DOCUMENT RESUME

ED 286 929 TM 870 570

AUTHOR	Schoen, Harold L.; And Others
TITLE	Measuring Computational Estimation Processes.
PUB DATE	Apr 87
NOTE	45p.; Paper presented at the Annual Meeting of the American Educational Research Association (Washington, DC, April 20-24, 1987).
PUB TYPE	Speeches/Conference Papers (150) -- Reports - Research/Technical (143) -- Tests/Evaluation Instruments (160)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Age Differences; Cognitive Processes; Cognitive Structures; *Cognitive Style; Concept Teaching; Constructed Response; Elementary Education; *Estimation (Mathematics); Interviews; *Mathematical Concepts; *Mathematics Instruction; Multiple Choice Tests; Response Style (Tests); *Test Format; *Test Items

ABSTRACT

The estimation processes used by fifth through eighth grade students as they responded to computational estimation test items were examined. Interview-based process descriptions were cross-validated using large group test data from an open-ended test and a multiple choice test. Five question formats were used to test different estimation processes: standard multiple choice; operation in foils; benchmark; and order of magnitude or operation in stems (for fractions). The mental processes tested were: (1) rounding by the usual rules to the closest power of ten or to the closest whole number; (2) front-ending, or rounding down to the power of ten of the leading digit or to the whole number of a mixed numeral; (3) other rounding, including all numbers up or some numbers up and others down; (4) using compatible numbers or numbers relatively close to the given numbers; and (5) compensating, or adjusting an estimate to reflect variations that might result from rounding or the use of some other adjustment process. Students had a strong mental set to round numbers to the nearest leading power of ten even when the items required other estimation processes. Performance differed by item format, types of numbers and operations in the items, and grade level of students. (The appendices include some test items). (Author/JGL)

Measuring Computational Estimation Processes

A Research Paper Presented at the

1987 Annual Meeting of the AERA

Washington, DC

by

Harold L. Schoen, University of Iowa

Glendon Blume, Pennsylvania State University

Eric Hart, Maharishi International University

ABSTRACT

The estimation processes used by students in grades five through eight as they responded to computational estimation test items were examined. Interview-based process descriptons were cross-validated using large group test data. Students had a strong mental set to round numbers to the nearest leading power of ten even when the items required other estimation processes. Performance differed by item format, types of numbers and operations in the items, and grade level of students.

Computational estimation has long been recognized as a basic mathematical skill, and recently it is receiving a great deal of attention in sets of curriculum recommendations (Reys, Rybolt, Bestgen, & Wyatt, 1981). While estimation can certainly be viewed as a skill, recent writers and researchers in the area emphasize its role in mathematical understanding. There appears to be an inextricable link between estimation in a number domain and understanding mathematical concepts in that domain such as order and number size, number properties, and meanings of operations. This general point is made repeatedly, although from different perspectives, by Trafton (1986), Reys (1986), and Carlow (1986). A more specific version of the same point is made by Leutzinger, Rathmell, & Urbatsch (1986), who emphasize critical links between estimation and conceptual understanding in the early stages of learning. Other writers make this connection in the domains of common fractions (Behr, Post, & Wachsmuth, 1986), decimal products (Vance, 1986), and percents (Allinger & Payne, 1986).

If this connection between estimation and conceptual understanding is as strong as these writers suggest, then it would seem to follow that an important by-product of learning to estimate is better conceptual understanding. Conversely, there are concepts that must be understood in order to acquire the flexible set of processes and decision-making rules needed to be a proficient estimator. Estimation should surely be a powerful and important corequisite for conceptual understanding.

Lending empirical support to the link between estimation and concept learning, Reys et al. (1981) found that good estimators in grades 7 through 12 and selected adults used a variety of estimation processes. In fact, the authors also point out that, among other characteristics, the good estimators had a good understanding of place value and number properties. Reys and his co-investigators identified the following three key processes used by the good estimators:

1. Reformulation (including various kinds of rounding and front-ending) or altering numerical data to produce a more mentally manageable form, but leaving the structure of the problem intact;

2. Translation or changing the mathematical structure of the problem to a more mentally manageable form, such as changing a sum of several nearly equal numbers to a product, and;

3. Compensation or adjusting an estimate to reflect numerical variation that came about as a result of translation or reformulation.

Unfortunately, what little computational estimation there is in textbooks and classrooms at present often fails to make this connection with conceptual understanding. Estimation is commonly viewed as equivalent to the following steps:

1. Round the numbers to be computed using the standard rules for rounding;

2. Mentally compute with the rounded numbers; and

3. Call the result the estimate.

This approach to estimation is one useful process that can give a reasonable estimate, yet it can surely be and sometimes is taught and learned as essentially a rote skill with no connection to understanding of any sort (Schoen, Friesen, Jarrett, & Urbatsch, 1981). However, Schoen et al. found that a meaningful approach to teaching estimation was better than rote practice in that it resulted in higher levels of transfer to verbal problem settings.

Estimation must clearly be taught from a broader perspective than just following the three rules above if the potentially powerful link between estimation and meaningful learning is to be made. Many authors, including those cited in the previous paragraphs, have put forth suggestions for teaching estimation in meaningful ways. However, as Reys (1986) points out, if estimation is to become a part of the curriculum then it is important that ways be designed to test the ability of students to estimate, and much needs to be done in that regard. Benton (1986) also cites testing difficulties as a major factor in limiting the number of research studies dealing with estimation. Few would argue with the need for concurrent development of the teaching and testing of estimation, but tests can only be viewed as facilitating meaningful teaching of estimation if the skill and understanding that are tested reflect accurately the processes and concepts that are the goals of teaching. For example, if a student can score very well on an estimation test by using the three rote skill steps described in the previous paragraph, the test may not facilitate the meaningful teaching of estimation. On the contrary, teachers and students may be tempted to focus on practicing this simple skill in order to do well on the test at the expense of the more important goals of estimation instruction. In that case, the test may affect the quality of instruction negatively, not positively.

Several different testing approaches and item formats have been used to test computational estimation (Reys, 1986). These approaches and item formats appear to test different processes and draw on the understanding of different concepts. It seems crucial that these testing approaches be studied to determine whether they test the broader goals of estimation instruction, thereby encouraging meaningful estimation instruction. One study that examined testing approaches was conducted by Rubinstein (1985). In a statistical analysis of eighth graders' performance on estimation items in different formats, she found a number of differences in difficulty. This suggests that there may be differences in requisite processes and understandings, too, but no attempt was made by Rubinstein to explain the reasons for these differences.

The purpose of the present study was to examine the processes that were used by middle school students as they responded to different types of estimation test items. The set of estimation processes examined was an adaptation from those identified by Reys et al. (1982) and described in general terms above. It is assumed in this study that mathematical concepts are inextricably linked to estimation skills. The same should, or at least could, be true in testing these skills. Thus, no attempt was made to write "pure estimation" items, but rather in many items conceptual understanding and estimation skill were deliberately combined.

METHOD

Subjects

In this study, there were three data collection phases, the open-ended, the interview, and the multiple-choice. The students in the open-ended phase were 65 sixth-grade students from two elementary schools and 57 eighth-grade mathematics students from one junior high school in a midwestern city. All of the participants were volunteers who had received no systematic instruction beyond their textbook's program.

For the interview phase, ten sixth- and ten eighth-grade students who had participated in the open-ended phase were chosen by their teachers. At each grade level, each student was identified as likely to be verbal in a one-to-one interview, two or three were judged by their teacher to be above average in mathematical ability, two or three below average, and the remainder about average.

The multiple-choice phase involved a total of 1376 students, 342, 336, 323, and 375 in grades 5, 6, 7, and 8, respectively. These were a randomly chosen 56% of all the students at those grade levels from 13 representative Iowa school districts, none of which had students who participated in the interviews. Thus, at each grade level about 70 students (62 to 79) completed each of the five forms of the test.

Instruments

Open-ended Test

A 13-item test was written with one item from each of the 12 number (whole number, fraction, decimal) by operation (addition, subtraction, multiplication, division) cells, except for division of fractions and division of decimals. The three remaining items mixed a whole number with either a fraction or a decimal.

Multiple-Choice Tests

Five equivalent forms of a 30-item multiple-choice test were constructed by first writing 30 item stems of which 25 were purely computational and five were word problems. The computational stems were written to fit the number and operation specifications in Table 1. The numbers and operations in the five word problems were two additions (one with decimals in the context of money, and one with whole numbers), one subtraction with decimals (money), one multiplication with whole numbers and one decimal (money) by whole number division.

Insert Table 1 about here

For each stem, five items were written, one in each of five formats designed to test different estimation processes or the same processes in different ways. The five formats for the 23 stems that contained no fractions were standard multiple choice (MC), operation in foils (OF), range in foils (RF), benchmark (BM), and order of magnitude (OM). Since order of magnitude choices are inappropriate for fractions, the OM format was replaced by operation in stem (OS), for the seven stems that contained fractions. The five item formats for two of the stems are given in Table 2.

Insert Table 2 about here

The estimation items were designed to elicit different processes depending not only on their stems but also on their choices or foils. For example, by including items for which correct choices were the result of different estimation processes the student was forced to use several processes. By making more than one foil in MC and OF items a result of a valid estimation process or by making one endpoint of the range in a foil in an RF item the result of a rounding and mental computation process, the student was forced to compensate.

The a priori specifications of processes that were tested by at least some items in each format are given in Table 3. These processes are (a) rounding by the usual rules to the closest power of ten or, in the case of mixed numerals, to the closest whole number (RC); (b) front-ending or rounding down to the power of ten of the leading digit or to the whole number part of a mixed numeral (FE); (c) other rounding including rounding all numbers up or some numbers up and others down (OR); (d) using compatible numbers or numbers relatively close to the given numbers for purposes of easily operating with the numbers in the item (CN); and (e) compensating or adjusting an estimate to reflect variations that might result from rounding or the use of some other adjustment process (CO). Furthermore, the rounding might be done to the leading digit, or a closer round might be used. This latter process was called refined rounding (RR). Finally, if a single estimate was required the student would need to use mental computation with the adjusted numbers (MC).

Insert Table 3 about here

The sample items given in Table 2 illustrate the steps that were taken toward making the five items for a given stem equivalent except for the format difference. In the following description of these steps, steps 1 and 2 apply to all item formats while step 3 is appropriate for all but BM and OM items.

1. The number of choices was always the same, namely four.

2. The correct answer, once placed randomly in an item, was kept in the same position for all five formats of the item.

3. The foils for MC items were written by using incorrect answers that appeared on the open-ended test and by analyzing the stem computation for likely process and execution errors. Once the foils for MC items were written, the foils in corresponding positions in other item formats were matched to them. Thus, the computation in a foil in an OF item yielded the number in the corresponding MC foil, the range in a foil in a RF item included the number in the corresponding MC foil, and the number in a foil in an OS item yielded results in the range given as the corresponding RF foil.

To compile the five equivalent forms of the multiple-choice test, the 30 item stems were placed in a reasonable order for a test, that is, items from stems that were likely to be easier were placed at the beginning with harder items near the end unless some logical content grouping suggested otherwise. Thus, the five word problem stems were placed at the end, the whole number computation stems were placed at the beginning and the decimal, fraction and mixed computation stems were placed in between. The forms of the test were then constructed using a Latin square procedure so that in each block of five consecutive item stems, each form contained one of each of the five item types. This is illustrated for stems 1 - 5 in Table 4. The pattern for the first five stems was then repeated five more times to complete the five 30-item test forms. For the seven item stems that contained fractions, OM items were replaced by OS items. Thus, each test form included the same 30 item stems in the same order and was comprised of five subtests each containing six items of the same format, except that the fifth format for whole number and decimal items was OM while for items containing fractions it was OS.

Insert Table 4 about here

Procedures

Open-ended Phase

The open-ended test was administered in early December by the classroom teachers using an overhead projector with a mask which allowed one item to be shown at a time. The teacher read the item (e.g., "2848 + 4123 is about?"), allowed ten seconds for the students to write their estimates on their answer sheets, and then displayed and read the next item. Thus, each item was read to the student and was visible to them for ten seconds. To further discourage exact computation, the students' answer sheets were darkened everywhere except for the spaces provided for the estimates.

The open-ended tests were scored using a scale that assigned 0, 1, or 2 points per item depending on whether the student's estimate fell within intervals determined by the highest and lowest numbers that would result from applying one of the following processes: front-end, round to closest, round up, or compatible number processes. Decisions to determine the acceptable intervals were made for each item, instead of trying to establish a general rule for all items. Students' responses to each open-ended item were also tabulated and used to infer the estimation processes employed by the students, an approach found by Schoen et al. (1981) to be in close agreement with process results from interviews. Independent judgments concerning the estimation process suggested by each answer were made by three doctoral students in mathematics education. Disagreements were discussed by the three along with the first two authors until concensus was reached.

Interview Phase

Three months after the open-ended test was administered, the ten sixth- and ten eighth-grade students were interviewed individually. Each student was asked to "think aloud" while responding to each of the ten interview items. These items were chosen to provide a good mix of the numbers, operations, and item formats on the entire test. In particular, two items were in each of the formats, MC, OF, RF, BM, and OM; four items involved multiplication while two involved each of the other three operations; and six items contained whole numbers while two each contained fractions and decimals. Items were placed in a different random order for each student interviewed.

The interviews were audio-taped and transcribed. A process sequence was coded for each student on each item using the following list: round to closest, front-end, round up, other rounding, compatible numbers, exact computation, refine or compensate, mentally compute with estimates, look at foils, compare, and choose from foils. Four coders independently coded the items for each student. Pairwise interrater agreements ranged from .86 to .94 based on a sample of 25 items, five from each student.

Multiple-Choice Phase

The multiple-choice test forms were administered at about the same time as the interviews. The five estimation test forms were stacked in order followed by four other experimental test units not related to this study. This stack of nine test forms was then repeated in order often enough to attain the number needed for the school districts to be tested. Each classroom set of tests was then formed by counting from the top of the stack. The teachers who administered the tests were directed to hand them out in the order in which they were stacked. In this way, the test forms were randomly distributed to students within classes, and about 11% of all the students completed each of the test forms.

While this 20-minute test was not timed by item, pilot work suggested that nearly all students did try to estimate and did not use rapid exact computation. Results in the interview phase of the study also support this contention. In part, this was accomplished by making the test too long to complete for virtually any fifth through eighth grader who made much use of exact computation. The directions also made it clear to the students that they would not have time to compute exactly.

The items were scored as right or wrong and individual item analyses were run. The item analysis included, for each item at each grade level, the difficulty index (percent of students answering correctly), discrimination index (biserial correlation between scores on the item and scores on the 30-item test form), and the percent who chose each foil. The item analyses were used, where appropriate, to provide support for the interview-based process and error analyses. A grade x test form x item format ANOVA using item difficulty and discrimination indices as dependent variables was also run to help describe item format and grade level effects.

RESULTS

Open-ended Phase

For each item, each answer, correct or incorrect, was analyzed to determine the estimation process the student most likely used to attain it, with a category for answers for which no process could be determined. The processes that were identified and recorded are RC, FE, OR, RR, CO, and exact computation (EC). In general, it was difficult to distinguish answers that arose from RR and those that were the result of CO so these categories were combined. Furthermore, in the two items in which RC and FE gave the same result, the answer was classified as RC. It is assumed that some mental computation was done in every case so this process was not recorded. A percent correct was also computed for each item based on a possible two points per item. Since students' processes and errors are dependent to a large extent on the types of numbers in the exercises, the results were analyzed separately for whole-number, decimal, and fraction items.

Results for each item are given in Table 5. On the whole-number items, students used the RC process 63% of the time. Students rarely used exact computation, and there was also little evidence of compensation or refinement. The division item was by far the most difficult, with 36 of 122 students making a place value error, usually giving a three digit estimate like 500.

Insert Table 5 about here

Students used the RC process on decimal items 50% of the time. See Table 6. Not surprisingly, exact computation was used by 35 students on the item, 5 + 6.43. On the item, 35 x 4.32, 50 students rounded 35 up to 40 and 4.32 down to 4 (classified as OR), while about the same number simply did the latter rounding (RC). More students used compensation or refinement than on whole number items, but this still only occurred about 8% of the time.

Insert Table 6 about here

Processes used to make estimates in fraction items were quite different from those used for whole numbers and decimals as Table 7 shows. Front-ending was used with about the same frequency as RC. A significant number of students used exact computation on the two items, 6 - 3 7/10 and 2 1/2 + 7 3/5, although on the latter item many students simply added numerators and denominators in the fractions and answered 9 4/7. On this same item, 20 students multiplied instead of added, a surprising result that was probably due to a mental set established by the preceding multiplication item, 3 7/8 x 6 1/2.

Insert Table 7 about here

Process Analysis for Multiple-choice and Interview Phases

The processes and errors on the items in the interview phase were analyzed for each item format. Each interview item was also included in the tests in the multiple-choice phase, and the item analyses for these items were used to cross-validate the interview-based process analysis.

Standard multiple-choice. The two interview items in the MC format along with their item analyses from the multiple-choice phase are given in Table 8. For item MC1, a few students who correctly chose 1300, first rounded 4329 to 4300 and 2847 to 3000 and mentally computed, thus using no compensation. More often, however, students rounded the given numbers to 4000 and 3000 or 2800, respectively. After checking the foils, they compensated for the rounding or simply chose 1300 because it was closest to their estimate. By far the most common error was 1000, because most interviewed students simply rounded to 4000 - 3000 and failed to check whether that was the closest of the given estimates. The item analysis indicates that few students answered this item correctly, and, especially in grades five and six, it did not discriminate well. Consistent with the interview results, about as many students chose the incorrect foil 1000 as chose the correct answer.

Insert Table 8 about here

For item MC2, most interviewed students added the whole numbers 7, 3, and 1 to get 11, although two first tried but failed at exact computation. Upon checking the foils, eight of the 20 simply chose 11. The others compensated, noting that the fractions they had dropped in the rounding process totaled about one-half. One student decided on 11 1/2 or 12 1/2 since both contained fractions, but then over-compensated and chose 12 1/2. The item analysis indicates that this, too, was a difficult item with a large grade level effect. About 36% of all students who took the multiple-choice test chose 12 1/2, while only about 10% chose 11, suggesting that an overcompensation error was more common than no compensation at all.

Operation in Foils. The two interview items that were in the OF format along with their item analyses from the multiple-choice phase are given in Table 9. Item OF1 was answered correctly by all interview students and by about 87% of all students in the multiple-choice phase. Students simply rounded 588 and 39 to 600 and 40, respectively, and chose 600 x 40.

Insert Table 9 about here

The analogous process with similarly successful results was used for item OF2 No compensation or mental computation was required or used for either item. Note also that these items discriminated very well in the multiple-choice phase.

The multiple-choice data, however, makes it clear that not all OF items were as straightforward as these two. For example, the first OF sample item given in Table 2 had an average difficulty of .52 across the four grades. On this item, many students chose 2000 + 8000 + 3000, which at first glance may appear to be the result of applying the RC process, even though 2000 + 1000 + 3000 was a much closer estimate.

Range in Foils. The two interview items in the RF format along with their item analyses from the multiple-choice phase are given in Table 10. For item RF1, students found it difficult to decide between 18 and 19, the correct answer, and 17 and 18. One interviewee rounded the given sum to 4 1/2 + 14 which fell into the correct range, but all other correct responses involved compensation up from 4 + 14 or down from 5 + 14. The most common error was to use 4 + 13 and either not compensate or under-compensate and choose 17 and 18. One student rounded to 5 + 14 then, confusing the direction, compensated up because both numbers had been rounded up. The item analysis from the multiple-choice phase shows that this was a very difficult item, but it discriminated well. Over twice as many students chose the incorrect range, 17 and 18, than chose the correct answer, suggesting that either no compensation or under-compensation was the norm on this item.

Insert Table 10 about here

For item RF2, only two of the 20 interviewees, both sixth graders, chose the correct answer, 900 and 1000. They both rounded the given numbers to 400 x 2.5, mentally computed to get 1000, and then compensated downward. The most common process was to round to 400 x 2 to get 800, then either under-compensate and choose 800 and 900, or compensate in the wrong direction and choose 600 and 800. The item analysis from the multiple-choice phase shows that only about 10% of students in any grade level chose the correct answer.

Not only was there no grade level effect, the item discrimination was -.46 in grade eight, indicating that the older and brighter students consistently chose 800 and 900.

Benchmark. The two interview items that were in the BM format along with their item analyses from the multiple-choice phase are given in Table 11. For item BM1, all interviewees rounded the given numbers to 800 - 600, got 200, and then tried to compensate. Correct compensation processes included (a) refining to 800 - 220 and noting that this is less than 800 - 200, (b) noting that 804 - 217 < 804 - 204 or 600, and (c) noting that 4 - 17 is negative so 804 - 217 must be less than 600. Three types of errors were made: (a) deciding that 800 - 200 is more than 600 since 804 and 217 were rounded down, (b) choosing c after noting correctly that 804 - 217 < 850 - 200, and (c) after incorrectly deciding on "more", choosing foil a rather than b because "800 - 300 looks more like an estimate than 800 - 250." In the multiple-choice phase, about 50% of the students chose the correct answer, and the item discriminated well. The most common error was foil c, chosen by about 23% of the students.

Insert Table 11 about here

The most common process for item BM2 was to round 521 x 29 to 500 x 30, then mentally compute to get 15,000 which is less than 18,000, and finally choose foil c because 600 x 30 = 18,000 or because "500 x 30 is closer to 600 x 30 than to 500 x 40" (in foil d). Errors included choosing foil b because 500 x 30 appears in it and choosing foil d because 500 x 30 < 500 x 40 (ignoring the fact that 500 x 40 is not 18,000) or because 500 x 40 is closer to 500 x

30 than is 600 x 30 since "it is only off in the smaller number." Interestingly enough, there was no grade level effect for this item in the multiple-choice phase with a little over 40% of the students at each grade level choosing the correct answer. The item discriminated moderately well except in the fifth grade, and the most common error was choosing foil b, the foil that contained 500 x 30.

Order of Magnitude. The two interview items that were in the OM format along with their item analyses from the multiple-choice phase are given in Table 12. Three different processes were used in item OM1. First, several students rounded 2475 ÷ 42 to 2000 ÷ 40, mentally computed to get 50, then upon seeing the foils chose 60 as the closest one. One student rounded the given numbers to compatible numbers, 2500 ÷ 50, and then chose 60 as above. Second, one student started as if to do exact long division and saw that the quotient would be "fifty some" and chose the closest foil, 60. Third, several students looked at the foils first and decided that 60 would be a good estimate since 2475 ÷ 42 was close to 2400 ÷ 40. All students used a variation of one of the above processes, but eight of the 20 made a place value error and chose 600. In the multiple-choice phase this item was essentially a two-choice item, with almost as many students choosing 600 as 60. The item was considerably easier for eighth graders than for fifth graders, but it discriminated better at grade five.

Insert Table 12 about here

For item OM2, most students simply rounded the given numbers to 8x1, looked at the foils, and chose 10. Two students made errors because they failed to understand that 1.27 was about 1 and eventually guessed at an answer. In the item analysis from the multiple-choice phase, fifth and sixth graders found this item to be much more difficult than seventh and eighth graders did, although the item discriminated very well at all grade levels.

Grade and Item Format Effects from Multiple-choice Phase

A grade x test form x item format ANOVA was run using item difficulty as the dependent variable. However, the significant higher order interactions that appeared made it clear that a confounding variable was causing noise in the data. Upon examination, it was found that items such as RF2 in Table 10 for which the rounding strategy led to an incorrect answer had much lower difficulty and discrimination indices than items with the same stem but a format for which rounding did not have such an effect. Two grade x test form x rule ANOVAs were run with item difficulty and discrimination indices as dependent variables. The rule variable had two levels, R1 and R2, depending upon whether or not the RC process led to an incorrect answer in at least one item format for that stem. Figure 1 shows the highly significant ordinal rule x grade interaction ($\underline{F}[3,12] = 34.23$, $\underline{p} < .0001$) when difficulty index is the dependent variable. Compared to other items, there was little improvement by grade level on items in which the usual rounding process led to a wrong answer. Table 13 gives the mean difficulties and discriminations by grade and rule. For difficulties, both the rule ($\underline{F}[1,4] = 74.56$, $\underline{p} < .0001$) and grade ($\underline{F}[3,12] = 158.36$, $\underline{p} < .0001$) main effects were significant. For discriminations, the grade main effect ($\underline{F}[3,12] = 6.33$, $\underline{p} < .0004$) was somewhat smaller, but both main effects were still significant (rule effect: $F[1,4] = 130.43$, $p < .0001$). Follow-up Tukey's Studentized Range Tests of pairwise differences between grade levels showed the expected result for mean difficulty indices, 8 > 7 > 6 > 5, each at the .05 level, and for discriminations, 8 is greater than any of the other grade levels but grade 5, 6, and 7 means did not differ significantly. Most students at all grade levels and all levels of overall performance on the test were using the RC process and were not compensating or refining when that was required.

Insert Table 13 about here

Insert Figure 1 about here

In order to eliminate the confounding rule variable and get a fair measure of the effects of the various item formats and their interaction with grade level on difficulty indices, the nine R1 item stems were eliminated. A grade x test form x item format ANOVA was then run with item difficulties of the remaining 21 item stems as the dependent variable. Only the four item formats that were common to all item stems were included, that is, MC, OF, RF, and BM. Thus, 336 individual item difficulty indices, 21 in each format at each grade level, were included in the ANOVA. Mean difficulties by grade and item format are given in Table 14. There were no significant interactions but both the grade ($F[3,12] = 52.96$, $p < .0001$) and format (($F[3,12] = 26.88$, $p < .0001$) main effects were significant. Tukey's Studentized Range Test of pairwise differences at the .05 level showed the expected grade differences in mean difficulty indices, 8 > 7 > 6 > 5. The significant pairwise format differences were OF easier than each of the other three and BM and MC both easier than RF.

Insert Table 14 about here

The OS format was used for the six fraction items, and the OM format for the other 24 items. The mean difficulties for the items in these formats for which rounding did not lead to an incorrect answer are given in Table 15. While OS items were quite difficult, it is important to note that items containing fractions were more difficult in general than those not containing fractions. In fact, the mean difficulty on these four fraction items across all grades and the four formats other than OS was .37. Similarly, the 17 items in the OM format only contained whole numbers and decimals. The overall mean difficulty for these 17 items in the four formats other than OM was .53, considerably less than the OM mean of .69.

Insert Table 15 about here

DISCUSSION

Consistent results from all three phases of this study provide a clear message that middle school students, regardless of grade and ablity levels, think of estimation with whole numbers and decimals as equivalent to the rote "round to the closest" approach. They round the numbers to the leading powers of ten and mentally compute, but they rarely compensate, refine, use compatible numbers, or illustrate any of the other estimation processes associated with conceptual understanding, even when the test item specifically requires them to do so.

If testing is to have a facilitative role in promoting meaningful estimation instruction, then it seems that one important criterion for judging an estimation test must be the extent to which it measures the process and concept goals of such instruction. Of the test item formats in this study, the open-ended would probably be judged to have the most face validity in that it appears to test only estimation. Yet the data show that students succeeded reasonably well on the open-ended test by using the "round to the closest" approach, making use of other estimation processes only rarely and then usually on fraction items. Therefore, such a test may motivate students to learn or teachers to teach only that process and not a broader understanding of estimation processes.

On the other hand, the multiple-choice items in all formats except the benchmark appeared to test in combination with related number concepts, various aspects of estimation, not just the "round to closest" process. The fact that most students failed to apply other fruitful processes when responding to these items points out (a) the strength of their belief that estimation is simply rounding to the closest leading power of ten and, hence, (b) the great extent to which estimation instruction presently falls short of its potential. Further efforts in the teaching and testing of computational estimation must focus on overcoming this narrow view of estimation held so strongly by most students.

REFERENCES

Allinger, G. D., & Payne, J. N. (1986). Estimation and mental arithmetic with percent. In H. L. Schoen (Ed.), Estimation and Mental Computation, 1986 Yearbook of the National Council of Teachers of Mathematics (pp. 141-155). Reston, VA: The National Council of Teachers of Mathematics.

Behr, M. J., Post, T. R., & Wachsmuth, I. (1986). Estimation and children's concept of rational number size. In H. L. Schoen (Ed.), Estimation and Mental Computation, 1986 Yearbook of the National Council of Teachers of Mathematics (pp. 103-111). Reston, VA: The National Council of Teachers of Mathematics.

Benton, S. E. (1986). A summary of research on teaching and learning estimation. In H. L. Schoen (Ed.), Estimation and Mental Computation, 1986 Yearbook of the National Council of Teachers of Mathematics (pp. 239-248). Reston, VA: The National Council of Teachers of Mathematics.

Carlow, C. D. (1986). Critical balances and payoffs of an estimation program. In H. L. Schoen (Ed.), Estimation and Mental Computation, 1986 Yearbook of the National Council of Teachers of Mathematics (pp. 93-102). Reston, VA: The National Council of Teachers of Mathematics.

Leutzinger, L. P., Rathmell, E. C., & Urbatsch, T. D. (1986). Developing estimation skills in the primary grades. In H. L. Schoen (Ed.), Estimation and Mental Computation, 1986 Yearbook of the National Council of Teachers of Mathematics (pp. 82-92). Reston, VA: The National Council of Teachers of Mathematics.

Reys, B. J. (1986). Teaching computational estimation: Concepts and strategies. In H. L. Schoen (Ed.), Estimation and Mental Computation, 1986 Yearbook of the National Council of Teachers of Mathematics (pp. 31-44). Reston, VA: The National Council of Teachers of Mathematics.

Reys, R. E. (1986). Evaluating computational estimation. In H. L. Schoen (Ed.), Estimation and Mental Computation, 1986 Yearbook of the National Council of Teachers of Mathematics (pp. 225-238). Reston, VA: The National Council of Teachers of Mathematics.

Reys, R. E., Rybolt, J. F., Bestgen, B. J., & Wyatt, J. W. (1982). Processes used by good computational estimators. Journal for Research in Mathematics Education, 13, 183-201.

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

Schoen, H. L., Friesen, C. D., Jarrett, J. A., & Urbatsch, T. D. (1981). Instruction in estimating solutions of whole number computations. Journal for Research in Mathematics Education, 12, 165-178.

Trafton, P. R. (1986). Teaching computational estimation: Establishing an estimation mind-set. In H. L. Schoen (Ed.), Estimation and Mental Computation, 1986 Yearbook of the National Council of Teachers of Mathematics (pp. 16-30). Reston, VA: The National Council of Teachers of Mathematics.

Vance, J. H. (1986). Estimating decimal products: An instructional sequence. In H. L. Schoen (Ed.), <u>Estimation and Mental Computation</u>, 1986 Yearbook of the National Council of Teachers of Mathematics (pp. 127-134). Reston, VA: The National Council of Teachers of Mathematics.

Table 1

Number of Computational Items by Number Type and Operation on Multiple Choice Tests

Operation	Whole Number	Fraction	Decimal	Whole & Decimal
Addition	2	2	2	1
Subtraction	2	2	2	2
Multiplication	2	3	2	1
Division	2	0	0	0

Table 2

Five Item Formats for Two Sample Stems

Item Format	Item for 1926 + 851 + 3273	Item for $6\frac{3}{4} \times 5\frac{1}{3}$
MC	The closest estimate of 1926 + 851 + 3273 is ______. 1) 5000 3) 7000 *2) 6000 4) 13,000	The closest estimate of $6\frac{3}{4} \times 5\frac{1}{3}$ is ______. *1) 35 3) 30 2) 42 4) 24
OF	The closest estimate of 1926 + 851 + 3273 is ______. 1) 1000 + 1000 + 3000 *2) 2000 + 1000 + 3000 3) 2000 + 1000 + 4000 4) 2000 + 8000 + 3000	The closest estimate of $6\frac{3}{4} \times 5\frac{1}{3}$ is ______. *1) 7 x 5 3) 6 x 5 2) 7 x 6 4) 6 x 4
RF	1926 + 851 + 3273 is between ______. 1) 4500 and 5500 *2) 5500 and 6500 3) 6500 and 7500 4) 12,500 and 13,500	$6\frac{3}{4} \times 5\frac{1}{3}$ is between ______. *1) 32 and 37 3) 27 and 32 2) 37 and 42 4) 22 and 27

Table 2 Continued

BM	Is 1926 + 851 + 3273 more or less than 7000? 1) Less, because 1926 + 851 + 3273 is less than 2000 + 8000 + 3000 *2) Less, because 1926 + 851 + 3273 is less than 2000 + 1000 + 4000 3) More, because 1926 + 851 + 3273 is more than 1000 + 8000 + 3000 4) More, because 1926 + 851 + 3273 is more than 1800 + 800 + 3000	Is $6\frac{3}{4} \times 5\frac{1}{3}$ more or less than 7×6? *1) Less, because $6\frac{3}{4} \times 5\frac{1}{3}$ is less than 7×6 2) Less, because $6\frac{3}{4} \times 5\frac{1}{3}$ is less than 7×7 3) More, because $6\frac{3}{4} \times 5\frac{1}{3}$ is more than 6×4 4) More, because $6\frac{3}{4} \times 5\frac{1}{3}$ is more than 6×5
OM, OS	The closest estimate of 1926 + 851 + 3273 is ______. 1) 600 3) 60,000 *2) 6000 4) 600,000	$6\frac{3}{4} \times$ ______ is between 31 and 37. *1) $5\frac{1}{3}$ 3) $7\frac{1}{3}$ 2) $6\frac{1}{3}$ 4) $4\frac{1}{3}$

Table 3

Estimation Processes Tested by Items in Different Formats

Format	Process						
	RC	FE	OR	RR	CN	CO	MC
MC	X	X	X	X	X	X	X
OF	X	X	X	X	X	X	
RF	X	X	X	X	X	X	X
BM	X	X	X	X	X	X	X
OM	X						X
OS	X	X	X	X	X	X	X

Table 4

Item Formats for the First Five Items in Each Multiple-choice Test Form

	Test Form				
Stem	A	B	C	D	E
1	MC	OF	RF	BM	OM
2	OF	RF	BM	OM	MC
3	RF	BM	OM	MC	OF
4	BM	OM	MC	OF	RF
5	OM	MC	OF	RF	BM

Table 5

Number of Occurrences of Processes and Errors on Whole Number Open-ended Items

		Processes					Errors		
Item	% Corr.	RC	FE	OR	EC	CO,RR	PV	WO	Other
2848 + 4163	75	76	20	4	5	5	0	5	18
6273 - 4926	67	75	25	5	6	7	6	8	20
32 x 68	59	70	19	2	10	1	12	2	27
4153 ÷ 79	32	87	0	5	0	9	36	1	40

Table 6

Number of Occurrences of Processes and Errors on Decimal Open-ended Items

		Processes					Errors		
Item	% Corr.	RC	FE	OR	EC	CO,RR	PV	WO	Other
5 + 6.43	75	58	--	17	35	5	0	7	17
1.8 + 4.37	68	63	16	17	8	13	1	7	23
19.13 - 7.84	62	61	22	10	4	5	1	7	34
35 x 4.32	62	49	--	50	0	16	8	3	26
8.8 x 3.3	68	72	19	2	10	8	5	1	18

Table 7

Number of Occurrences of Processes and Errors on Fraction Open-ended Items

		Processes					Errors		
Item	% Corr.	RC	FE	OR	EC	CO,RR	PV	WO	Other
$2\frac{1}{2} + 7\frac{3}{5}$	61	26	31	6	16	1	0	20	10
$6 - 3\frac{7}{10}$	68	65	17	0	26	1	0	1	27
$13\frac{1}{3} - 8\frac{4}{5}$	74	50	36	6	8	1	0	3	22
$3\frac{7}{8} \times 6\frac{1}{2}$	56	25	38	37	8	30	1	3	16

Table 8

Items and Partial Item Analyses (% choosing each foil) for MC Interview Items

	(MC1) The closest estimate of 4329 - 2847 is ______.					(MC2) The closest estimate of $7\frac{1}{10} + 3\frac{1}{3} + 1\frac{1}{5}$ is ______.				
	a) 2000		c) 1000			a) 11			c) $12\frac{1}{2}$	
	b) 2300		d) 1300			b) $11\frac{1}{2}$			d) 14	
Grade	a	b	c	d*	Disc	a	b*	c	d	Disc
5	19	43	21	16	.12	10	26	35	19	.50
6	6	24	35	33	-.01	9	34	37	14	.37
7	16	9	36	38	.30	10	38	35	10	.29
8	16	17	33	33	.39	11	49	36	4	.51
Total	14	23	31	30		10	37	36	12	

Table 9

Items and Partial Item Analyses (% choosing each foil) for OF Interview Items

	(OF1) The closest estimate of 588 x 39 is ______. a) 500 x 40 c) 600 x 30 b) 500 x 30 d) 600 x 40					(OF2) The closest estimate of 5927 ÷ 32 is ______. a) 6000 ÷ 25 c) 5000 ÷ 40 b) 6000 ÷ 30 d) 5000 ÷ 50				
Grade	a	b	c	d*	Disc	a	b*	c	d	Disc
5	6	6	6	83	.48	3	76	13	7	.64
6	4	1	6	88	.88	6	83	0	6	.39
7	5	2	8	85	.51	1	93	5	16	.75
8	3	1	4	91	.67	0	92	7	0	.75
Total	5	3	6	87		3	86	6	7	

Table 10

Items and Partial Item Analyses (% choosing each foil) for RF Interview Items

	(RF1) $4\frac{1}{2} + 13\frac{4}{5}$ is between ______.					(RF2) 397.8 x 2.49 is between ______.				
	a) 18 and 19	c) 16 and 17				a) 900 and 1000	c) 800 and 900			
	b) 17 and 18	d) 19 and 20				b) 1000 and 1200	d) 600 and 800			
Grade	a*	b	c	d	Disc	a*	b	c	d	Disc
5	9	41	14	8	.38	11	33	17	36	.31
6	25	53	13	4	.68	7	28	33	31	-.13
7	29	47	6	15	.50	10	15	37	39	.13
8	35	51	9	4	.55	10	15	49	25	-.46
Total	25	48	11	8		10	23	34	33	

Table 11

Items and Partial Item Analyses (% choosing each foil) for BM Interview Items

	(BM1) Is 804 - 217 more or less than 600?					(BM2) Is 521 x 29 more or less than 18,000?				
	a) More, because 804 - 217 is more than 800 - 300					a) More, because 521 x 29 is more than 500 x 20				
	b) More, because 804 - 217 is more than 800 - 250					b) More, because 521 x 29 is more than 500 x 30				
	c) Less, because 804 - 217 is less than 850 - 200					c) Less, because 521 x 29 is less than 600 x 30				
	d) Less, because 804 - 217 is less than 800 - 200					d) Less, because 521 x 29 is less than 500 x 40				
Grade	a	b	c	d*	Disc	a	b	c*	d	Disc
5	9	19	23	45	.44	15	28	44	13	.10
6	14	13	21	50	.29	15	26	44	15	.30
7	11	13	27	47	.44	9	41	42	8	.31
8	11	8	20	61	.44	4	32	42	21	.26
Total	11	13	23	51		11	32	43	14	

Table 12

Items and Partial Item Analyses (% choosing each foil) for OM Interview Items

	(OM1) The closest estimate of 2475 ÷ 42 is ______. a) 6000 b) 600 c) 60 d) 6					(OM2) The closest estimate of 7.85 x 1.27 is ______. a) .10 b) 1.0 c) 10 d) 100				
Grade	a	b	c*	d	Disc	a	b	c*	d	Disc
5	9	48	31	8	.55	10	17	30	38	.60
6	15	32	46	4	.32	12	18	45	23	.61
7	11	21	65	3	.50	9	11	70	9	.67
8	3	41	54	3	.20	8	13	69	7	.59
Total	10	36	49	5		10	15	54	19	

Table 13

Mean Difficulties (and Discriminations) for Items by Rounding Rule and Grade Level

		Grade Level				
Rule	N	5	6	7	8	Total
R1	9	.26(.26)	.28(.23)	.32(.26)	.34(.34)	.30(.28)
R2	21	.40(.52)	.54(.52)	.61(.57)	.68(.58)	.56(.55)
Total	30	.35(.43)	.45(.42)	.50(.46)	.56(.50)	.47(.45)

Table 14

Mean Difficulties by Grade and Item Format for 21 R2 Items

	Grade Level				
Format	5	6	7	8	Total
OF	.48	.60	.68	.74	.62
MC	.31	.46	.54	.63	.49
BM	.38	.49	.51	.57	.49
RF	.29	.41	.50	.56	.44
Total	.36	.49	.56	.63	.51

Table 15

Mean Difficulties of R2 Computational Items in the OM and OS Format by Grade Level

		Grade Level				
Format	N	5	6	7	8	Total
OM	17	.51	.63	.78	.84	.69
OS	4	.26	.30	.41	.49	.37

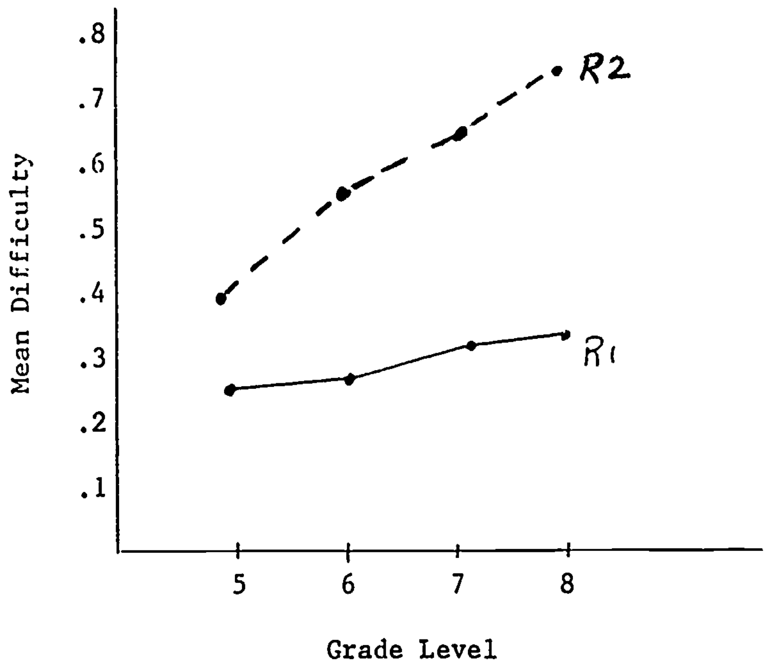

Figure 1. Mean Difficulties of R1 and R2 Items by Grade Level