1-1-2001

Criterion validity of an estimation general outcome measure for middle school mathematics

Kristin Sue Harrington
Iowa State University

Criterion validity of an estimation general outcome measure for middle school mathematics

by

Kristin Sue Harrington

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Major: Education (Special Education)

Major Professor: Anne Foegen

Iowa State University

Ames, Iowa

2001

Graduate College
Iowa State University

This is to certify that the Master's thesis of

Kristin Sue Harrington

has met the thesis requirements of Iowa State University

Signatures have been redacted for privacy

TABLE OF CONTENTS

ABSTRACT	iv
CHPATER 1. INTRODUCTION	1
CHAPTER 2. LITERATURE REVIEW	4
CHAPTER 3. METHODS	11
CHAPTER 4. RESULTS	16
CHAPTER 5. DISCUSSION	21
APPENDIX A. ESTIMATION PROBE	27
APPENDIX B. ESTIMATION TEST	29
APPENDIX C. TEACHER RATING FORM	32
REFERENCES	34

ABSTRACT

The present study addressed the importance of developing general outcome measures in mathematics using an estimation task. The study examined a general outcome measure estimation probe using data from 111 sixth grade students from a suburban area in Iowa. The estimation probe was compared to a formal test of estimation to establish validity. A moderately strong relation was obtained between the estimation probe and the formal estimation test. Further comparisons were made between the estimation probe and three criterion measures including teacher ratings, grade point in mathematics and composite scores in mathematics from the Iowa Test of Basic Skills. The estimation test in this study was also compared to the same criterion variables. The difference between each of the correlations was determined to be statistically insignificant. These results indicated the probe in this study could be used interchangeably with the formal estimation test as an indicator of student proficiency in mathematics.

CHAPTER 1

INTRODUCTION

In recent years there has been increased emphasis on student achievement and assessment in education (Ysseldyke, Thurlow, & Shriner, 1992). The reauthorization of the Individuals with Disabilities Act (IDEA) in 1997 also made teachers in the field of special education look more closely at their practices. This focus on demonstrating student outcomes has created an increased need for efficient and consistent ways to determine student achievement in specific content areas.

Although many different types of assessments are available to teachers, they vary in their usefulness. Standardized tests, given annually, can be helpful, but are limited in their ability to show regular student progress within an academic year (Wesson, King, & Deno, 1984). Teacher-made tests vary in their consistency, reliability and validity and therefore are limited in their usefulness (Fuchs & Deno, 1991). Curriculum-based measurement (CBM) has been a popular and empirically supported tool for frequently monitoring student growth within basic skills areas at the elementary level (Shinn, 1989). Given the advantages and empirical support for CBM, this type of assessment may prove to be a useful addition to the assessment tools available to monitor the progress of secondary students.

There has not been a great deal of research on CBM at the secondary level largely because secondary settings emphasize application of known skills rather than mastery of basic skills (Espin & Tindal, 1998). Although designing a CBM system for secondary students may be difficult, it could be extremely useful for documenting student growth and achievement in a given curricular area. Based on the concept of

CBM, a more general outcome measure indicating achievement in global concepts within a curricular area may be an appropriate way to show student growth at the secondary school level. Although measures appropriate for secondary students have been developed for reading, written expression and content area learning, little work has been done in the area of mathematics (Espin & Tindal, 1998).

Therefore, the need remains for an effective way for teachers to quickly and accurately monitor secondary students' progress in mathematics. If such a program could be designed it would have to be reliable and valid, simple and efficient, easily understood and inexpensive (Deno, 1985). Foegen and Deno (in press) have explored the technical adequacy of two potential general outcome measures for secondary mathematics. One task involved fluency with basic facts and the other involved estimation. Although the reliability and criterion validity of the measures were within acceptable ranges in this initial study, additional questions remain about the technical adequacy of the measures.

This study extends Foegen and Deno's initial research by further examining the criterion validity of the estimation probes. The study addresses the following research questions:

1. What is the relation between performance on the estimation probes and performance on a formal test of estimation?
2. To what extent do the estimation probes and estimation test differ in their relationship to teacher ratings of students' math proficiency?

3. To what extent do the estimation probes and estimation test differ in their relationship to students' grades in math?

4. To what extent do the estimation probes and estimation test differ in their relationship to composite scores in mathematics from the Iowa Test of Basic Skills?

Definitions of Key Terms

To assist the reader, the following sections includes definitions of key terms used in the study.

Criterion Measures: indicators of student success in a given content area measured in ways other than the assessment being investigated.

Estimation Probe: a 3-minute, 40 question, multiple-choice test that involves choosing a correct estimate from three options that differ in magnitude.

Estimation Test: a 10 minute, 40 question, open-ended test of estimation that involves generating an estimate for a problem displayed on an overhead transparency.

General Outcome Measure: a task drawn from a skill or content area and for which a student's performance on that task is related to the student's proficiency in the skill or content area; growth evidenced on a general outcome measure is hypothesized to relate to overall growth in the skill area.

The following chapter includes a review of literature related to mathematical estimation and general outcome measures. The chapter discusses the background of general outcome measures and their importance. It also describes the relation between mathematical estimation and other aspects of mathematical proficiency.

CHAPTER 2

LITERATURE REVIEW

The purpose of this study was to investigate the criterion validity of timed, multiple-choice mathematical estimation probes (subsequently referred to as estimation probes). This study examined how student performance on the estimation probes relates to performance on a timed, open-ended test of mathematical estimation (subsequently referred to as the estimation test). In addition, the study explored relations between these two forms of estimation tests and other criterion measures, including teacher ratings, grades, and standardized test scores. The context of the study requires that one understand the importance of general outcome measurement, as well as the importance of mathematical estimation.

General Outcome Measurement

General outcome measurement originated as a way to provide teachers with a reliable, valid, and efficient means of monitoring students' progress and instructional programs (Fuchs & Deno, 1991). General outcome measurement is standardized; it has clear prescriptive procedures and long-range consistency. It measures the extent to which a student is mastering the global outcomes that are intended to result from a given curriculum. As a student progresses through a curriculum, the student's scores naturally improve, but to what extent depends on the level of overall mastery. General outcome measurement differs from subskill mastery measurement (described further below) in that the former uses long range goals to establish the domain for assessment (Fuchs & Deno, 1991).

To illustrate how general outcome measurement is used in practice, envision a seventh grade mathematics teacher. This teacher has created a series of math tasks, each of which reflects the broad goals of the seventh grade curriculum and requires students to integrate the skills they are learning. The tasks are parallel in problem types and difficulty level, but differ in the exact content of the individual problems. The tasks are short; they may be done weekly or monthly. At the beginning of the school year the teacher would expect that students would not perform well since the material is new. However, as students progressed through the school year and received instruction in new math concepts, growth would become evident by the increased number of correct answers on the probes. The students' scores would improve as overall proficiency in mathematics increased.

In contrast, subskill mastery measurement breaks a curriculum down into specific skills that lead to a larger skill. These skills are generally hierarchical and vary in complexity. The skills are usually assessed in isolation. This type of assessment becomes problematic in that retention and generalization of skills are not measured. When used in a gatekeeping manner, subskill mastery testing also inhibits students from moving through a curriculum if they are struggling with a concept. Other problems with this approach include unknown consistency in teacher made tests and difficulty developing an overall perspective of where the student is with regard to the larger scheme of the curriculum (Fuchs & Deno, 1991). General outcome measurement can address many of these concerns by assessing students with tasks that represent the entirety of a curriculum. Each parallel form of the task requires students to integrate their learning across the component elements of the curriculum.

As described earlier, one form of general outcome measurement (known as CBM) has an extensive research base supporting its use at the elementary level (Shinn, 1989). Although initial research has begun on the extension of this model to the secondary level, very few studies have addressed the area of mathematics. General outcome measures of mathematics may be very useful for secondary educators. If a general outcome measure possesses 1) reliability and validity, 2) simplicity and efficiency, 3) understandability, and 4) inexpensiveness, then it meets the criteria for on-going progress monitoring identified by Deno (1985). Measures of this quality can provide educators with valuable information.

General Outcome Measures for Middle School Mathematics

Foegen and Deno (in press) investigated two types of mathematics probes for use at the middle school level. The investigation was done with urban middle school students from a large metropolitan area. Approximately 100 students of diverse racial and ethnic backgrounds attended the school. Review of records from the school showed approximately 75 percent received free or reduced-priced lunches and 9.9 percent received special education services. The mathematics probes explored in the study were a Basic Mathematics Operations Task (BMOT) and a Modified Estimation Task (MET).

The BMOT was a basic facts fluency task similar to the types of CBM probes used with elementary students. The MET required students to select the best estimate to a computation or word problem from three choices. The response options differed by an order of magnitude (e.g., 2, 20, 200). Students were to quickly determine which option was approximately correct and circle that response.

Foegen and Deno's initial findings with these probes revealed acceptable levels of reliability and criterion validity. Internal consistency coefficients for the BMOT ranged from .91-.92. Test–retest reliability coefficients ranged from .80-.85 and parallel forms reliability coefficients ranged from .70-.82. The internal consistency coefficients for the MET ranged from .77-.93. The test-retest and parallel forms reliability coefficients ranged from .67-.88 and .67-.86, respectively.

To examine criterion validity, the measures were correlated with group standardized achievement tests, teacher ratings and student grades in mathematics. Correlations between the BMOT and math grade point average, standardized test scores in mathematics and teacher ratings were .44, .44-.63, and .16-.54, respectively. Correlations between the MET and math grade point average, standardized test scores in mathematics and teacher ratings were .39, .45-.56, and .15-.54, respectively. While these initial results are promising, additional work is needed to establish the technical adequacy of middle school mathematics general outcome measures.

Estimation

The use of mathematical estimation as a general outcome task is preferable to the use of fluency with basic facts because the process of estimating involves various thought processes and can be applied to a wide range of mathematical domains. The literature on estimation documents strong correlations between estimation and number sense (McIntosh, Reys, & Reys, 1992), estimation and computation skills (Rubenstein, 1985), and estimation and real life applications of mathematics (B. Reys, 1992). In addition, the National Council of Teachers of Mathematics (NCTM, 1989) has urged

schools to use a more general curriculum that applies mathematical knowledge rather than focusing on computation skills.

Estimating is more than guessing or rounding off numbers to an easy figure. Estimation can be used to study the development of mathematical competence. Upon deeper investigation estimation can show the extent of a student's procedural and conceptual knowledge (Bisanz & LeFevre, 1990).

Estimation involves the manipulation of numbers and procedures to approximate an answer and then determine its appropriateness (LeFevre, Greenham, & Waheed, 1993). Students who are good estimators are able to "reformulate" problems by changing numbers through rounding. They are able to "compensate" by adjusting either of the numbers to estimate. For example, a student may round one number up and one number down to get a closer approximation. Good estimators also "translate" by changing processes. Rather than adding three similar numbers they may simply multiply by three. Estimation skills seem to develop simultaneously with other mathematical skills (LeFevre et al., 1993)

Estimation and Number Sense

Estimation is interrelated with number sense. Number sense is a broad encompassing term that indicates one's ability to understand numbers and operations along with the ability to apply this knowledge and skill (McIntosh et al., 1992). Individuals with number sense are able to manipulate numbers and operations and make judgments about the results of their efforts (McIntosh et al., 1992). Estimation is one part of this ability to make sense of situations involving mathematical

manipulations. The ability to estimate allows a person to know whether or not an answer is approximately correct.

Problem solving is so intricately entwined in mathematics the two are inseparable (R. Reys, 1985). The ability to problem solve and estimate go hand in hand with number sense. This assertation is supported by LeFevre et al. (1993) who found that adults scored considerably higher than middle school students on tests of estimation.

Estimation and Computation

The relationship between computation and estimation is complex. Estimation correlates with computation. In a study conducted by Rubenstein (1985), students who had difficulty operating with tens and making comparisons had difficulty with estimation. However, the relationship is not that simple. In a study of Taiwanese students, R. Reys and Yang (1998) found that students scored significantly better on written computation than on questions relying on number sense. This would indicate that being able to follow procedural rules for computation does not necessarily mean that a student can manipulate operations in order to form an accurate estimate.

Rubenstein (1985) noted the difficulty of the mathematical operation was reflected in the ability to estimate. For example, estimation problems involving addition and subtraction were easier than estimation problems involving multiplication and division. As indicated by LeFevre et al. (1993) it would appear that overall mathematical competence, including proficiency with computational skills, surfaces in estimation. Students may know procedural ways to solve problems, but when forced to look at problems and logically manipulate the operation in order to estimate they

become uncertain. Therefore, students who are skilled at computation do not necessarily excel at estimation. The process of estimation requires additional manipulation skills.

Estimation in Daily Living

Do I have enough money for that item? How much paint will I need? Will fifty dollars be enough for groceries this week? Estimation is a skill that is used by adults on a regular basis (B. Reys, 1992). For practical reasons, students need to learn to be efficient at estimation. The new push for practical application of mathematics by the NCTM (1989) helps educators understand the importance of common math skills.

Given the skills required of proficient estimators and empirical relations between estimation and number sense, computation, overall mathematical competence and daily living applications, the choice of estimation as a general outcome measure is justified.

Summary

This review of literature has established a need for middle school mathematics general outcome measures, but little work has been done to date. The mathematics education literature supports estimation as a logical choice for assessing general mathematical proficiency. Moreover, preliminary research on estimation probes has yielded positive results regarding the technical adequacy of these measures. The present study extends the current knowledge base on general outcome measures for middle school mathematics by further exploring the technical adequacy of the estimation probe developed by Foegen and Deno (in press). In the next chapter, a description of the methods used for the study is provided.

CHAPTER 3

METHOD

The purpose of this study was to gather information with regard to these research questions:

1. What is the relation between performance on the estimation probes and performance on a formal test of estimation?
2. To what extent do the estimation probes and estimation test differ in their relationship to teacher ratings of students' math proficiency?
3. To what extent do the estimation probes and estimation test differ in their relationship to students' grades in math?
4. To what extent do the estimation probes and estimation test differ in their relationship to composite scores in mathematics from the Iowa Test of Basic Skills?

An existing set of data was used in order to investigate these questions. The data were gathered as part of a study conducted by Dr. Anne Foegen during the spring of 1997. Human subjects approval was requested and obtained for the original study. Students were only included in the data set if a parent/guardian and the student himself/herself responded affirmatively on their respective consent forms.

Participants and Setting

The participants in this study were 111 sixth grade students (67 male, 44 female) in a suburban middle school in the Midwest. Approximately 825 students attended the middle school for sixth and seventh graders. The students were

predominantly white (97%) and came from middle to upper-middle class families. Within this building 8% of the students received free or reduced-priced lunches.

Following the middle school philosophy, each grade was organized into three multidisciplinary teams. Each team included teachers representing language arts, reading, mathematics, science, social studies and special education. Support for students requiring special education services in mathematics was provided through a co-teaching model in which a special educator was present daily in the general education mathematics classroom. For students unable to participate in general education mathematics, a separate math class (taught by a special education teacher) was provided.

Measures and Materials

Two types of measures were used in the study: the estimation measures and additional criterion measures. The estimation measures were the estimation probes and the estimation test. The additional criterion measures included teacher ratings, grades in math and standardized test scores. Each of these variables is described further in the section below.

Estimation Measures

The estimation probes consisted of forty multiple-choice questions (20 computational and 20 word problems). The problems involved addition, subtraction, multiplication and division. The problems involved two- and three- digit numbers and included whole numbers, decimals and fractions. Figure 1 is an example of a computational problem and word problem. A copy of an estimation probe is provided in Appendix A.

59.4 – 30.8 is about			82 students tried out for the football team. 59 players were selected. About how many did not make the team?		
.3	3	30	2	20	200

Figure 1. Sample computational problem and word problem from probe

The students were given three minutes to complete each forty-question estimation probe. The scores were based on the number correct. The probes were graded promptly so student progress could be charted at that time. Ten parallel forms of the estimation probes were developed for the original study.

The estimation test was the CET (Computational Estimation Test) developed by R. Reys, Reys, Trafton and Zawojewski (1984). This test consisted of twenty computation and 20 application questions, all of which were open-ended. Sample questions are provided in Appendix B. To administer this test, the researcher showed forty separate overheads (one per item) for fifteen seconds each. The students recorded their answers on narrow slips of paper. The narrow slips of paper were intended to reduce students' ability to actually compute correct answers and then round them. The test included an equal number of addition, subtraction, multiplication and division problems. Answers to the problems ranged from two-digit numbers to six-digit numbers. Some more difficult problems involved fractions and decimals. The test was designed to reflect the instructional content for the grade level being tested. Two example problems are given in figure 2. According to the test developers the test-retest

37,689 – 18,812	The airplane traveled **6,153** miles on **19** gallons of fuel. ABOUT how many miles per gallon?

Figure 2. Samples from estimation test

reliability estimates for the CET ranged from .78-.88. No information on the validity of the test was reported.

Additional Criterion Measures

The students' cumulative records were reviewed for demographic information as well as standardized test scores. Math grades and composite scores for mathematics from the Iowa Tests of Basic Skills were obtained.

The mathematics teacher rated each student on a scale from one to five. One represented the lowest level of general mathematics proficiency and five represented the highest. A sample teacher rating form is provided in Appendix C.

Procedures

In the original study one form of the estimation probe was administered each week for ten consecutive weeks. Students were told they could skip questions and come back to them if they had time. An unattempted problem did not count against them. Students had three minutes to complete as many problems as possible. The CET was administered after the final week of the study. Teacher ratings of students' overall mathematics proficiency were gathered using a Likert scale instrument developed by the researcher; ratings completed during the final week of estimation probes were used in this study.

Although 10 weeks of data were collected for the original study, only scores from weeks nine and ten were used in the present study. The probes were selected for comparison against the CET because all three measures were administered within a two week period. The teacher ratings were gathered at that time as well.

Scoring and Data Analysis

Foegen (in press) determined that the multiple-choice estimation probes were susceptible to random error associated with guessing. To control this error, Foegen recommended a correction for guessing. This formula [number correct – (number incorrect/2)], suggested by Mehrens and Lehman (1991), was applied to the number correct and number incorrect scores to compute corrected estimation scores. Only corrected estimation scores were used for the analyses. The corrected estimation scores for weeks nine and ten were averaged to obtain a more stable estimate of student performance. The CET was scored by comparing students' responses to ranges of acceptable answers provided by the test developers. The CET score was the total number of problems correct.

The first criterion variable was teacher ratings of overall mathematics proficiency. This consisted of a number from one to five, with one representing the lowest level of performance and five representing the highest. The second criterion variable was students' grades in math for the semester immediately preceding the study. These were recorded using a 4.0 = A, 3.0 = B, 2.0 = C and 1.0= D scale. Each student's mathematics composite score on the Iowa Tests of Basic Skills was reported using national percentile ranks.

The data were analyzed using Pearson's Product-moment correlation coefficients; the significance of differences between the correlation coefficients was tested using procedures described by Howell (1992).

CHAPTER 4

RESULTS

The purpose of this study was to determine relations between two types of estimation assessment and three other criterion variables believed to be indicators of student success in mathematics. Four research questions were addressed in the study:

1. What is the relation between performance on the estimation probes and performance on a formal test of estimation?
2. To what extent do the estimation probes and estimation test differ in their relationship to teacher ratings of students' math proficiency?
3. To what extent do the estimation probes and estimation test differ in their relationship to students' grades in math?
4. To what extent do the estimation probes and estimation test differ in their relationship to composite scores in mathematics from the Iowa Test of Basic Skills?

The following chapter presents the results of the study.

Prior to conducting any analyses, the accuracy of the scoring for the estimation probes and tests was examined. Thirty estimation probes (approximately 15% of the total number of probes) were randomly selected and rescored by another graduate student. A problem by problem comparison of each scorer's results was used to tally the number of problems on which the scorers agreed and disagreed. The interrater reliability was calculated by dividing the number of agreements by the number of

agreements plus the number of disagreements. The two scorers agreed 98.9 percent of the time on the scoring of the estimation probes.

The same procedure was then applied to the estimation test. Thirty tests (approximately 30% of the total) were randomly selected and rescored. The level of agreement for the two scorers was 99.8% agreement for the estimation test.

Descriptive Data for Study Variables

Means and standard deviations for each of the variables in the study are listed in Table 1. Although both types of estimation assessments contained 40 problems, the students' mean scores were higher on the multiple-choice estimation probe than they were on the open-ended estimation test. It should be noted that this group of students scored above average on each of the criterion variables. The average teacher rating was 3.5 on a scale of one to five. The average grade in math was nearly 3 on a scale of one to four. They were also above the national average percentile rank of 50 on the ITBS. The students' average percentile rank was nearly 70.

Correlation Between the Estimation Measures

To answer research question one, the Pearson r correlation coefficient between the estimation test and the estimation probe was computed. The obtained coefficient ($r = .65$, $p < .01$) was based on the 95 participants for whom I had complete data. This correlation is defined as moderately strong according to Hinkle, Wiersma, and Jurs. (1998).

Table 1. Means and Standard Deviations

Measure	N	Mean	Standard Deviation
Estimation probe	99	17.23	7.57
Estimation test	106	11.57	6.80
Teacher rating	111	3.52	1.14
Math grade	109	2.96	.78
ITBS total math	100	69.79	23.93

Correlations Between Estimation Measures and Criterion Variables

Table 2 lists the correlations between the estimation measures and the criterion variables. All of the correlations could be classified as moderately positive using the standards outlined by Hinkle, et al. (1998). As the table indicates, all of the correlation coefficients were statistically significant. The correlations between the estimation test and the criterion variables ranged from .52 to .60. The correlations between the estimation probe and the criterion measures were slightly higher, ranging from .54 to .65. For two of the criterion variables (math grade and ITBS total math), the correlations with the estimation probe were slightly higher than the correlations with the estimation test. For the teacher ratings, the estimation test produced a slightly higher correlation. In general, however, the relationships appeared to be very similar.

The correlations obtained in this study align with the findings of Foegen and Deno (in press), who obtained coefficients between an estimation probe and several criterion measures that ranged from .16 - .54. On average, the assessment measures in

Table 2. Sample sizes and correlations

Measure	Estimation Test		Estimation Probe	
	N	r	N	r
Teacher rating	106	.60**	99	.58**
Grade point average	92	.52**	96	.54**
ITBS total math	96	.57**	89	.65**

**p-value <.01.

the present study have a slightly stronger relationship to the criterion variables than did the measures in the Foegen and Deno study.

Differences in Relations between the Estimation Measures and the Criterion Measures

The intent of research questions two through four was to determine whether the correlations between the estimation probe and each of the criterion variables differed significantly from the correlations between the estimation test and each of the corresponding criterion variables. The procedures described by Howell (1992) for testing two non-independent correlation coefficients were used for the analyses. In order to provide the most accurate relations, only the scores from students with complete data for all three variables were used in the calculations. Table 3 has been included to show the differences in the correlation coefficients when the analyses were limited to students with complete data.

For the teacher rating scale, the correlations with the estimation probe and the estimation test were not significantly different $N = 95$, $t = .49$, $p > .50$. For a sample

Table 3. Coefficients for 3-way complete data

Measure	N	Estimation test	Estimation probe
Teacher rating	95	.60**	.56**
Math grade	93	.52**	.55**
ITBS score	86	.56**	.66**

**p-value <.01

size of 95, the t value must exceed 1.98 for the correlation to be statistically significant at the .05 alpha level.

For math grade, the correlations with the estimation test and estimation probe were not significantly different (N=93, $t = .32$, $p > .50$). The correlations of the ITBS scores with the estimation probe and the estimation test were not significantly different either (N=86, $t = .93$, $p > .40$). None of the three criterion variables showed a significant level of difference in their relations to the two types of estimation measures.

In summary, the results of this study identified a moderately strong relation between the estimation probe and an open-ended test of estimation. The two measures were similar in the strength of their relations to three criterion measures, including teacher ratings, grades in mathematics, and composite mathematics scores on a standardized test. The findings were consistent with the results obtained by Foegen and Deno (in press) who showed moderate relations between math probes and criterion variables. The following chapter will discuss the significance of these findings and their application for classroom use.

CHAPTER 5

DISCUSSION

The purpose of this study was to examine the criterion validity of a general outcome measure for middle school mathematics. More specifically, the study explored the relations between an estimation probe (general outcome measure) and an estimation test and then further explored relations between the two estimation measures and criterion variables thought to be indicators of student success in mathematics. Positive correlations would indicate that the estimation probe may be a valid indicator of the mastery of global mathematics concepts. With sound research results, this type of assessment could be used in the future in much the same way educators use tests such as the Iowa Tests of Basic Skills to report student achievement. Positive correlations could also support the repeated use of the measure by educators and assessment specialists to determine a student's growth in mathematics proficiency.

The results from this study provide evidence that the estimation probes are sound assessment devices. In each comparison between the estimation probes and another measure, a moderately strong correlation was found. The following sections discuss what those correlations mean to researchers and educators alike.

Correlation Between the Estimation Measures

The first issue explored in the study was the relation between students' performance on the estimation probe and an open-ended test of estimation used by mathematics education researchers. One way to establish the criterion validity of the estimation probe is show that it provides results similar to other established tests of estimation. The findings of this study indicate that there is a moderately strong positive

relation between the estimation test and the estimation probe. This means that both measures would provide a similar indication of the level of proficiency in estimation. The measures could be used interchangeably to show student achievement. This is useful to educators because it provides teachers with an alternative way to assess student achievement in mathematics, specifically in the area of estimation. The efficiency of estimation probes make them very viable as a new form of assessment at the middle school level.

Relations Between Estimation Measures and Criterion Variables

The other important focus of this study was an exploration of the differences between correlations as they pertain to each external criterion variable. The study found that the estimation probe and the estimation test correlated similarly to each of the criterion variables. Some of the correlations between the probe and the criterion variables exceeded those obtained for the estimation test and the same criterion variable. For example, the correlation between the probe and the ITBS total scores in math was higher than the correlation between the test and the ITBS scores. Although differences existed, none of the differences were deemed statistically significant. This would indicate that both the estimation probe and the estimation test deliver similar results for measuring student achievement in mathematics. Either of these tests could be used to provide a teacher with information about student mastery of mathematics concepts. These probes could be used in much the same way as teacher ratings, grades, and standardized tests.

It is also important to note that when the estimation probe was compared to indicators of mathematics proficiency much broader than the estimation test it

continued to show moderately strong relations. This is an important consideration because it supports the use of estimation as an indicator of more generalized proficiency in a range of mathematical domains. The probe's usefulness is not limited to estimation.

Implications for Practice

This type of research is important because educators are accountable to parents and the general public. There are frequent editorials and news articles demanding accountability from teachers. People want to know what students in their school district are learning. More importantly, parents want to know that their student is mastering the material within the curriculum. This type of assessment may be an excellent way to show people how well students are doing.

If general education outcome measures such as the estimation probes used in this study are to be useful in practice, they must be valid, reliable, simple, efficient, understandable, and inexpensive (Deno, 1985). The results of this study add to the research by Foegen and Deno (in press) to demonstrate that such measures are valid and reliable. Another major advantage of the estimation probes is efficiency. The testing time for a probe is three minutes and the teacher time to correct such a probe is minimal, especially compared with other mathematics assessments. By comparison, the estimation test, although efficient to administer, takes more time to correct because one must decipher the variety of students' responses and compare them to an acceptable range of answers. Grading open-ended mathematics assessments can be quite time consuming. Probes, as a form of general outcome measure, are also easy to understand. There are not elaborate administration procedures or voluminous sets of instructions.

The probes are very easy to use. It is also inexpensive for school districts to use probes as a form of assessment since the probes can be copied year after year to use with the same grade level students.

Probes such as these can provide mathematics and special educators with valuable information about the performance of their students. These types of probes are especially valuable for getting frequent feedback about student understanding of mathematics concepts. It is easy and feasible to construct and use multiple forms of an estimation probe. In doing so, teachers can monitor progress on a regular basis and are not limited to the feedback from test forms such as ITBS or the CET. These tests can only provide educators with information about student achievement at one point in time and are generally administered only once per academic year.

Limitations

This is only one study, with a sample drawn from one group of sixth graders from a middle school in an upper-middle class neighborhood. The students as whole scored above the national average on the ITBS tests which may indicate they do well at testing. It is unclear whether the results obtained in this study would generalize to other groups of students more diverse in race/ethnicity, socioeconomic levels and achievement status.

This study was done in February and March and the ITBS tests were administered the preceding October. This gap in time may have impacted the relations between the ITBS scores and the other measures.

Although this study had positive results at the middle school level it is unclear if such a measure could be transposed to a high school setting. Due to the variety of

curriculum that students encounter in high school mathematics it may be difficult to design a general outcome measure.

Finally, this study was done once with one group of students. To be used for actual progress monitoring we would need to know that student scores change over time. This study did not look at the probes in this manner.

Directions for Future Research

This study adds to a small, but growing knowledge base on the validity and reliability of general outcome measures at the secondary level. Future research is needed to explore whether the estimation probes are sensitive to changes in student performance over time. A function of general outcome measurement is monitoring student progress to formatively evaluate teaching effectiveness. If proficiency with estimation is a static characteristic that does not change as students become more proficient in mathematics, the usefulness of the probes will be extremely limited.

Further study of the application of estimation measure to other grade levels is also needed. This research should not only include middle school students in grades seven and eight, but high school students as well.

Conclusion

It would appear that although no measure of mathematical assessment is perfect, each one has its merits. The estimation probe examined in this study does correlate with a formal test of estimation. Both forms of assessment correlate similarly with other indicators of student success in mathematics. The correlations with various indicators of student achievement included in this study indicate that different measures of student achievement produce comparable results. Each measure has its own

purpose, but collectively they can provide educators with a more complete picture of student ability.

Using this type of general outcome measure could be a quick and accurate way to monitor student success within a given curriculum. This is an important implication for teachers who need frequent, yet accurate feedback on student mastery. The simplicity of grading probes makes them efficient to use, which is important to classroom teachers. Using probes is not costly, an important consideration for school districts. Based on the correlations obtained in this study, probes may portray student success in mathematics as accurately as other commonly used indicators. They also posses the advantage of being quick and inexpensive.

In summary, the study presented here indicated that estimation probes correlated well with one form of a formal test of mathematical estimation. The findings also indicated that the differences between each of the estimation assessments and another criterion measures were insignificant. As we look at the importance of assessing student achievement it would appear that probes such as these could be prove to be a very worthwhile alternative.

APPENDIX A. SAMPLE ESTIMATION PROBE

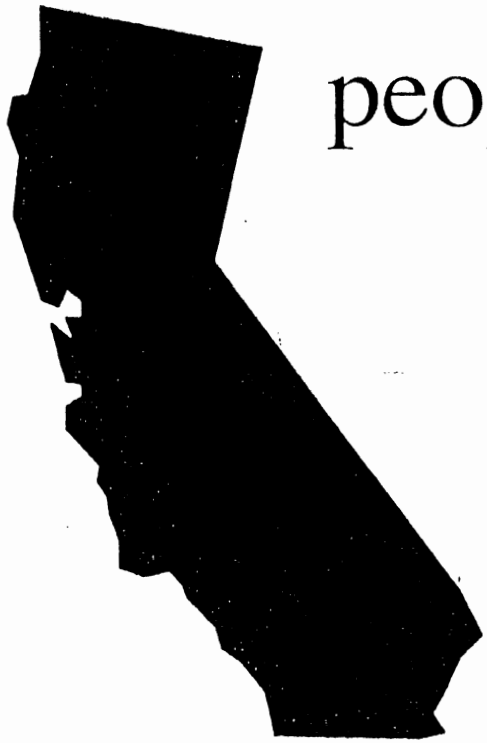
Estimation Probe 4

Evan's baseball card collection has 326 cards from 11 teams. About how many cards does he have of each team?	3	30	300	767 + 31.4 is about	800	8,000	80,000	Matt and Trina rode 2 1/2 miles and then walked their bikes 1 3/4 miles. About how many miles did they travel?	0.4	4	40	0.31 - 0.19 is about	0.001	0.01	0.1	Trisha makes $4.15 an hour at MacDonalds. If she is given a 10% raise, about how much more will she earn per hour?	$0.04	$0.40	$4
96 - 27 is about	70	700	7,000	Liz bought a new car for $12,900. The salesman gave her $4,800 for her old car. About how much did she pay for the new car?	$80	$800	$8,000	29 + 52 is about	8	80	800	The recipe makes 24 cookies per batch. Sarah needs to make 100 cookies. About how many batches should she make?	4	40	400	46.3 - 25.5 is about	0.2	2	20
The pep club raised $92.50 from the car wash and $39.75 from the raffle. About how much did they raise altogether?	$13	$130	$1,300	$88\overline{)29.6}$ is about	0.3	3	30	Dan worked 18 hours and made $100 last week working at the grocery store. About how much did he get paid per hour?	$0.05	$0.50	$5	223 x 71 is about	140	1,400	14,000	This past summer 57 students took drivers training. 49 students got their licenses this fall. About how many did not get their licenses?	1	10	100
$4\overline{)431}$ is about	1	10	100	Heather needs 8 new wheels for her inline skates. She paid $3.75 for each wheel. About how much did the repair cost her?	$0.32	$3.20	$32	$21\overline{)836}$ is about	4	40	400	Skippy ate 1/4 of the 18 inch sub sandwich at the Super Bowl party. About how many inches long was his part of the sandwich?	5	50	500	51 x 32 is about	15	150	1,500

APPENDIX B. SAMPLE ESTIMATION TEST ITEMS

California has **43** representatives in Congress.

ABOUT how many people does each representative represent?



Population of California

23,669,435

$$8{,}127 \overline{)474{,}257}$$

APPENDIX C. TEACHER RATING FORM

Teacher ____________________

Developing Indicators of Student Achievement in Mathematics
Teacher Rating

<u>Directions</u>: Below is a list of the students you teach. Please rate the mathematics achievement of each student, considering that student in comparison to <u>all</u> the students you teach, not just those in the same class period. Students who have the lowest mathematics achievement in your classes should be rated "1," while those who have the highest mathematics achievement should be rated "5." Please try to use all the numbers in the scale. There should be approximately the same number of students who are rated "1" and "2" as there are students who are rated "4" and "5."

1 = very <u>low</u> mathematics achievement   5 = very <u>high</u> mathematics achievement

<u>Student</u>	<u>Mathematics Achievement</u> Very Low				Very High
____________________	1	2	3	4	5
____________________	1	2	3	4	5
____________________	1	2	3	4	5
____________________	1	2	3	4	5
____________________	1	2	3	4	5
____________________	1	2	3	4	5
____________________	1	2	3	4	5
____________________	1	2	3	4	5
____________________	1	2	3	4	5
____________________	1	2	3	4	5
____________________	1	2	3	4	5
____________________	1	2	3	4	5
____________________	1	2	3	4	5
____________________	1	2	3	4	5
____________________	1	2	3	4	5
____________________	1	2	3	4	5
____________________	1	2	3	4	5
____________________	1	2	3	4	5
____________________	1	2	3	4	5
____________________	1	2	3	4	5
____________________	1	2	3	4	5
____________________	1	2	3	4	5
____________________	1	2	3	4	5
____________________	1	2	3	4	5
____________________	1	2	3	4	5
____________________	1	2	3	4	5
____________________	1	2	3	4	5
____________________	1	2	3	4	5
____________________	1	2	3	4	5
____________________	1	2	3	4	5
____________________	1	2	3	4	5
____________________	1	2	3	4	5
____________________	1	2	3	4	5
____________________	1	2	3	4	5
____________________	1	2	3	4	5
____________________	1	2	3	4	5

REFERENCES

Bisanz, J., & LeFevre, J. (1990). Strategic and nonstrategic processing in the development of mathematical cognition. In D. Bjorklund (Ed.), Children's strategies: Contemporary views of cognitive development (pp. 213-243). Hillsdale, N.J.: Lawrence Erlbaum Associates.

Deno, S. (1985). Curriculum-based measurement: The emerging alternative. Exceptional Children, 52, 219-232.

Espin, C. & Tindal, G. (1998). Curriculum-based measurement for secondary students. In M. Shinn (Ed.), Advanced applications of curriculum-based measurement (pp. 214-253). New York: Guilford.

Foegen, A. (in press). Technical adequacy of general outcome measures for middle school mathematics. Diagnostique.

Foegen, A. & Deno, S. (in press). Identifying growth indicators for low-achieving students in middle school mathematics. Journal of Special Education.

Fuchs, L., & Deno, S. (1991). Paradigmatic distinctions between instructionally relevant measurement models. Exceptional Children, 57, 488-500.

Hinkle, D., Wiersma, W., Jurs, S. (Eds.). (1998). Applied statistics for the behavioral sciences (4th ed.). Boston: Houghton Mifflin Company.

Howell, D. (1992). Statistic methods for psychology (3rd ed.). Boston: PWS-Kent Publishing.

LeFevre, J., Greenham, S., & Waheed, N. (1993). The development of procedural and conceptual knowledge in computational estimation. Cognition and Instruction, 11(2), 95-132.

McIntosh, A., Reys, B., & Reys, R. (1992). A proposed framework for examining basic number sense. For The Learning of Mathematics, 12(3), 2-8, 44.

Mehrens, W., & Lehman, I. (1991). Measurement and evaluation in education and psychology. Fort Worth, TX: Holt, Rinehart and Winston.

National Council of the Teacher of Mathematics. (1989). Curriculum and evaluation standards for school mathematics. Reston, VA: Author.

Reys, B. (1992). Estimation. In T. Post (Ed.), Teaching mathematics in grades K-8 research based methods (pp. 279-301). Boston: Allyn and Bacon.

Reys, R. (1985). Estimation. Arithmetic Teacher, 32(6), 37-41.

Reys, R., Reys, B., Trafton, P., & Zawojewski, J. (1984). Developing Computational Estimation Materilas for the Middle Grades. Final Report. (National Science Foundation Grant No. NSF-81-13601). Columbia, MO: University of Missouri. (ERIC Document Reproduction Service No. ED 242 525)

Reys, R.E., & Yang, D.C. (1998). Relationship between computational performance and number sense among sixth and eighth-graders in Taiwan. Journal for Research in Mathematics Education, 29(2), 225-237.

Rubenstein, R. N. (1985). Computational estimation and related mathematical skills. Journal for Research in Mathematics Education, 16(2), 106-119.

Shinn, M.R. (1989). Curriculum-based measurement: Assessing special children. New York: Guilford.

Wesson, C., King, R. & Deno, S. (1984). Direct and frequent measurement of student performance: If it's good for us, why don't we do it? Learning Disability Quarterly 7, 45-48.

Ysseldyke, J., Thurlow, M., & Shriner, J. (1992). Outcomes are for special educators too. Teaching Exceptional Children, 25(1), 36-50.