AMAL HUSSAIN ALAJMI and ROBERT REYS

EXAMINING EIGHTH GRADE KUWAITI STUDENTS' RECOGNITION AND INTERPRETATION OF REASONABLE ANSWERS

Received: 2 March 2008; Accepted: 18 May 2009

ABSTRACT. This research documents Kuwaiti eighth grade students' performance in recognizing reasonable answers and the strategies they used to determine reasonableness. The results from over 200 eighth grade students show they were generally unable to recognize reasonable answers. Students' performance was consistently low across all three number domains (whole numbers, fractions, and decimals). There was no significant difference in students' performance on items that focused on the practicality of the answers or on items that focused on the relationships of numbers and the effect of operations, or on both. Interview data revealed that 35% of the students' strategies were derived from two criteria for judging answers for reasonableness: the relationships of numbers and the effect of operations, and the practicality of the answers. They used strategies such as estimation, numerical benchmarks, real-world benchmarks, and applied their understanding of the meaning of operations. However, over 60% of the students' strategies were procedurally driven. That is, they relied on algorithmic techniques such as carrying out paper-and-pencil procedures. Additionally, some of the students' strategies reflected misunderstandings of how and when to apply certain procedures. Given these findings, mathematics education in Kuwait should shift the emphasis from paper-and-pencil procedures and provide systematic attention to the development of number sense and computational estimation so Kuwaiti students will be more adept at recognizing reasonable answers.

KEY WORDS: estimation, international studies, middle grade, number sense, reasonableness, reasonable answers, students' strategies

Recognizing reasonable answers is an important component of number sense that underlies the success of computational results, including exact computation and estimation. It is practical and used in a wide array of daily activities, ranging from determining prices, making change, deciding on the dosage of a medicine, to choosing proper sizes of clothing.

Limited research has been conducted on students' ability to judge the reasonableness of answers (Sowder, 1992; Verschaffel et al., 2007). Reasonableness has been primarily studied as a secondary aspect of research focused on students' understanding of estimation (Reys et al., 1980; Vance, 1986). A few studies have investigated the reasonableness of exact calculations, particularly when working with decimals (Bell et al., 1981; Hiebert & Wearne, 1986).

Criteria for Determining Reasonableness

Theoretical discussions of the reasonableness of answers have focused on the meaning of reasonableness and the criteria upon which students judge the reasonableness of answers. In this paper, recognition of a reasonable answer refers to the likelihood that an answer that has been obtained (or provided) is an acceptable or a fairly good response for a mathematics problem. Research suggests that students use one of two criteria to judge answers for reasonableness (Gagne, 1983; Johnson, 1979; Hiebert, 1984; Reys, 1985; McIntosh & Sparrow, 2004): (1) *number relationships and the effect of operations* and (2) *practicality of the answer*. The first criterion reflects an understanding of number relationships and the effect of operations. Using this criterion, an individual will be able to identify the boundary of a reasonable answer. For example, consider the subtraction of two fractions, $9/11-4/5$. Research shows that some students will subtract numerators and denominators and report 5/6 as a result. Yet students utilizing number sense will immediately see that both fractions are close to 1, so the difference must be small. They also realize that a difference of 5/6 is close to 1 and, therefore, unreasonable.

The second criterion is related to the practicality of the answer. Students using this criterion compare the answers with what makes sense in their daily lives. They examine whether the magnitude of the answer and the type of numbers make sense based on their own experience. For example, the price of six tickets for a movie cannot be $61. Or a fraction cannot be an answer for a number of people on a trip or the number of cars on the highway. The extent to which students utilize this criterion depends on the range of their real-world experiences and their willingness to reflect on them to make a judgment.

These two criteria for judging an answer for reasonability are interrelated. When an individual reflects on an answer, he or she can judge the result using one or both criteria, depending on the type of problem as well as the numbers and operations involved.

A theoretical underpinning for recognizing reasonableness of answers is provided by the characterization of mathematical proficiency (Kilpatrick et al., 2001) as well as a framework for number sense (McIntosh et al., 1992). More specifically, identifying and recognizing reasonableness requires an amalgamation of qualities from these two models—conceptual understanding (number sense and work with operations), strategic competence (work flexibility with numbers), adaptive reasoning (observe and explain relationships), procedural fluency (operation with whole numbers, fractions, and decimals), and productive disposition (make connections to the real world). It seems likely that the extent to which students possess these qualities, these students will be proficient in identifying and recognizing the reasonableness of answers.

Recognizing reasonable answers is valued as an important learning goal in major mathematics documents in the United States of America (National Committee on Mathematical Requirements, 1923; National Council of Supervisions of Mathematics, 1989; National Council of Teachers of Mathematics, 1980, 1989, 2000). However, attention to reasonable answers is not a high priority in countries around the world (Reys & Nohda, 1994; Yang, 2005). Kuwait is one of the countries that has given very little attention in the mathematics curriculum to developing reasonableness.

The Educational System in Kuwait

This section provides some background about the educational system in Kuwait. It is based on the experiences of one of the authors, first as a Kuwaiti student, later as a Kuwaiti mathematics teacher, and now as an educator preparing future teachers in Kuwait.

Kuwait is a small country that has a small population and great opportunities for jobs. Therefore, more than 100 different nationalities work in Kuwait, including teachers and educators. Both public and private schools were created to accommodate the needs of the variety of students. Public schools, limited to Kuwaiti students and to children of those who work in the Kuwaiti government, have a national textbook at each grade level for all subjects, including mathematics. The private schools include both an Arabic and a non-Arabic curriculum; it is open for all students, including Kuwaitis. While the majority of Arabic private schools follow the same curricula as the public schools, non-Arabic private schools such as British, American, Indian, and Pakistani schools use curricula from their home countries. All private schools follow the administrative policies of the Kuwaiti Ministry of Education (Ministry of Education, 2006a, b).

Due to the shortage of mathematics teachers, the Kuwaiti Ministry of Education employs teachers from other Arabic countries, such as Egypt and Syria. These teachers usually are graduates of mathematics education programs that are similar to those offered in Kuwait.

All mathematics teachers in public schools follow the national textbook and the curricular plan of the Ministry of Education to determine what, when, and how to teach mathematics. They focus on goals specified by the ministry. There is little variation in the content taught and how it is taught.

Two levels of support are provided for teachers to help unify their teaching methods: school-level and district-level support. School-level support is provided by senior teachers who have experience in teaching mathematics and demonstrated excellence in teaching. These senior teachers help direct other teachers in the school to reach the goals that are set by the ministry, working closely with teachers in planning and carrying out their lessons. They identify general lesson plans for each topic they teach, specifying the goals for each lesson and identifying exercises, homework, assignment plans, and an examination plan.

The school district provides mathematics advisors who visit schools monthly. Advisors meet with teachers to clarify the plan of the ministry and the goals for teaching mathematics topics, as well as the mathematics knowledge, teaching strategies, and any new issues suggested by the ministry. They also monitor the number of students performing at a low level and the performance of new teachers in the school. Both mathematics advisors and senior teachers visit classrooms and provide advice.

The Kuwaiti mathematics curriculum focuses on procedural knowledge, following standard algorithms and finding exact answers (Alajmi, 2004, 2009). As a result, rules and procedural knowledge dominate computation, with no systematic attention given to developing computational estimation or encouraging students to reflect on answers to determine if they seem reasonable. Teachers stress procedural rules to solve the problem as the way to check the reasonableness of an answer. They often encourage students to evaluate the accuracy of their answers by performing another written computation, such as using multiplication to check the results of a division problem or addition to check the answer of subtraction (Alajmi & Reys, 2007). Neither the curricular or instructional focus in Kuwaiti classrooms is on the *reasonableness* of answers, but on verifying the *correctness* of exact calculations.

This approach does not promote the development of the number sense necessary to judge the reasonableness of an answer (Reys & Yang, 1998, Yang, 2005). Students in such a school setting are unlikely to examine the reasonableness of their answers because they do not see any need for this evaluation. They assume their answers are correct (Hiebert, 1984; Hope, 1989).

There has been no previous research conducted specifically about Kuwaiti students' ability to recognize reasonable answers. However, the Third International Mathematics and Science Study (Beaton et al., 1996) revealed that the number sense of Kuwaiti middle school students was below the international average. For example, on an item that assessed understanding the relative size of fractions, students were asked to write a fraction larger than 2/7. Only 37% of the Kuwaiti students provided a correct response, compared with 75% of students internationally.

This study reports students' performance in an area that does not depend on exact solutions. More specifically, this study addresses two research questions:

1. What is the performance of Kuwaiti eighth grade students in recognizing the reasonableness of answers to mathematical problems?
2. What strategies do Kuwaiti eighth grade students use to determine the reasonableness of an answer?

It is hoped this report will bring attention to the need for more curricular and instructional consideration for developing students' ability to determine the reasonableness of answers in Kuwaiti mathematics programs.

METHODOLOGY

The study made use of two instruments, the Reasonable Answer Test (RAT) and the Reasonable Answer Interview (RAI). The RAT was used to assess students' performance in recognizing the reasonableness of answers. The RAI focused on examining the processes students used to determine the reasonableness of answers. Validity of the instruments was provided through face validity. Mathematics educators from the United States reviewed the draft of the RAT. The modified instruments were then field tested with Kuwaiti eighth grade students and revised from information gathered from the piloting. Then, the final version of the RAT was tested for reliability through a test–retest method. One eighth grade girls' class and one eighth grade boys' class took the RAT; after 2 weeks, the same students took the RAT for a second time. For the girls, the reliability was 0.82; and for the boys, it was 0.74.

Context for the Study

This study was done in Kuwaiti public schools. There are six school districts in Kuwait. There is little variation among these school districts with regard to socioeconomic status. The overwhelming majority of Kuwaiti families can be classified as middle class; the remainder are wealthy. There is no difference in the achievement level among these school districts in elementary and middle school levels as reported by the Minister of Education. Students are assigned to schools according to their local residency. Students in Kuwaiti schools are also separated by sex. Thus, girls go to a different school than boys. In Kuwait, the collection of data from students does not require parents' permission. Permission to work with these students was gained through the Ministry of Education (Ministry of Education, 2006a). Given the similarity among the school districts and the schools in Kuwait, the researchers selected one district as a research site.

Participants

From the 36 middle schools in the chosen district (18 boys' schools and 18 girls' schools), two middle schools for boys and two middle schools for girls were randomly selected to participate in this study. The first girls' school and the first boys' school were selected randomly to pilot the RAT and the RAI and to examine the reliability of the RAT.

The data to answer the research questions were collected from the second girls' school and the second boys' school. There were four eighth grade classes in each of these two schools and each of these classes had between 25 and 30 students. All eighth grade students who were in their classrooms on the test day took the RAT, a total of 115 eighth grade girls and 108 eighth grade boys (a total of 223 eighth grade students). A sample of 24 eighth grade students (12 girls and 12 boys) was selected for the interview, including the one student who scored the highest on the RAT and 23 other students who were randomly selected (12 girls and 11 boys).

Reasonable Answer Test

The RAT included 25 items. Each item required the eighth grade students to judge the reasonableness of an answer provided for a mathematics problem or select a reasonable answer for a mathematics problem and then explain how they arrived at their judgment. The items on the RAT made use of the four basic operations (addition, subtraction, multiplication, and division) with whole numbers, fractions, and decimals. The RAT included items that were presented on mathematics statements and problems in context, including multiple choice, short answer items, and items that had more than one reasonable answer (see Table 1). All items on the RAT were developed by the researchers except for two items that were adapted from other research (Hiebert & Wearne, 1986; Yang, 1995).

TABLE 1

Examples of RAT items that represent three number domains and two criteria for reasonable answers and students' responses on these items

Item	*Incorrect response*	*Correct response*	*Correct response with adequate explanation*
1. Dalal used a calculator to solve the following mathematics problem. Your task is to look at Dalal's answer and tell if it can be right or not and why? $689 \times 33 = 22{,}737$[a]	69.0	27.8	3.2
2. 1,221 workers need to be transported to a construction location. How many buses do they need if each bus holds 31? A worker suggested 39.4 buses. Can his suggestion be right? Why?[b]	53.4	42.6	4.0
3. The P.E. teacher has a 210 KD budget to prepare for the school's annual party. He spent 4/5 of the money; he asked the management team to find out how much money he spent. The managing team did some calculations, and then decided that 262 KD was spent. Can their suggestion be right? Why?[c]	34.1	41.7	24.2
4. Bedoor has 3 KD. She got a notebook for 1 3/4 KD. Now she wants to buy a story for 1 3/4 KD. Do you think she has enough money? Why?[a]	13.9	32.3	53.8
5. Farah has 42 KD. She wants to buy 21 jars of juice. Each jar of juice costs 1.950 KD. Do you think she has enough money? Why?[a]	35.4	41.7	22.9
6. Choose from the numbers between the parentheses a number that makes the statement right for you. Then explain why you chose this number. $\left\{3\frac{3}{4},\ 1.54,\ 1\frac{1}{5}, \frac{3}{5},\ 0.9,\ 0.05,\ 2.5\right\}\ 16.48 \div \_\_ < 8$[a]	47.0	39.5	13.5

[a]Item can be judge using first criterion
[b]Item can be judge using second criterion
[c]Item can be judge using both criteria

In designing the RAT, the researcher utilized suggestions from the literature about how to help students judge the reasonableness of answers. The items were presented in contexts in which students need to think about the reasonableness of the given answer (Johnson, 1979; Reys, 1985) such as homework problems, answers for a mathematics test, and checking bills. These items provided a context that allowed students to reflect on both criteria for judging the reasonableness of answers. The students were given a limited time to judge the answer (less than a minute to judge and write their thinking of the reasonableness of the answer for each question), so that they would reflect on the reasonableness of the answer rather than attempt to calculate the exact answer. Additionally, students were asked to explain their strategies for judging the reasonableness of the answer (Wickett, 1997). It should be noted that asking students to explain their answers is not something that is generally observed in Kuwaiti classrooms, so this expectation would likely be new and challenging to these eighth grade students.

Reasonable Answer Interview

The RAI was designed to examine strategies that students used in determining the reasonableness of answers. The researcher conducted all interviews on a one-on-one basis. The interview took about 35 minutes or one classroom period. During the interview, each student was presented with 13 items that were selected from the RAT. Six were presented as mathematical statements and seven were presented in context. At the beginning of the interview, the researcher explained the purpose of the interview to the students and then asked them to describe their thinking out loud about the reasonability of the presented answer or the answer they chose. Students were also asked if they had another way to judge the reasonableness of the answer. The researcher asked follow-up questions to clarify and provide more detail about their thinking (e.g., Can you explain what you mean by …?). The researcher continued to ask follow-up questions until the students provided an explanation that the researcher understood or they simply stopped providing any productive answers. For example, when students said “the answer is reasonable,” the researcher asked them to provide more detail to explain why they thought the answer was reasonable. If students continued to make statements that did not clarify or provide insight into their thinking, such as “I think it is reasonable” or “it seems reasonable to me,” the researcher stopped asking any follow-up questions and moved to another question.

The students were not allowed to use calculators or to work the problem using paper-and-pencil procedures (for further details about the interview protocol, see Alajmi, 2004). All interviews were conducted by the author in Arabic, and all translations from Arabic to English were done by the author. All interviews were audio taped and later transcribed.

Data Analysis

Both quantitative and qualitative methods were used to analyze the data. The RAT was scored according to the following scoring system: Each item was assigned a maximum of two points. If both a correct response and a clear, correct explanation were provided, two points were awarded. An explanation was considered adequate if it reflected the use of at least one of the two criteria to judge reasonable answers. If a correct answer was provided but the explanation was unclear or omitted, the item was assigned one point. If the response was incorrect and the explanation was unclear or missing, the item was assigned zero points. After scoring the RAT, the mean and standard deviation were calculated for the eighth grade students' performance on the test. An item analysis was done to describe students' performance on each item.

The analysis of the transcripts of the RAI provided information to answer the second research question. Content analysis of the transcripts focused on coding the strategies used by the students in determining the reasonableness of their answers. In particular, the researcher was attentive to the following strategies predicted by the literature:

1. Using real-world benchmarks, e.g., buses cannot carry fractions of passengers.
2. Connecting the answer with real-world data: this cannot be the price for six movie tickets.
3. Using numerical benchmarks: 9/11 is a little bit less than 1 and 8/9 is a little bit less than 1, so the sum is less than 2.
4. Monitoring the effect of operations: this number is less than 1 so the product should be less (ex. $94\times2/3<94$).
5. Using the meaning of operations and number magnitude: there will be four "0.8s" in the number "3.22."
6. Estimating the expected answer:

    (a) Rounding: the answer will be around 21,000 ($700\times30=21{,}000$).
    (b) Clustering: all these number are around 29,000, so the sum will be around $29{,}000\times7$.

(c) Front-end: for $3{,}684 \div 7$, estimate $36 \div 7 = 5$ and then the answer will be around 520.
(d) Using compatible numbers: 0.38×8, it would be around $0.4 \times 8 = 3.2$.

The researcher was also open to other strategies that could emerge from the data. Through the analysis, the researcher developed a coding system for students' strategies. The reliability of the coding was insured by a series of reviews. Five mathematics educators were asked to apply the coding system. They were provided with examples of students' responses for each interview item (two to three examples for each item) and asked to independently code their responses. For almost 90% of the examples, there was 100% agreement on the coding of the strategies that students used.

Results

Students' Performance on the RAT

The RAT had 50 points possible. The mean score was 15.2 with a standard deviation of 4.8. On average, these eighth graders scored less than one third of the points possible on the RAT. Overall, these results suggest that Kuwaiti eighth grade students were unable to recognize reasonable answers.

On ten of the items on the RAT, less than half of the students recognized a reasonable answer. Additionally, the students' ability to explain their answers was extremely low. In all of the items, more than half of the students provided general statements, such as "it is reasonable" and "it is the right answer," but did not explain their thinking. In addition, some students provided incomplete statements that did not clarify their reasoning. Overall, the eighth graders provided few adequate explanations. There was only one item (no. 4; Table 1) on which more than half of the students were able to provide adequate explanations for their thinking. About one fourth of the students provided adequate explanations on the two other items (see item nos. 3 and 5; Table 1).

Students' Performance by Number Domain

The RAT included items that represent three number domains: whole numbers, decimals, and fractions. Most items on the RAT focused on decimals and fractions because these numbers are emphasized in the Kuwaiti middle school mathematics curriculum. Table 2 shows the mean percent of correct responses and the standard deviation for each number domain. An analysis of variance (ANOVA) showed no statistically significant differences ($p>0.05$) among the mean percent of correct responses on the RAT by the three number domains. Eighth grade students' performance in recognizing reasonable answers was consistently low in the three number domains as they showed consistent difficulty in recognizing reasonable answers for problems involving whole numbers, fractions, and decimals.

The students were challenged even on items that focused on whole numbers. For example, item no. 1 involved a computation involving multiplying whole numbers (see Table 1). The answer provided in item no. 1 ($689\times33=22{,}737$) was correct, but almost 70% of the students judged it as unreasonable. Of the students determining the answer to be reasonable, only seven (about 3% of the total sample) students provided an acceptable explanation for their judgment. These students rounded both numbers and then multiplied. For example, one wrote "700×30 is 21,000" and concluded "then the answer seems right."

Slightly more than half of the students who considered 22,737 an unreasonable answer provided an explanation for their answer. Their explanations reflected a misunderstanding of the magnitude of the answer. For example, many thought an answer of 22,737 was too large for the problem. They provided explanations that were supported by their personal intuition rather than thinking based on mathematical principles. For example, they made statements such as "the answer is thousands," "it should be smaller," or "this is a very large answer" to justify their conclusions. These students provided no evidence of estimation skills or any other processes to defend their decision.

These eighth grade students' were also challenged in recognizing reasonable answers for problems dealing with fractions and decimal

TABLE 2

Mean percent of correct responses and standard deviation for each number domain

Number domain	*Number of items*	*Mean percent of correct response*	*Standard deviation*
Whole number	4	27	0.09
Fraction	10	30	0.17
Decimal	11	32	0.13

Mean percent of correct response is the proportion of mean scores to the possible scores on the items on each number domain numbers. Item no. 6 (see Table 1) focused on dividing rational numbers and had more than one reasonable answer. In order to get a reasonable answer, students needed to recognize a basic idea of division: to get a small result, you need a large divisor. There were two acceptable answers to this problem because, in order to get a result smaller than eight, students could divide by any number larger than two.

About 40% of the students chose one of the reasonable answers. Out of these 89 students, only nine of them provided a clear and acceptable explanation for their answers. One student, who chose 3 3/4, said "16÷3<8." His explanation was based on rounding 6.48 and 3 3/4 to whole numbers.

Students' explanations for selecting unreasonable answers revealed several major misunderstandings. For example, one student argued that 3/5 was a reasonable answer because "3/5<8." Another wrote, "3/5 is the smallest number; it will give an answer less than 8." Other students based their choice on the similarity of the form of the numbers. They chose 1.54 as a reasonable answer because it has two decimal digits, the same as 16.48.

Students' Performance by the Two Criteria for Judging the Reasonableness of Answers

The RAT included items that could be judged using the first criterion (number relationships and the effect of operation) or the second criterion (practicality of the answer), or both criteria. Table 3 displays the mean percent of correct scores on items related to the two criteria used to judge the answer. An ANOVA revealed no statistically significant difference ($p>0.05$) in item difficulty among items based on the different criteria.

TABLE 3

Mean percent of correct responses for the three types of problems that used two criteria for judging answers for reasonableness

Criteria	*Number of items*	*Mean percent of correct response*
Items judged by number relationships and the effect of operations	18	28
Items judged by practicality of the answer	3	30
Items judged by number relationships and the effect of operations and practicality of the answer	4	43

Eighth grade students were equally challenged in examining answers for reasonableness using number relationships and the effect of operation as well as the practicality of the answer. The majority of students appeared to be unaware of the connection between mathematics in the classroom and real life. In items judged using the practicality of the answer, they did not realize the suitableness of some answers in a real-world context. For example, in item no. 2 (see Table 1), less than one half of the students considered 39.4 an unreasonable answer for the number of buses, and only 4% provided an adequate explanation for their thinking. The correct explanations students provided were based on recognizing that decimal numbers cannot represent the number of buses. For example, one reasoned, "there are no 0.4 bus," while another noted "a bus cannot carry decimal numbers." Thirty-five of the students who considered 39.4 an unreasonable answer provided unacceptable explanations for their choice; 60 students provided no explanation at all. Twenty students wrote statements that did not describe their thinking about the answer, such as "it is wrong," or "it is not reasonable."

Slightly more than half of the students considered 39.4 buses a reasonable answer (53.4%). Forty percent of these students provided an explanation for their judgment. None of these explanations described how the students arrived at their decision. They either stressed that 39.4 is a reasonable answer by saying "it is right" or "it is enough" or provided facts such as "each bus carries 31 workers" or "the number of workers is more than the number of buses." Four students provided a decimal answer for the problem: 31.2, 49.5, 166.4, and 378.5. These answers revealed the students' lack of understanding of the appropriate use of decimal numbers to represent the numbers of objects in a real-world setting.

Results of the Interview

The RAI was designed to reveal strategies students used to determine the reasonableness of answers. The RAI interview guide included 13 items from the RAT. The interview transcripts were coded according to the strategies that students used to judge the reasonableness of the answers. Strategies used by eighth grade students were classified into categories: effective strategies and ineffective strategies (see Table 4). A strategy was considered effective if it reflected a student use of one of the two criteria for judging the reasonableness of an answer: number relationships and the effect of operations and the practicality of the data such as estimation and numerical benchmark. A strategy was considered ineffective if it reflected

TABLE 4

The strategies that students used on the RAI

	Strategies	*Frequency of strategies observed*	*Percent*
Effective strategies			
Number relationships and the effect of operations	Estimation	44	24.7
	Meaning of operation	19	
	Numerical benchmark	4	
	Decomposition/recomposition	9	
	Mental computation	1	
		3	
Practicality of the answer	Real-world benchmark	33	10.7
Ineffective strategies			
Mathematical procedure	Wanting to do the paper-and-pencil calculation	36	38.6
	Wrong procedure	33	
	Counting rule for the number of digits in the answer	26	
	Form of the answer	9	
	Misunderstanding a concept (number magnitude, operation)	14	
	Did the paper-and-pencil procedure mentally	1	
Others strategies and responses	Guessing	30	26.0
	I do not know	45	
	Misinterpreting the problem	5	
Total		308	100

a procedural rule-based method for determining the reasonableness of the answer or employed incorrect mathematics concepts.

Effective Strategies. Slightly more than one third of the students' strategies in recognizing reasonable answers were effective. One fourth of the strategies were based on their understanding of number relationships and the effect of operations. The most common strategy was estimation, for example, on item no. 2 (see Table 5). Thirteen students rounded 1.950 KD to 2 KD and then 12 of them used multiplication. For example, one student said, "let's say 1.90 is 2, so 21 jars, each costing 2, will be 42."

Estimation was also used as a second strategy on item no. 3 (see Table 5). After the students placed the decimal in the wrong spot to get 29.1357 as an answer for the problem, the researcher asked the students to reflect on the answer they got (534.6×0.545=29.1357). Only one student reflected on his understanding of the numbers and operation involved in the problem and used estimation to judge the answer.

Researcher: Look at the answer now: 29.1357. Is it a reasonable answer?
Student: It is not reasonable.
Researcher: Why?
Student: Because this number is almost 500 and the other number almost 0.5. If we multiply them we get 250.
Researcher: Where would you put the decimal point?
Student: After 3 digits, 291.357.

Students also used a numerical benchmark as a strategy to judge answers for reasonableness. In this strategy, students used their understanding of the size of a number as an anchor to judge an answer for reasonableness. For example, on item no. 2 (see Table 5), four students identified the maximum value for an individual jar of juice. They divided the money they had by the number of jars they wanted. One said, "42÷21=2, so they can buy anything that costs less than 2 KD." The same method was used by five students on the second part of item no. 4, (Table 5); one of them said "16÷2=8, but when you divide by 2.5 the answer will be less than 8."

Another approach used by eighth grade students depended on utilizing meanings of operations. Students reflected on their understanding of what the operations mean to determine whether the answer is reasonable or not. For instance, on the first part of item no. 4, statement a (see Table 5), a student clarified choosing 0.05 as a reasonable answer "The answer is 0.05, because, if we look for how many 0.05 in 3 8/9, it will be more than 4."

TABLE 5

Students' strategies for judging reasonable answer on the RAI

Items	*Strategies used* — *Effective strategies*	*Strategies used* — *Ineffective strategies*
1. The P.E. teacher has a 210 KD budget to prepare for the school's annual party. He spent 4/5 of the money; he asked the management team to find out how much money he spent. The managing team did somecalculations, and then decided that 262 KD was spent. Can their suggestion be right? Why?	Real-world benchmark (19) Meaning of the operation (1)[a]	Guessing (1) Paper-and-pencil procedure (3) Inappropriate paper-and-pencil procedure (3)
2. Farah has 42 KD. She wants to buy 21 jars of juice. Each jar of juice costs 1.950 KD. Do you think she has enough money? Why?	Estimation (14) Numerical benchmark 42÷21=2, 1.95<2 (4)	Guessing (4) Do not know (2)
3. Dalal used a calculator to solve the following problem; she got the right numerical numbers. However, she forgot to place the decimal point in the results. Your task is to place the decimal point so the answer is right, and then explain you answer. 3.04×5.3=16,112	Estimation (1)[a]	Counting rules (24)
4. Choose from the numbers between the parentheses a number that makes the statement right for you. Then explain why you chose this number. $\{3\frac{3}{4}, 1.54, 1\frac{1}{5}, \frac{3}{5}, 0.9, 0.05, 2.5\}$. Statements: (a) $3\frac{8}{9} \div \_ > 4$;	Statement a Meaning of the operation (2)	The form of the answer (3) Paper-and-pencil procedure (3) Inappropriate paper-and-pencil 2procedure (4) Guessing (4) Do not know (8)
(b) $16.48 \div \_\_ < 8$	Statement b Estimation (3) Numerical benchmark (5)	The form of the answer (2) Paper-and-pencil procedure (2) Inappropriate paper-and-pencil procedure (4) Guessing (1) Do not know (7)

[a]Secondary strategy provided when prompted by researcher

Eleven percent of the strategies were based on their understanding of the practicality of the data. They judged the size of the answer by what they knew about data in the real world, namely, using benchmarks. Students used this strategy on three items that can be judged using practicality of the data. For example, on item no. 1 (see Table 5), students realized that 262 KD is more than the budget, so they used the budget as a benchmark to explain that the answer should be less than 210.

Ineffective Strategies. Table 4 reports that almost 65% of the total number of strategies was ineffective and almost 40% of these eighth grade students' strategies was procedurally driven. This finding is not surprising in an educational system that focuses on mathematics procedures. Eleven percent of the students' responses claim that they needed to work the paper-and-pencil calculation as the only way to judge the reasonableness of the answer. For example, one student said: "I need to work the problem to see whether this answer is reasonable or not." On item no. 4, statement a (see Table 5), a student said, "I need to try all these numbers. I need to solve them all, then see which one will be more than 4." Students seemed to trust the paper-and-pencil calculations for finding the *exact* answer as the only way to determine a *reasonable* answer.

It is also worth noting that nearly 12% of the procedures suggested by the students were inappropriate. Students' responses revealed that some of these eighth grade students were unaware of the effect of an operation or had no idea when a mathematical procedure was appropriate. For instance, on the first part of item no. 4 (see Table 5), students described two different and wrong mathematical procedures. One described dividing fractions $\left(3\ \frac{8}{9} \div 3\ \frac{3}{4}\right)$ by using a part-by-part procedure. He said, "3÷3=1, 8÷4=2, 9÷3=3. Answer 1 3/2 is bigger than 4." This student did not realize that the answer he got was actually less than 4. Similarly, on item no. 4, statement b (see Table 5), students suggested dividing decimals with a part-by-part procedure.

Another type of inappropriate procedure was converting a division to an addition problem. Two students on item no. 4, statement a (see Table 5) described a process in which they changed the division to addition and then added both numerators and denominators. One said, "1 1/5 is the answer, because when we change the operation into addition, we get 4 9/14>4."

Another ineffective strategy focused on the form of the number. Students judged or chose the reasonable answer based on how the numbers were presented in the problem, such as both are decimals or fractions. On item no. 4, statement b (see Table 5), a student chose the answer 1.54 to be an answer that makes $16.48 \div \_\_ < 8$ true because "I chose 1.54 so the decimal will be the same, then when we divide, the answer will be less." Some students chose specific forms for the answer that they thought was helpful in working the problem, such as a similar denominator or numerator. For example, on item no. 4, statement a (see Table 5), students chose 0.9 to be a reasonable answer. One student said "0. 9=9/10 so we can make cancelation when we divide." Another reason provided by two students was that "we need to have two numbers of the same denominator."

Students also relied heavily on rules, such as counting rules, to examine the reasonableness of an answer. For example, on item no. 3 (see Table 5), all students tended to count the number of digits to place the decimal on the answer. Only one student provided an explanation for placing the decimal, after a follow-up prompt from the interviewer. She reasoned, "Here in 0.545, the decimal digit is thousands and in 534.6 is ten, so $1{,}000 \times 10 = 10{,}000$ so the decimal point is after four digits." The student here used $10 \times 1{,}000 = 10{,}000$ as a short way for $1/10 \times 1/1{,}000 = 1/10{,}000$. This explanation clarifies the counting rule, but fails to consider the reasonable magnitude of the result of multiplying these two decimal numbers. The students' strategies in approaching this item are consistent with the findings of Hiebert & Wearne (1986).

Slightly more than a quarter of the students either based their responses on guessing or admitted that they did not know how to approach the problem (see Table 4). In guessing, students did not provide enough information that could help to classify their thinking. Table 4 shows about 15% of the student admitted they did not know how to examine answers for reasonableness.

Conclusions

This research has documented the level of Kuwaiti eighth grade students' performance in recognizing reasonable answers and the strategies they used to determine reasonableness. The results from over 200 eighth grade students showed that they were generally unable to recognize reasonable answers. They were even challenged in recognizing reasonable answers with problems that dealt with whole numbers, a computational topic that Kuwaiti students have been developing for 8 years.

These eighth grade students faced challenges in using the two criteria in judging the reasonableness of answers. Although evidence of using these criteria was found, it was not necessarily used often nor was the evidence used by a majority of the eighth graders. In fact, students' performance was consistently low on items that focused on number relationships and the effect of operations and on items that focused on the practicality of the answer. The challenges students faced in judging the reasonableness of answers using the practicality of data provides evidence that many students do not make connections between mathematics problems and real-world data. While an effort was made to provide familiar real-world contexts for these students, it is recognized that students' daily experiences vary greatly. The lack of alignment between real-world and the questions used in this research would impact their use of one of the criterion, namely, "practicality of the answer" and the extent to which this happened is a limitation of this research. Nevertheless, these findings support Hiebert's (1984) argument that students are unaware of the consistency between the results of mathematics problems and how things work outside their classrooms. The results are also consistent with Greeno (1991) who argued that students who do not see the conflict between their answers and data in the outside world usually focus on the arithmetic operation "without a sense of what the numbers were meant to be about" (p. 173). Kilpatrick et al., (2001) reported that this misconnection is due to classroom practices in which the focus is on applying algorithms without reflecting on the meaning of the answer.

While some students used effective strategies to determine the reasonableness of answers, these students were the exception rather than the rule as a majority of the students relied exclusively on algorithms that tended to be applied in a rote fashion. When effective strategies were used, they reflected students' understanding of the relationships of numbers, the effect of operations, and the practicality of the answer, such as estimation, numerical benchmarks, and real-world benchmarks. The most commonly used effective strategy for judging the reasonableness of an answer was estimation. On the RAI, 14% of students' strategies involved estimation. The estimation strategies they used varied, as did their ability to effectively use a particular strategy. Although little attention is given to estimation in the Kuwaiti national mathematics curriculum, this research confirms that estimation is a skill that a few students develop on their own. This finding is in line with previous research findings on students' use of estimation (Reys et al., 1980; Rubenstein, 1985). It also provides support for the argument that estimation should be taught as a means of improving students' ability to recognize reasonable answers (Hiebert, 1984; Onslow et al., 2005; Reys, 1985; Trafton, 1994). Bonotto (2005) found that introducing estimation in the fourth grade "allowed students to discuss their own work whenever the procedure used gave results that were incompatible with the predictions previously made. This method fostered a connection between solutions and their reasonableness" (p. 338).

Forty percent of all the strategies used to examine answers were procedurally driven; the most common response was carrying out exact calculations. This result confirms Byers & Erlwanger's (1984) findings that students use procedural rules to try to verify their solutions because they believed these rules *must* be applied to get the problem right. It also aligns with Yang's (2005) finding that Taiwan students tend to use paper-and-pencil procedure to explain and verify their thinking. Additionally, Kuwaiti students confused the procedural rules they learned at school, such as applying a rule for addition and subtraction to a multiplication problem. These students did not have a clear understanding of the effect of these operations nor when to apply certain rules. Resnick (1987) called these mixed procedures "buggy" algorithms. She blamed students' use of "buggy" algorithms on the school learning environment where students use mathematics symbols without fully understanding the meaning of the symbols. Galen & Gravemeijer (2003) argued that misapplying rules is related to learning these arithmetic rules as isolated facts and procedures.

Students were challenged in explaining their mathematics thinking. Their responses in the RAT and RAI revealed that students face difficulty in articulating their mathematical reasoning. On the RAT, most students wrote simple statements such as "the answer is reasonable." During the RAI, some students provided fragments of their thinking. Mathematics teachers in Kuwait explained that students do not often have opportunities to reflect on their mathematical thinking. Teachers rarely ask students for written explanations of their thinking. Teachers acknowledged that students cannot express themselves very well. Additionally, they have difficulty with language. Consequently, most of the explanations that students provided were in an oral form and focused on explaining standard procedures. In general, they reflected low-level thinking (Alajmi & Reys, 2007).

Students' explanations of their reasoning on the RAT and in the RAI indicated that they were unaware of the magnitude and relationships of numbers and the effect of the basic operations. This lack of understanding presented obstacles for them to recognize or determine reasonable answers. This finding provides further evidence that focusing on standard computational algorithms does not necessarily help students develop an understanding of numbers and operations, and that is consistent with earlier research in this area (Reys & Yang, 1998; Yang & Huang, 2004).

If improvement in students' interpretation and recognition of reasonable answers is to occur in Kuwait, then major changes need to be made in mathematics programs. This means that the mathematics curriculum must be adjusted to provide opportunities for students to experience situations that encourage students to reflect on results and assess the reasonableness of answers. It means that instruction must also change. Teachers must help students value the importance of determining reasonable answers and foster effective strategies to decide if answers are reasonable. Additionally, teachers must dispel the notion popular among Kuwaiti eighth graders that finding the exact answer was the only way to find a reasonable answer. There is a need to shift the emphasize in the Kuwaiti mathematics programs to focus more on developing students understanding of the numbers and operation and their relationship and providing students with the opportunity to think of alternative methods to determine reasonable answers.

This research addressed the notion of reasonableness and provided, for the first time, a detailed profile of Kuwaiti student performance. Hopefully, this study will be informative to the international community of mathematics education and also help stimulate some changes in future mathematics curriculum and teaching in Kuwaiti schools.

References

Alajmi, A. (2004). Eight grade Kuwaiti students' performance in recognizing reasonable answer and strategies they used to determine reasonable answer. Unpublished doctoral dissertation, University of Missouri–Columbia.

Alajmi, A. (2009). Addressing computational estimation in the Kuwaiti curriculum: Teachers' views. *Journal of Mathematics Teacher Education*, (in press)

Alajmi, A., & Reys, R. (2007). Reasonable and reasonableness of answers: Kuwaiti middle school teachers' perspectives. *Educational Studies in Mathematics, 65*, 77–94.

Beaton, A., Mullis, I., Martin, M., Gonzalez, E., Kelly, D., & Smith, T. (1996). *Mathematics achievement in the middle school years: IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA: TIMSS International Study Center, Boston College.

Bell, A., Swan, M., & Taylor, G. (1981). Choice of operation in verbal problems with decimal numbers. *Educational Studies in Mathematics, 12*, 399–420.

Bonotto, C. (2005). How informal out-of-school mathematics can help students make sense of formal In-school mathematics: The case of multiplying by decimal numbers. *Mathematical Thinking & Learning: An International Journal, 7*, 313–344.

Byers, V., & Erlwanger, S. (1984). Content and form in mathematics. *Educational Studies in Mathematics, 15*, 259–275.

Gagne, R. M. (1983). Some issues in the psychology of mathematics instruction. *Journal for Research in Mathematics Education, 14*, 7–18.

Galen, F., & Gravemeijer, K. (2003). Facts and algorithms as products of students' own mathematical activity. In D. Schifter, J. Kilpatrick & W. Martin (Eds.), *A research* *companion to principles and standards for school mathematics* (pp. 114–122). Reston, VA: National Council of Teachers of Mathematics.

Greeno, J. G. (1991). Number sense as situated knowing in a conceptual domain. *Journal for Research in Mathematics Education, 22*, 170–218.

Hiebert, J. (1984). Children's mathematics learning: The struggle to link form and understanding. *Elementary School Journal, 84*, 497–513.

Hiebert, J., & Wearne, D. (1986). Procedures over concepts: The acquisition of decimal number knowledge. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics* (pp. 199–223). Hillsdale, NJ: Erlbaum.

Hope, J. (1989). Promoting number sense in school. *Arithmetic Teacher, 36*, 12–16.

Johnson, D. C. (1979). Teaching estimation and reasonableness of results. *Arithmetic Teacher, 27*, 34–35.

Kilpatrick, J., Swafford, J., & Findell, B. (eds). (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: National Academies.

McIntosh, A., Reys, R., & Reys, B. (1992). A proposed framework for examining basic number sense. *For the Learning of Mathematics, 12*, 2–8.

McIntosh, A. & Sparrow, L. (Eds.) (2004). *Beyond written computation*. Perth, Western Australia: Mathematics, Science & Technology Education Centre (MASTEC).

Ministry of Education. (2006a). *The development of education in Kuwait: Historical perspective*. Kuwait: Dar Albalak.

Ministry of Education. (2006b). *Educational statistical group 2005–2006*. Kuwait: Films.

National Committee on Mathematical Requirements. (1923). *The reorganization of mathematics in secondary education: A report of the National Committee on Mathematical Requirements under the auspices of the Mathematical Association of America*. Oberlin, OH: The MAA.

National Council of Supervisions of Mathematics. (1989). Essential mathematics for the twenty-first century: The position of National Council of Supervisors of Mathematics. *Arithmetic Teacher, 37*, 44–46.

National Council of Teachers of Mathematics. (1980). *Agenda for action: Recommendations for school mathematics of the 1980's*. Reston, VA: National Council of Teachers of Mathematics.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation for school mathematics*. Reston, VA: National Council of Teachers of Mathematics.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics.

Onslow, B., Adams, L., Edmunds, G., Waters, J., Eady, J., Chapple, N., et al. (2005). Are you in the zone? *Teaching Children Mathematics, 11*, 458–463.

Resnick, L. (1987). Constructing knowledge in school. In L. S. Lynn (Ed.), *Development and learning: conflict or congruence?* (pp. 19–50). Hillsdale, NJ: Erlbaum.

Reys, R. (1985). Estimation. *Arithmetic Teacher, 32*, 37–41.

Reys, R., Bestgen, B., Rybolt, J., & Wyatt, W. (1980). *Identification and characterization of computational estimation processes used by in school pupils and out-of-school adults*. Columbia, MO: Missouri University.

Reys, R., & Nohda, N. (1994). *Computational alternatives for the twenty-first century: Cross-cultural perspectives from Japan and the United States*. Reston, VA: National Council of Teachers of Mathematics.

Reys, E., & Yang, D. (1998). Relationship between computational performance and number sense among sixth- and eighth-grade students in Taiwan. *Journal for Research in Mathematics Education, 29*, 225–37.

Rubenstein, R. (1985). Computational estimation and related mathematical skills. *Journal for Research in Mathematics Education, 16*, 106–19.

Sowder, J. (1992). Estimation and number sense. In D. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 371–389). New York: Macmillan.

Trafton, P. (1994). Computational estimation: Curriculum and development efforts and instructional issues. In R. Reys & N. Nohda (Eds.), *Computational alternatives for the twenty-first century: Cross-cultural perspectives from Japan and the United States* (pp. 76–86). Reston, VA: National Council of Teachers of Mathematics.

Vance, J. H. (1986). Estimating decimal products: An instructional sequence. In H. L. Schoen & M. J. Zweng (Eds.), *Estimation and mental computation (1986 Yearbook of the National Council of Teachers of Mathematics)* (pp. 127–134). Reston, VA: The Council.

Verschaffel, L., Greer, B., & De Corte, E. (2007). Whole number concepts and operations. In F. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 557–628). Charlotte: Information Age.

Wickett, M. S. (1997). Serving up number sense and problem solving: Dinner at the panda palace. *Teaching Children Mathematics, 3*, 476–80.

Yang, D. (1995). Number sense performance and strategies possessed by grades 6 and 8 students in Taiwan. Unpublished doctoral dissertation, University of Missouri–Columbia.

Yang, D. C. (2005). Number sense strategies used by 6th-grade students in Taiwan. *Educational Studies, 31*, 317–333.

Yang, D. C., & Huang, F.-Y. (2004). Relationships among computational performance, pictorial representation, symbolic representation and number sense of sixth-grade students in Taiwan. *Educational Studies, 30*, 373–389.

Amal Hussain Alajmi

Curriculum and Instruction, College of Education
Kuwait University, Kuwait, Kuwait
E-mail: Alajmi.a@ku.edu.kw

Robert Reys

Mathematics Education, College of Education
University of Missouri, Columbia, MO, USA
E-mail: reysr@missouri.edu