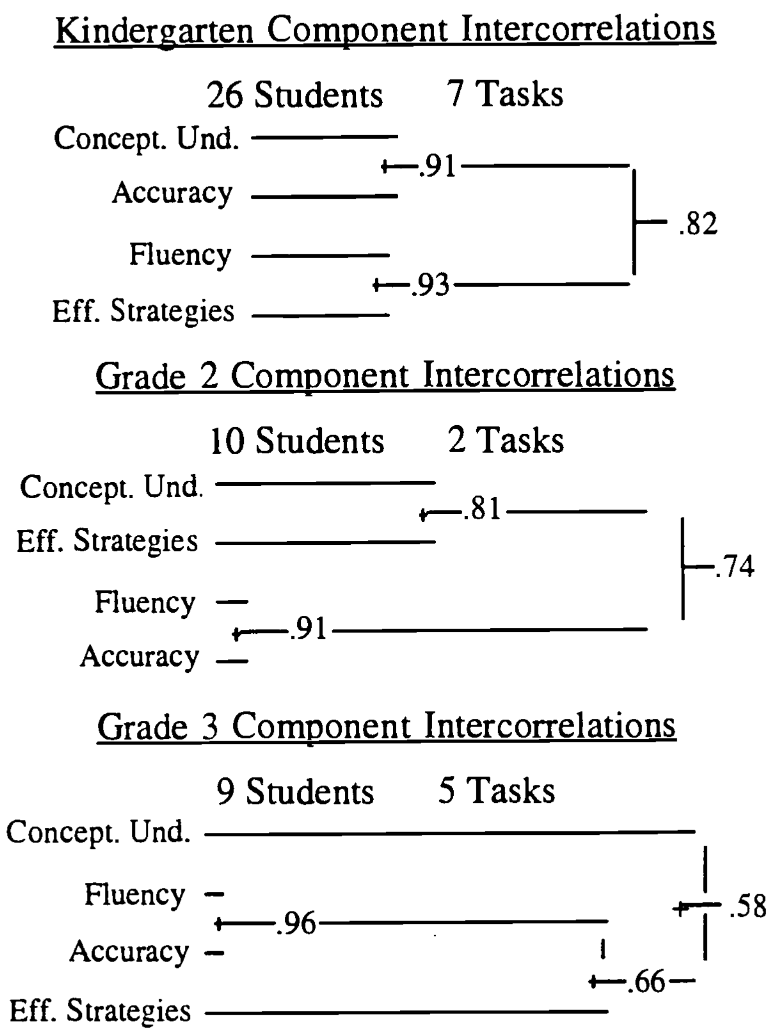
Note. Tabled numerals are means for 1-5 ratings, combinations of the four subtest scores: (a) conceptual understanding; (b) use of valid, effective strategies; (c) accuracy of task completion; (d) fluency & Speed of task completion.

*p value based on paired t-tests.

**E.S. = Effect Size, standardized mean difference.

Figure 1.

Cluster trees for the interrelationship of four task ratings across grade level tasks for Kindergarten, Grade 2, and Grade 3.



*Note: The numerals on the cluster trees are Pearson correlation coefficients.

* Concept. Und.= Conceptual understanding; Accuracy = Accuracy of task completion; Fluency = Fluency & speed of task completion; Eff. Strategies = Use of valid, effective strategies.

Figure 2. Patterns of Component Skills for First and Second Assessments at Each Grade Level.

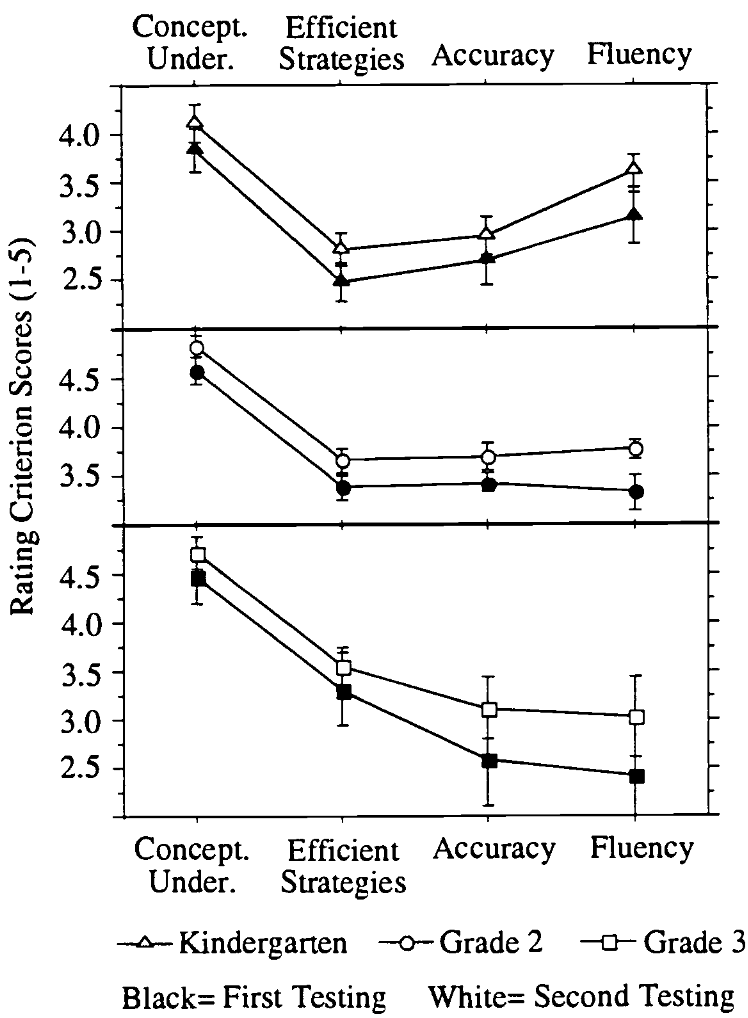
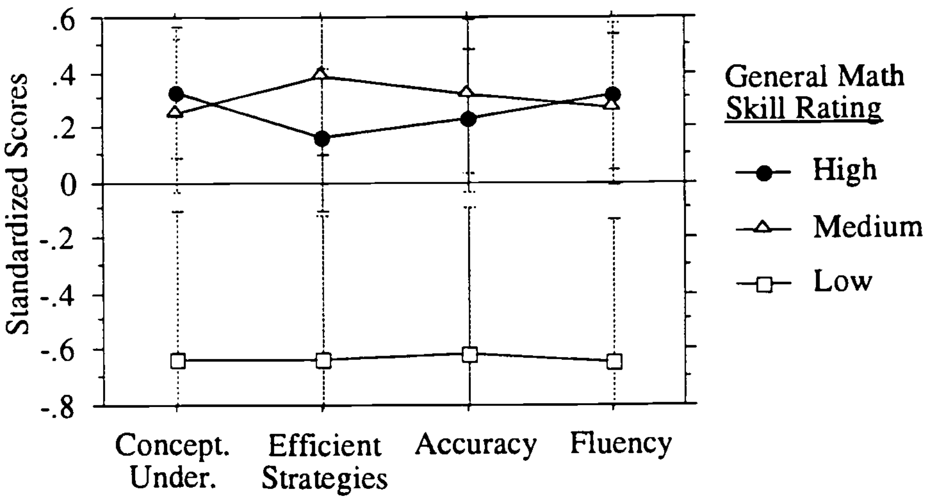


Figure 3. Relationship between Subscores on Manipulatives-based Test* and General Math Skills.



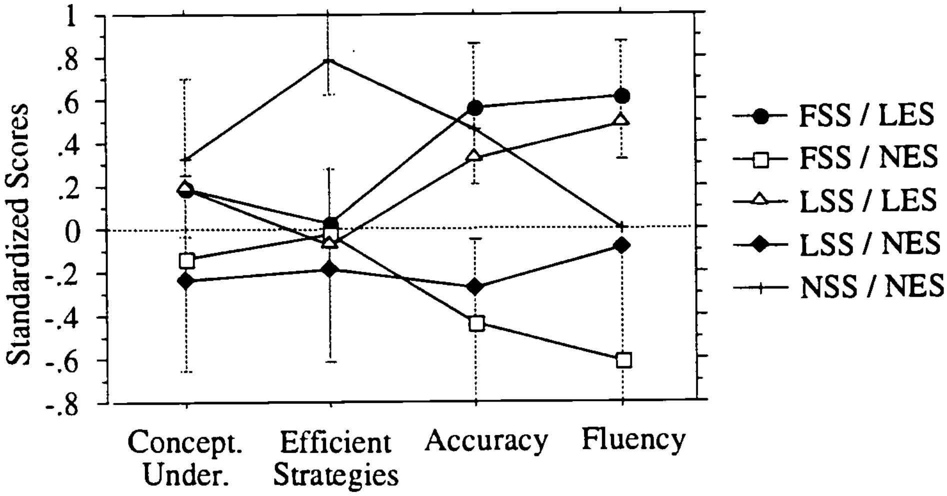
*High (n=15): 87th-70th %iles.

Medium (n=16): 58th-39th %iles.

Low (n=14): 30th-12th %iles.

Note: Error bars mark 85% Confidence Intervals

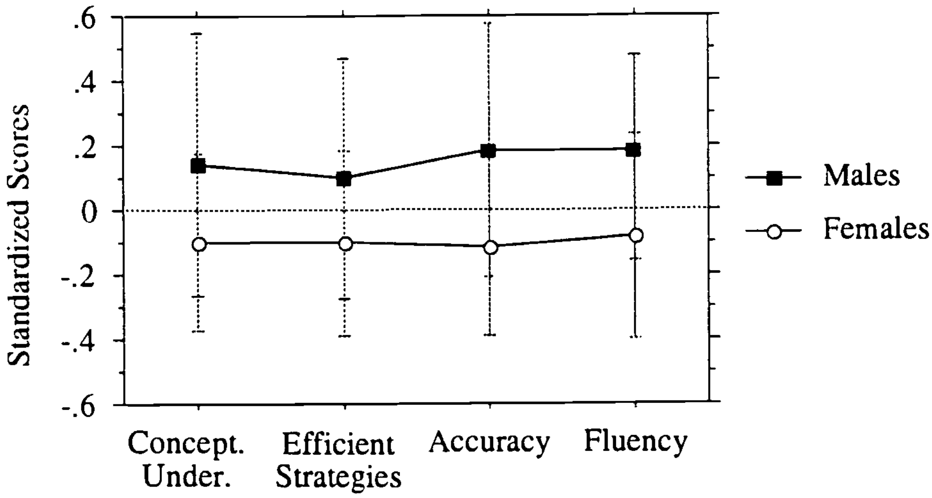
Figure 4. Prediction of component scores on manipulatives-based test based on five English and Spanish Skill groupings*



*FSS/LES (n=5): Fluent Spanish/Limited English
FSS/NES (n=13): Fluent Spanish/No English
LSS/LES (n=12): Limited Spanish/Limited English
LSS/NES (n=11): Limited Spanish/No English
NSS/NES (n=4): No Spanish/No English

Note: Error bars mark 85% Confidence Intervals for mean scores. Error Bars are provided only for highest and lowest scores, to reduce confusion.

Figure 5. Prediction of component scores based upon gender.



Note: Error bars mark 85% confidence intervals around mean scores.

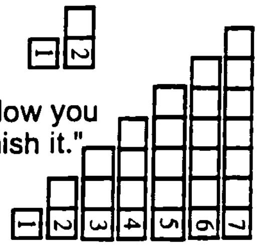
Sample Portfolio Assessment Tasks for K-3 Mathematics

1. Making Stair-steps. [K]

"Here are seven rods. I want to line them up evenly on the bottom, so they make stair-steps.

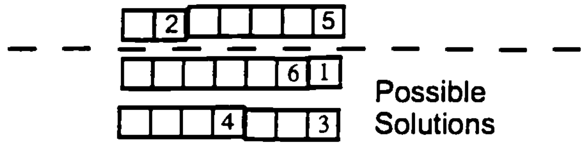
"I started it for you."

"Now you finish it."



2. Equal Trains. [K]

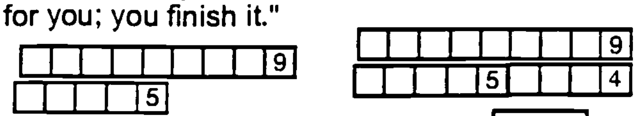
"Here I have made a train out of the two rod and the five rod. Now you make a different train the same size, but you have to use different rods."



Possible Solutions

3. Find the missing rod. [K]

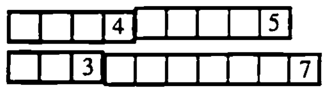
"Here is the 10 rod. I want to make a train with two cars which is just as big as the 10 rod. I started it for you; you finish it."



Answer: 4 rod

4. Which is more? [K]

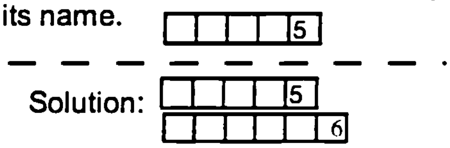
"Make a train out of the 4 rod and the 5 rod. Now make a different train out of the 3 rod and the 7 rod. Show me which train has more. "



Answer: 3 & 7

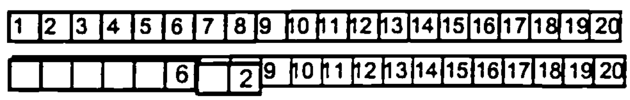
5. Finding "one bigger". [K]

"This is the 5 rod. I want to find the rod that is just one bigger, and put it beside this 5 rod. Then say its name.



6. Find the length of a train. [K]

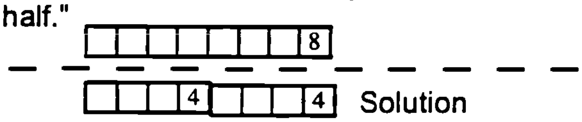
"Make a train out of the 6 rod and the 2 rod. Put the train on the numberline. Then tell me how long the train is."



Answer: 8

7. Half the size. [K]

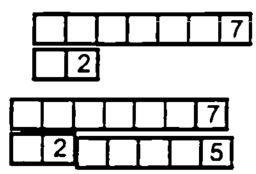
"Find the rod which is exactly half the size of this 8 rod. Show me how you know that it is half."



Solution

8. Missing addends. [Gr. 2]

"Show the missing rod that will sum to make 7. Then complete and read the equation."

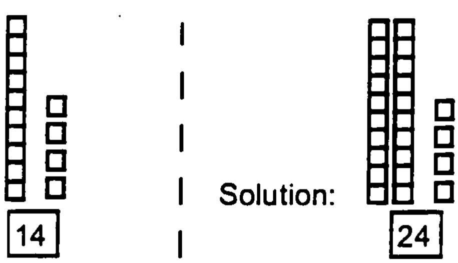


2 + __ = 7

2 + 5 = 7

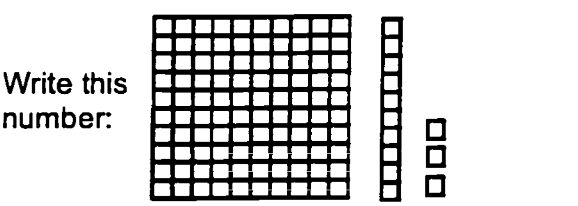
9. Ten More Than. [Gr. 2]

This number is 14. Add ten more, and tell me what new number you have made.



10. Writing Three-Digit Numbers. [Gr. 3]

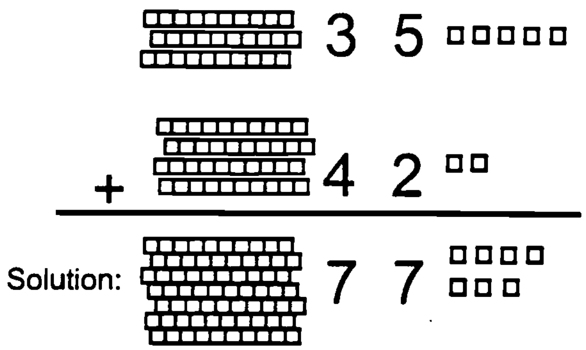
Write this number:



Solution: 113

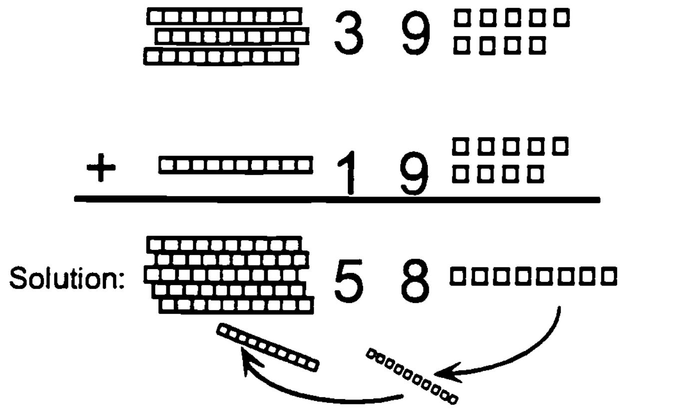
11. Simple Two-Digit Addition. [Gr. 3]

Solve the following problem by combining the rods. Then write and read your answer.



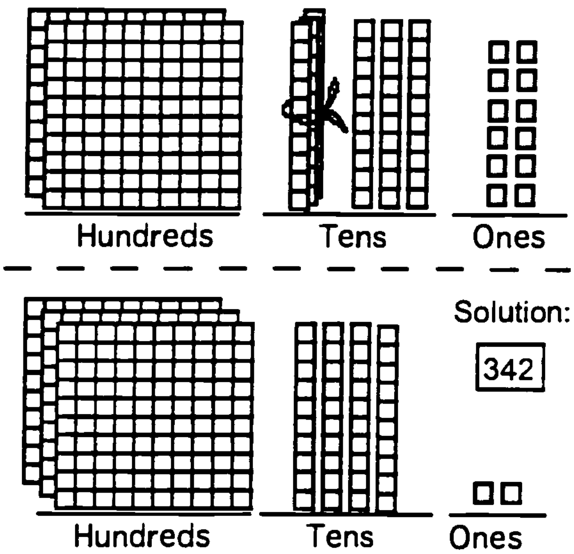
12. Two-Digit Addition with Regrouping. [Gr. 3]

Solve this addition problem by combining and trading the rods. Then write and read your answer.



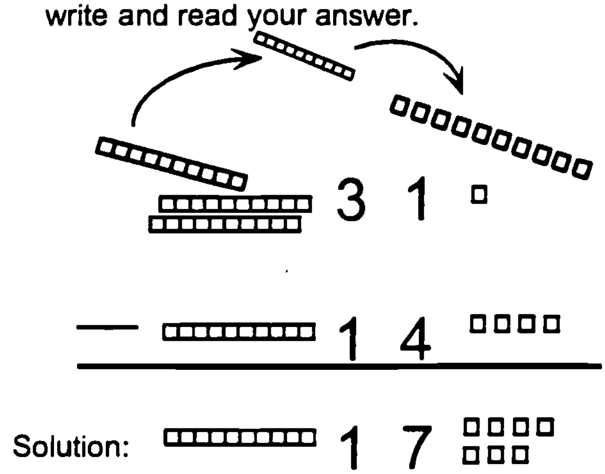
13. Trading Up. [Gr. 3]

Trade up from ones to tens and from tens to hundreds. Then write the number.



14. Two-Digit Subtraction with Regrouping. [Gr. 3]

Solve this subtraction problem by combining and trading the ones and tens rods. Then write and read your answer.



FL024277

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
Educational Resources Information Center (ERIC)

ERIC®

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Math Assessment with Manipulatives for Hispanic Limited English Proficient Students.

Author(s): Lara-Alecio, R., Parker, R., Irby, B., Mason, S., & Avila, C

Corporate Source:

Publication Date:

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic/optical media, and sold through the ERIC Document Reproduction Service (EDRS) or other ERIC vendors. Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following two options and sign at the bottom of the page.

[X] Check here **For Level 1 Release:** Permitting reproduction in microfiche (4" x 6" film) or other ERIC archival media (e.g., electronic or optical) and paper copy.

The sample sticker shown below will be affixed to all **Level 1** documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Level 1

The sample sticker shown below will be affixed to all **Level 2** documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Level 2

[ ] Check here **For Level 2 Release:** Permitting reproduction in microfiche (4" x 6" film) or other ERIC archival media (e.g., electronic or optical), but *not* in paper copy.

Documents will be processed as indicated provided reproduction quality permits. If permission to reproduce is granted, but neither box is checked, documents will be processed at Level 1.

"I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic/optical media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries."

Sign here→ please

Signature: Rafael Lara

Printed Name/Position/Title: Rafael Lara-Alecio Assist. Professor & Director Bilingual/ESL Programs.

Organization/Address: Bilingual Education Texas A&M University, College Station, TX 77843-4232

Telephone: (409)845-3467

FAX: (409)845-9663

E-Mail Address:

Date: 2/12/97

(over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

ERIC Clearinghouse on
Languages & Linguistics
1118 22nd Street NW
Washington, D.C. 20037

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
1100 West Street, 2d Floor
Laurel, Maryland 20707-3598

Telephone: 301-497-4080
Toll Free: 800-799-3742
FAX: 301-953-0263
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com