

Asking Questions in First-Grade Mathematics Classes: Potential Influences on Mathematical Thought

Michelle Perry, Scott W. VanderStoep, and Shirley L. Yu

This study investigated the types of questions that are asked in 1st-grade addition and subtraction lessons in Japan, Taiwan, and the United States. Some researchers have argued that knowledge is, in part, constructed through questions and that these may be used differently in U.S. than in Asian classrooms. Thus, each question about addition or subtraction in 311 observed lessons was coded as 1 of 6 types of questions. Analyses revealed that the Asian teachers asked significantly more questions about conceptual knowledge and about problem-solving strategies than did U.S. teachers. In addition, Chinese teachers asked significantly more questions that were embedded in a concrete context than did U.S. teachers. These findings allow speculation that the kinds of questions typically asked in Japanese and Chinese classrooms may contribute to the construction of more sophisticated mathematical knowledge for the children in those classrooms.

The differences in mathematics achievement between children from the United States and children from Asia are well established in the educational and psychological literatures. Studies have consistently found U.S. children scoring below children from Asia on achievement tests (e.g., Crosswhite, Dossey, Swafford, McKnight, & Cooney, 1985; Husen, 1967; Robitaille & Garden, 1989; Stevenson et al., 1990; Stevenson, Lee, & Stigler, 1986). Although this is a well established fact, researchers and educators are only beginning to understand why these differences have emerged. The present study represents an attempt to examine potential influences on children's mathematical understanding that may lead to differences in performance between Asian and U.S. children.

The specific purpose of this study was to examine whether the degree to which teachers attempt to engage students in higher order thinking, through the kinds of questions they ask students, differs between Asian and U.S. cultures. We chose to look at the questions that teachers ask because evidence suggests that students who are asked to respond to questions that require higher order thinking per-

form better than students who do not have to answer such questions (and this holds true even when students do not actually answer the questions; see, for example, Pressley, Symons, McDaniel, Snyder, & Turnure, 1988). We investigated whether the amount of higher order thinking that is encouraged in students differs across the three countries by examining the questions that teachers ask.

Several lines of research led us to pursue this investigation. Specifically, the following three areas informed our research agenda: (a) past research that documented achievement differences between Asian and U.S. children, (b) differences in classrooms that appear to be related to achievement differences, and (c) other classroom practices that we would expect to be related to achievement differences. Each of these areas is briefly outlined in the following sections.

Cross-Cultural Achievement Research

Much of the research exploring differences between Japan, Taiwan, and the United States in children's mathematics achievement at the elementary school level has been conducted by Stevenson, Stigler and their colleagues (for a review, see Stevenson et al., 1986, 1990). Their first large-scale study (e.g., Stevenson et al., 1990) involved first- and fifth-grade classrooms from Japan (Sendai), Taiwan (Taipei), and the United States (Minneapolis, Minnesota). The researchers constructed an achievement test on the basis of their analyses of the curriculum. Their analyses of test scores showed that, in both first and fifth grades, the children from the United States scored significantly below children from Taiwan and Japan. Although the scores achieved by children from the United States showed a great degree of improvement between first and fifth grade, they still lagged far behind the scores of their Asian peers.

The second large-scale study undertaken by Stevenson and his colleagues (e.g., Stigler, Lee, & Stevenson, 1990) involved first- and fifth-grade classrooms from Japan (Sendai), Taiwan (Taipei), and the United States (Cook County, Illinois, including Chicago). Stigler et al. (1990) and Stigler and Perry (1988b) described two new tests that were devised to measure (a) computational skills and

Michelle Perry, Departments of Educational Psychology and Speech Communication, University of Illinois; Scott W. VanderStoep and Shirley L. Yu, Combined Program in Education and Psychology, University of Michigan.

Portions of this report were presented in July 1991 at the biennial meeting of the International Society for Behavioural Development, Minneapolis, Minnesota.

Support for conducting some of the analyses presented in this study was provided by the University of Michigan Center for Research in Learning and Schooling. The data on which this report is based came from a study that was funded by the National Science Foundation (Grant BNS8409372) to Harold Stevenson, whom we thank for his generosity in sharing these data for our analyses and for his insightful comments on a draft of this article. We also thank Terri Graham, George Kamberelis, James Stigler, and an anonymous reviewer for their help in clarifying some of the ideas presented here in this article.

Correspondence concerning this article should be addressed to Michelle Perry, Department of Educational Psychology, University of Illinois, 1310 South Sixth Street, Champaign, Illinois 61820.

(b) broader mathematical knowledge. Comparisons conducted with these tests (Stigler et al., 1990) showed that children in Taiwan and Japan consistently outperformed the U.S. children in computational skills. In this second study, the first graders in the Asian countries scored better than the U.S. first graders, and by the fifth grade, the difference between the Asian and U.S. children was even greater.

Stigler et al. (1990) also gave individually administered tests of mathematics knowledge to the children. These tests included problems typically found in the mathematics curricula of the three countries, as well as problems testing mathematical knowledge, more broadly conceived. Children at both grade levels were tested on word problems, concepts and equations, estimation, operations, graphing, visualization, mental folding, and mental calculation. An oral test of reasoning was given only to the first graders, and a geometry test was given only to the fifth graders. As in previous comparisons, Stigler et al. (1990) noted that, for all problem types, Japanese children scored the highest and U.S. children scored lowest, with Chinese children generally falling between the Japanese and U.S. children. In addition to replicating the results of the first study, Stigler et al. (1990) found the same pattern of differences across a wide range of mathematical skills, beyond computational skills.

Differences in Classroom Structure

Several attempts have been made to examine the variables that may account for the achievement differences described above. For example, Stigler, Lee, Lucker, and Stevenson (1982) analyzed the curriculum in the three countries (also see Stevenson & Bartsch, 1992; Stigler, Fuson, Ham, & Kim, 1986). By looking at the national textbook series used in Japan and Taiwan and a popular textbook series used in the United States, Stigler, Lee, and Stevenson (1987) compiled 320 topics covered in these books by at least one of the countries in grades one through six. They found that significantly more topics were covered by the Japanese textbooks (91% of the 320 topics) than by the Chinese or U.S. books (80% and 78%, respectively). They also found that, of the 320 concepts covered in at least one of the countries before the middle of the first grade, Japanese and U.S. texts covered an equal number of those topics, whereas Chinese texts covered fewer. However, by fifth grade, Japanese texts had covered 86% of the topics covered by at least one country, whereas U.S. texts had covered only 66%. The authors concluded that, in terms of curriculum, the U.S. keeps pace with Japan until some point beyond the first grade but falls behind by the fifth grade.

As another example of classroom differences, Stigler et al. (1987) found that children worked alone more than half of the time in U.S. classrooms; in contrast, Japanese and Chinese children spent most of their time paying attention to the teacher. Stigler et al. found that Asian classes were more tightly organized, with a high frequency of whole-class responses, whereas the U.S. classrooms were loosely struc-

tured with a lot of irrelevant activity taking place and children often being out of their seats.

Asian children also have more opportunities to receive instruction than their U.S. peers because Asian children spend more time in school (i.e., a longer school day and more days in school) than their U.S. peers. Although it is clear that spending more time in school and receiving more instruction would at least partially account for any cross-national differences, in the present study, we sought to explore whether another factor—teacher questions—could also help to explain these differences.

Differences in Classroom Processes

Classroom process differences have also been examined cross-culturally. For example, Stigler and Perry (1988b) analyzed classroom observations conducted in Sendai, Japan; Taipei, Taiwan; and Chicago, Illinois. They found that classrooms in the three countries differed in many ways, such as their approach to problem solving and the ways in which problem solutions are evaluated. As an example, in both first and fifth grades, more segments of the mathematics lessons were devoted to student evaluation in the Asian countries than in the United States. Furthermore, more of this evaluation was public and visible to the entire class in Japan and Taiwan than was the case in the United States. The significance of this is that public evaluation allowed all children to see how one might approach a particular problem, and thus could serve as a learning experience, whereas private evaluation typically allowed only for the individual student to know whether he or she got the right or wrong answer (also see Stevenson & Stigler, 1992).

Both Graham (1991) and Perry (1989, 1992) have investigated the types of mathematical explanations to which first- and fifth-grade children are exposed in Japan, Taiwan, and the United States. As an example of the differences that have been reported, Perry (1989) found that first-grade children in Taiwan learned about addition and subtraction most often through the processes of decomposition and recomposition of addends to 10 (e.g., when adding 5 to 8, decompose the 5 into 2 + 3, add the 2 to the 8 to make 10, then just add the 3, to make 13). The knowledge used to solve problems in this way can be generalized across addition and subtraction problems. In contrast, one of the frequent ways in which first-grade children in the United States learned about addition and subtraction was through a “families” metaphor (e.g., 8, 5, and 13 make a family). However, this metaphor easily could be—and actually was in one classroom—taken to the extreme to include “a mommy, a daddy, and a baby,” which could lead to confusion about the operations of addition and subtraction.

It seems clear that, at the very least, the environments in which the teaching occurs and the practices that occur in the classrooms are quite different across the three sites. We now take up the issue of where else to look to pinpoint differences in classroom practices that actually lead to cognitive differences in the children.

Classroom Practices That Lead to Cognitive Differences

One of the most intriguing findings about classroom practices that could lead to positive student cognitive outcomes has centered around the effects of asking higher order cognitive questions (Bloom, 1956; Redfield & Rousseau, 1981; Samson, Strykowski, Weinstein, & Walberg, 1987; Winne, 1979). Theoretically, the reason that asking more conceptually challenging questions leads to better student achievement is that these questions engage children in integrative thought, which leads to better learning than either answering questions that deserve rote responses or passively taking in material (e.g., Martin & Pressley, 1991).

The most powerful evidence in support of this expectation comes from research that has documented the powerful effects of verbal elaboration on learning (e.g., Brown & Kane, 1988; Slamecka & Graf, 1978). As an example of this research, Brown and Kane (1988) showed that preschool children who either formulate an idea on their own or who are coached to verbalize the idea that an experimenter described to them outperform children who merely hear the idea from the experimenter and do not participate in verbalization.

In several related studies, Pressley and colleagues (e.g., Martin & Pressley, 1991; Pressley et al., 1988) found that merely being asked a "why" question promoted learning, even when the learners did not succeed in generating answers. This suggests that question asking may, in and of itself, lead subjects to elaborate mentally, which eventually leads to the formation of new knowledge.

Some researchers have suggested that higher level learning is epiphenomenal of the intent of effective verbalization. Benware and Deci (1984), for example, told one group of students to learn course material from a psychology class with the purpose of being tested on the content. They told another group to learn the material with the intent of teaching it to other students. They found no differences between the two groups on recall of low-level information, such as the recall of facts, but the group that learned the material in anticipation of teaching it to others did better on high-level, integrative questions. Thus, it appears that preparing to communicate one's knowledge to others may activate high-level cognitive processes. We explored this assumption in this project.

Much of the research that suggests that elaboration is the key to significant learning has been done outside of the classroom. When attempting to investigate elaboration in the classroom, it is critical to investigate the questions that teachers pose, as students' elaborations are likely to be dependent on the questions addressed to them.

In general, our goal in the present study was to take advantage of some of the achievement differences that exist cross-culturally and to examine whether the classroom practices are related to the known patterns of achievement. In particular, we examined whether asking higher order cognitive questions, as an important classroom practice that leads to elaboration, is related to superior patterns of achievement. Given that Asian students, particularly Japa-

nese, have higher achievement scores than U.S. students, we predicted that Asian teachers would be found to ask questions requiring high-level thinking in a greater proportion of their lessons than U.S. teachers.

Method

Subjects

The data on which this report is based came from a study that was conducted in three locations: Japan (Sendai), Taiwan (Taipei) and the United States (Cook County, Illinois, including Chicago). Ten schools participated in each of the Asian cities and 20 schools participated in the United States. A larger number of U.S. schools was used because classrooms in the United States were more diverse ethnically and linguistically, and this diversity could be captured by including a larger number of schools. A full description of the demographic characteristics of these schools is provided in Stigler et al. (1990).

In each of the schools, two first-grade classrooms (and teachers) were observed. Four 40-min lessons were observed in each classroom. Thus, 80 lessons were observed in Japan (i.e., 10 schools \times 2 teachers \times 4 lessons for each teacher), 79 in Taiwan, and 152 in the United States (as 20 schools were used),¹ for a total of 311 lessons in the three countries.

Procedure

Lessons were recorded by trained observers who were native to one of the three countries. There were 16 observers in Japan, 10 in Taiwan, and 4 in the United States. Each observer took records in several schools in one of the three countries. These observers were instructed to write down as much as possible during the entire mathematics lesson. Some of the observers from each of the three sites had been brought together for initial training so that agreement about what was to be observed was reached before the actual observations took place; this information was then used to train the remainder of the observers in the three countries, who initially were not gathered together. The observers were trained to record the ongoing flow of behaviors and to include descriptions of the lesson, including the materials used and the specific nature of the content. The observers were instructed to pay special attention to the verbal remarks made by teachers and students and to include as many of these remarks as possible in their notes. None of the observers was aware of any of the hypotheses of this study.

The analyses reported here are based on the summaries produced from the classroom transcripts. As Stigler and Perry (1988b) reported, the decision to rely on summaries of the observations was made to deal with inconsistencies in the amount of detail the observers recorded. The summaries were constructed both to carry the flow of lesson and to recapitulate in detail what took place during the lesson. Standardized procedures were outlined so that certain things would always be included in the summary if they occurred during the lesson. To achieve reliability, all lessons were

¹ Teacher illness prohibited the collection of all 80 planned observations in Taiwan, and illness or premature birth of one teacher's baby prohibited the collection of all 160 planned observations in the United States.

coded by two coders, and any disagreement between the two coders was resolved by discussion among the entire group of coders ($n = 6$). Further detail about how the summaries were constructed was reported by Stigler and Perry (1988a, 1988b).

Coding

We developed a coding scheme to categorize all of the questions teachers asked their students. We created six categories of questions, with the goal of making the categories mutually exclusive. (The categories are described in Table 1 and in the following paragraphs.) After the coding scheme had been devised by the first two authors, all three of us coded 30 lessons together (10 randomly chosen from each country) to ensure that we would be applying the coding scheme similarly. Next, we randomly chose 45 lessons (15 from each country); each of these lessons was coded by two raters. Reliability (percent agreement) on these 45 lessons was 87%. The two raters who had the lowest overall reliability (85.5%) randomly chose an additional 12 lessons to code, and their reliability on this set of lessons was 90%. All disagreements on double-coded lessons were discussed and reconciled. The remaining 66 lessons were coded only once. All coding was performed with information identifying the country omitted to prevent any bias in our coding.

The coding scheme was applied to each of the questions that teachers asked about addition or subtraction in the lessons.² The main reason for selecting only these lessons was that the types of questions that can be asked change with the topic or the content of the question. For example, it is virtually impossible to ask a computing in context question about calendars (which was a frequently occurring topic in the Chinese mathematics lessons). Furthermore, the amount of coverage of addition and subtraction topics was fairly high and relatively equivalent across the observation sites (59% of the lessons in Japan, 73% of those in Taiwan, and 63% of those in the United States covered addition, subtraction, or both). This was not true for the other topics that were taught in the first grade. In fact, no other topic was covered in all three sites in the first grade (e.g., lessons on geometry were observed in Japan and the United States but not in Taiwan). If we had included all topics and all questions that were asked in the service of teaching those topics, we could have distorted our results such that some of the sites would have seemed not to have included a certain type of question with great frequency, when actually they had simply covered topics that did not allow for certain types of questions. To avoid this problem, we restricted all analyses to lessons covering addition, subtraction, or both, which were also the most frequently covered topics at each of the three observation sites.

Table 1 presents a summary of the coding scheme, along with examples of each question type. Descriptions of the six categories that were used to classify teachers' questions follow.

Computation or rote recall. The teacher asks students to compute a solution to an addition or a subtraction problem (e.g., "What is $7 + 5$?"). Also included in this category were place-value questions and yes-no or multiple-choice questions, such as, "Is this number in the ones column or the tens column?"

Rule recall. The teacher asks students to recall a particular rule they have been taught. For example, some teachers instructed children that the way to perform two-digit addition was to "add the ones with the ones and the tens with the tens." These teachers frequently would ask, "How do you do two-digit addition?" and the students almost always would respond "You add the ones with the ones and the tens with the tens." The question was coded as *rule recall* because of the high consistency with which the students provided the same answer.

Computing in context. This type of question is similar to computation because the required calculations could be the same. However, in this case the question had to involve either actual concrete manipulatives or a realistic context for the particular problem. As an example of a computing in context question, one teacher asked, "There are 30 pieces of red folding paper, 20 green. How many altogether?"

Make up a problem. The teacher asks the students to make up problems. This question requires students to generate a problem but not necessarily or merely to solve one. As a typical example, we found that teachers would often introduce a numeric equation and then ask the children to generate appropriate contexts from which that equation could be derived. For example, one Chinese teacher put " $8 - 3 - 2$ " on the blackboard and asked students to create problems that would lead to this problem. Students came up with some of the following problems: "I bring eight apples to school. I eat three and then eat two. How many have I left?"; "There are eight birds. Three fly away and then two more fly away. How many are left?"; and "There are eight fish. The cat eats three and the dog eats two. How many left?" The act of asking students to create these contextual problems was coded as student makes up a problem.

Problem-solving strategies. We coded questions as problem-solving strategies when teachers asked their students to go beyond actual computation and to explain how a problem was solved. These questions potentially enlist the student to think about the problem at a higher level than do the previously described categories. For example, a Japanese teacher first asked her students to stand up if they had gotten 46 for their answer and then asked them to explain how they arrived at this answer. We argue that verbalizing how one obtained an answer potentially requires a deeper understanding of addition and subtraction concepts than either computing the answer, reciting a rule, or generating a problem.

Conceptual knowledge. For the conceptual knowledge category, we attempted to capture questions that asked for the principles underlying addition and subtraction. A question was considered a conceptual knowledge question if it asked students to transcend the problem at hand and to attempt to construct generalizable, abstract, or conceptual knowledge (e.g., why a particular method does or does not apply to a particular problem or what each operation accomplishes).

Because it was possible for more than one question type to occur in a lesson, all lessons were coded for all six types of questions. For example, a teacher could ask a computing in context question and then later ask a problem-solving strategies question. If this occurred, then both the computing in context and problem-solving strategies types would be recorded as occurring. Thus, we obtained the proportion of lessons that contained each question type.

² We did not code every question asked in these lessons; we only coded those that we agreed were questions about addition, subtraction, or both of these operations. In general, two types of questions were ignored by our coding system: (a) questions that were nonmathematical, independent of whether they were central to the lesson (e.g., "Can you read the problem for us?") or incidental to the lesson (e.g., "Do you live in a red house?") and (b) questions that asked for agreement (e.g., "Do you agree or disagree?"). Questions such as these were considered to be beyond the bounds of our investigation.

Table 1
Examples of the Six Question Types

Question type	Example
Computation/rote recall	What is $7 + 5$?
Rule recall	What is the rule for two-digit addition?
Computing in context	There are 30 pieces of red folding paper, 20 green. How many altogether?
Make up a problem	Create a problem that would lead to the equation $8 - 3 - 2 = 3$
Problem-solving strategies	How did you arrive at this answer?
Conceptual knowledge	Why do you use subtraction for this problem?

Results

General Results

In this section we address whether the proportion of the types of questions that teachers asked in addition and subtraction lessons differed across the three countries. We tested for these differences with one-way analyses of variance on an arcsine transformation of the proportions of each teacher's addition and subtraction lessons in which certain question types appeared (Winer, 1971). Note that for these analyses we chose to examine all lessons that contained instruction about addition or subtraction. We reasoned that it would be less valid to include lessons that would not be likely to contain a question about addition or subtraction (e.g., lessons devoted to geometric shapes or the calendar) and also that it would be unfair to omit lessons about addition or subtraction that did not include questions (e.g., lessons in which the children spent most of their time hearing explanations). The problem in the first case (i.e., including lessons other than addition or subtraction) is that it would be unfair to claim that teachers do not ask high-level questions when in fact they may, but the questions may be about the particular topic they are teaching. The problem in the second case (i.e., including only those lessons with addition or subtraction questions) is that we found the omission of questions to be meaningful in the comparisons we chose to pursue in this study when the lessons were focused

on addition and subtraction and the teachers could have asked questions but did not. Table 2 shows the untransformed proportions for all six question types across all three countries.

As is clear from Table 2, no significant differences in the proportions of computation or rote recall questions, rule recall questions, or requests to make up problems were found across the three countries. However, substantial differences in the proportions of computing in context questions, problem-solving strategies questions, and conceptual knowledge questions were found. We present the remainder of the results in two sections. In the first section, we examine the question types that received comparable attention across the three countries and in the second section we examine the question types that received differential attention across the three countries.

Comparable Questioning Across Japan, Taiwan, and the United States

Computation or rote recall. The first-grade students from each of the three sites were engaged often in computation and rote recall questions, and the teachers in the three countries did not differ significantly in the proportion of lessons in which they asked these questions (61%, 46%, and 46% in Japan, Taiwan, and the United States, respectively).

We also examined the character of this question type in the three different countries and found them to be similar

Table 2
Mean Proportion of Teachers' Use of Different Types of Questions in Addition and Subtraction Lessons

Question type	Country			F (df = 2, 66)	MS _e
	Japan (n = 16)	Taiwan (n = 20)	U.S. (n = 33)		
Computation/rote recall	.61	.46	.46	0.75	1.065
Rule recall	.08	.01	.03	1.31	0.073
Computing in context	.58	.64	.31	3.99*	1.247
Make up a problem	.07	.14	.08	0.89	0.248
Problem-solving strategies	.63	.50	.24	7.66**	0.857
Conceptual knowledge	.37	.20	.02	11.34**	0.423

Note. F values were obtained by one-way analyses of variance of the arcsine-transformed proportions of each teacher's addition and subtraction lessons in which question types appeared. The proportions reported in this table are untransformed, whereas the F ratios and MS_es are based on the arcsine-transformed proportions.

* $p < .05$. ** $p < .001$.

at the three sites. For example, in Japan, one of the lessons involved the teacher asking students to answer problems such as $64 - 20 = 44$." In another lesson in Japan, a teacher said to the students, "So far we have thought of 50 as 5 tens. Then how do we view 85? How many tens and how many singles?" In the Chinese lessons, for example, one teacher wrote four problems on the blackboard, such as $3 + 6$ and $5 + 5$, and asked the students to answer the problems orally, one at a time. The character of the computation and rote recall questions in the United States was similar to the character of these questions asked in the Asian classrooms. For example, in one U.S. lesson on subtraction facts, the teacher asked the students to figure out the subtraction problem " $124 - 121 = \underline{\quad}$."

Rule recall. As can be seen in Table 2, very few teachers asked the rule recall type of question (it was asked, on average, by teachers in only 2.6% of all addition and subtraction lessons, across all three countries). The Japanese students were asked more of these questions than either the Chinese or U.S. students. However, these differences were not significant.

In examining the questions that were coded as rule recall, we found little that distinguished the character of the rule recall questions among the three countries. For example, the Japanese teachers would often ask their students questions such as, "What is the rule for two-digit addition?" The students would then answer, "Add the ones with the ones and the tens with tens." An example of a rule recall question asked in Taiwan involved the teacher asking, "What should we do with taking away?" The students answered in choral response, "Subtraction." Similarly, an example of a rule recall question in the United States involved the teacher writing " $45 + 33 = 78$ " on the blackboard and asking, "It is very important that we start in which house?" As is clear from these examples, there was consistency in the rule recall questions across the three countries.

Make up a problem. Very few teachers asked students to make up problems. No significant differences were found in the proportion of lessons in which students were asked to make up problems. Moreover, the character of the questions of this sort was similar in the three countries.

General observations about comparable questions. It seems that the students in the three countries of concern were exposed to comparable proportions of questions that make up the "basics" of addition and subtraction lessons: namely, questions about computation or rule recall. These students also were asked infrequently to recall rules for addition and subtraction or to make up addition and subtraction problems. In general, it appears that teachers, across the three countries, emphasized computation and rote recall questions (appearing in at least 50% of all addition and subtraction lessons in each of the countries) and did not devote much effort to getting students to recall rules or to generate problems.

Differences in Questioning Across Japan, Taiwan, and the United States

Computing in context. As Table 2 indicates, the Asian students were engaged by their teachers in a greater proportion of lessons containing computing in context questions than were the U.S. students. The only significant difference indicated by post hoc Fisher least significant difference (LSD) comparisons occurred between Chinese and U.S. teachers ($p < .05$). These comparisons are displayed in Figure 1.

We found that the character of the computing in context questions that were asked in the three countries was not necessarily equivalent. Although all of the questions coded as computing in context were reliably coded as this type of question, our impressions were that the success of the representation fostered or the contexts to which the mathematical problem was to be applied seemed to vary across the three countries. We provide several examples to substantiate the apparent differences we sensed across the countries.

The computing in context questions presented to Japanese students typically included contexts and materials with which they were familiar and that made sense in terms of their world knowledge. For example, many of the computing in context questions were about the cost of familiar items. Here is an example from the observations. One teacher asked her students: "You had 100 yen but then you bought a notebook for 70 yen. How much money do you still have?" In addition, the Japanese teachers seemed to stress that computing in context questions must first be understood before they can be solved. Here is how one Japanese teacher introduced a problem about which she would eventually require actual computation:

First the teacher read the following problem: "There are 45 apples on the apple tree. Picking 25 of those, how many will be left on the tree?" The teacher then asked her students to copy the problem (i.e., the words that make up the problem,

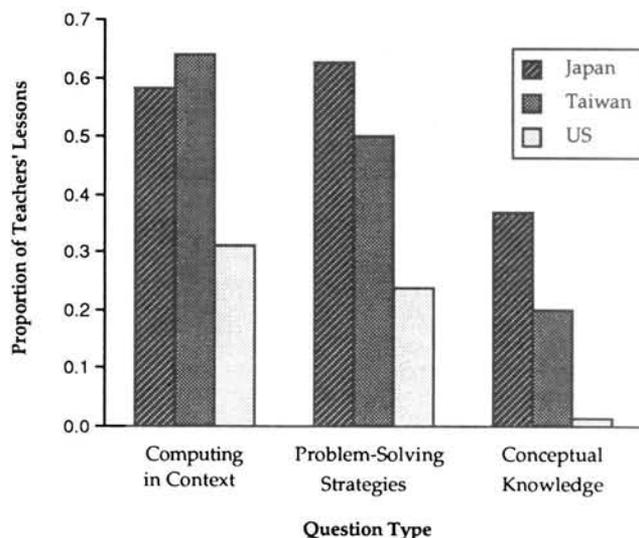


Figure 1. Profile of question types with significant differences across countries. (See Table 2 for statistical results.)

not the equation) into their notebooks and to silently read the problem that they had written. The teacher then asked the students to put down their pencils, look up at the blackboard, and read the problem silently once again while the teacher read it aloud. The teacher then asked two of the students to read the problem aloud.

In this example, the teacher allowed her students to understand what the problem meant; she made sure that the children tied the computation problems to things that were familiar or to things that had real-world applicability.

Many of the Chinese lessons seemed to be directed toward providing a concrete representation of the problem to be solved. Here is an example from the observations:

The teacher says, "Look at the first problem on page 15. Originally there are three kids playing ball. Two more are coming later and one more is also coming to join them. How many kids are playing now?" The teacher drew little people and wrote the equation down on the blackboard as she was speaking.

As is evident from this excerpt, when the Chinese children were given computing in context questions, teachers attempted to tie the computation problems to things that were familiar or to manipulatives that clearly represented the problem at hand. In many ways, the computing in context questions were used similarly in both Japan and Taiwan.

The picture appeared somewhat different when we scrutinized the computing in context questions asked by many of the U.S. teachers. In the U.S. examples, we encountered many questions in which the context appeared to be arbitrary rather than familiar. Here is the first example: The teacher read a problem out loud: "She walked into the store with 8 cents, earned 4 more, how much does she have now?" In this example, it seems as if the problem (earning more money and knowing how much money someone has) was an afterthought in an attempt to provide context to the symbolic problem " $8 + 4 = \underline{\quad}$." We feel that it may have been an afterthought because it is unlikely that a customer or even an employee will earn 4 cents by merely walking into a store (at least by legal means).

Here is a second example: "We have six yellow houses and two green houses. How many houses in all?" Here it is unclear what the significance of yellow versus green houses has to these children, why the children would concern themselves with adding together yellow and green houses, or why they would have reason to consider having multiple houses.

Here is a third example: "Mother bought ten hats. Aunt Nell bought three hats. How many more hats does Mother have than Aunt Nell?" In all of these cases, the context is more or less plausible but we believe not very meaningful to first graders.

We also found that, in many cases in the U.S. classrooms, more time and trouble was spent creating a context for the problem than on the meaning of the problem. Here is an example: "One apple, two more apples; how many in all?" The teacher then told the students to make a number sentence and to write the numbers clearly. We can imagine that with such a simple problem (in which most students would not even need to imagine apples to solve the problem), this

sort of elaborate representation is, at best, not necessary and, at worst, counterproductive by distracting children from the core of the problem.

We also noticed teacher confusion occurring in U.S. classrooms when attempts were made to use representations to motivate arithmetic understanding. For example, one lesson occurred as follows: "There were seven snowmen and three more came. No, three melted. So cross out three snowmen. How many are left?"

Despite the relatively impoverished or confusing examples we have mentioned thus far in our discussion of the U.S. lessons, we also found several examples of effective computing in context questions. Consider the following example:

The teacher had one row of students come to the front of the room and split them into two groups, one of four and one of seven. The teacher asked the students "Four plus seven equals what?" The teacher then asked a few more problems in the same manner, using the students in the front of the room as props.

In general, it appeared that there was much more variability in the ways in which U.S. teachers used computing in context questions as compared with their Asian counterparts. We found that these sorts of questions were used consistently to build meaningful representations for the children in the Asian countries and were used toward this end relatively less consistently in the United States.

Problem-solving strategies. As Table 2 indicates, the Japanese and Chinese students were engaged by their teachers in a significantly greater proportion of lessons containing problem-solving strategies questions than were the U.S. students by their teachers. Fisher LSD comparisons indicated that both Japanese and Chinese teachers asked this type of question in proportionately more lessons than did U.S. teachers ($ps < .01$ and $.05$, respectively). These comparisons are displayed in Figure 1.

A closer examination of the content of the problem-solving strategies questions revealed that teachers seemed to use this type of question similarly across the three countries. As an example, one Japanese teacher asked the students to give their answers to worksheet problems, which she wrote on the blackboard. Then, for the problem " $14 + 32 = \underline{\quad}$," the teacher had all students who got the correct solution (46) stand up and asked them to explain how they got that answer. In a similar vein, one teacher in Taiwan had the students read a problem out loud and then asked if it was an addition or a subtraction problem. Comparable examples were found in the U.S. data; for example, one teacher wrote " $5 + 5$ " on the overhead projector and asked the students to explain to the class how to use the number line to solve the problem.

Conceptual knowledge. Japanese students were engaged by their teachers in a significantly greater proportion of lessons containing conceptual knowledge questions than were Chinese students ($p < .05$, by the Fisher LSD procedure) or U.S. students ($p < .01$, by the Fisher LSD procedure), and Chinese students were engaged in a significantly greater proportion than were U.S. students

($p < .05$, by the Fisher LSD procedure). These comparisons are displayed in Figure 1.

The character of the conceptual knowledge questions was similar in Japan and Taiwan (and almost nonexistent in the United States). The conceptual knowledge questions asked by the Asian teachers most often dealt with evaluation of methods or determination of whether a certain word problem implied addition or subtraction. As an example, one Japanese teacher placed 5 rows of 10 tiles (a manipulative used in Japanese classrooms for counting) on the desk. The teacher then took away 3 rows of 10 tiles and asked a student how many were left. The student responded "20." The teacher then asked the class how they knew it was a subtraction problem. By asking students why the problem was a subtraction problem, the teacher was asking for conceptual knowledge about subtraction, in general. Effectively, students were asked to rely on conceptual knowledge about subtraction (and to go beyond the computations that the teacher had also asked the students to do).

The conceptual knowledge questions had a similar character in Chinese classrooms. For example, a Chinese teacher asked whether a problem out of the workbook was addition or subtraction. The class responded, "Addition," in unison. The teacher then asked why the problem was an addition problem.

General observations about question-asking differences. Asian students were exposed to more of the challenging kinds of questions (or questions requiring high-level thinking, that is, problem-solving strategies questions and conceptual knowledge questions) than were the U.S. students. Indeed, the Asian students were asked more questions that exploited existing complex mathematical knowledge or required the construction of such knowledge (as opposed to arithmetic knowledge, which may be exploited by, for example, computation questions).

Although we noted more variability in the ways in which U.S. teachers used computing in context questions than among the Asian teachers, we argue that the Asian teachers used these questions consistently to build meaningful representations for the children in the Asian countries (and were used toward this end only occasionally in the United States). This sort of tool may be especially critical for forming mathematical knowledge in the early school years, given that most of the mathematical knowledge with which children begin school is based on concrete or contextualized referents.

Discussion

Much of the previous research exploring cross-cultural differences in mathematics achievement has focused on social and cultural differences (e.g., Stevenson et al., 1990). We see these variables as crucial in understanding these differences. However, it is also critical to explore the different types of activities that take place in mathematics classrooms if we are to understand cross-cultural achievement differences. The differences in the cognitive processes in which children are engaged may also help us understand

why Asian children have higher mathematics achievement than other children. The present study adds to this line of research by expanding on how classroom discourse can affect cognition.

The results of applying our classification scheme suggest that the types of questions that the children were expected to grapple with differed across the three countries. Most notably, the Asian teachers (especially the Japanese teachers) behaved as though they expected their students to be able to deal with complex conceptual questions, whereas the U.S. teachers did not. The Asian teachers acted as though they expected the children not only to engage in a dialogue about how a solution was achieved (i.e., problem-solving strategies questions) but also to compare solutions across problems and to explain differences across operations (i.e., conceptual knowledge questions). This contrasts sharply with the view of the U.S. child held by U.S. teachers that we infer from the types of questions U.S. teachers posed. What we learned from the comparisons in this study is not only that teachers act as though they hold different beliefs about the level of thought of which first graders are capable but also that first-grade children are capable of being engaged in conceptual and abstract mathematical thought. Through observation of U.S. classrooms alone, this inference could not have been formulated.

It may be argued that noting whether a question type occurred is limited in certain ways, and under other circumstances, it would be advisable to compute actual proportions of the various question types. Indeed, if this were possible, analyses of this sort would allow inferences about the amount of emphasis for particular question types. The issue of whether asking many questions of a certain type is beneficial for building mathematical knowledge may deserve investigation, in addition to the issues that we have raised here.

We add a caveat with regard to choosing to examine teachers' questions, as opposed to students' responses to those questions. After all, the inferences we wish to make concern students' cognitive activity that may be fostered by the instructional activity of the teachers. We have two responses to this concern. First, others (e.g., Benware & Deci, 1984) have shown that it is the mere anticipation of verbalization that leads to desirable learning outcomes. Second, the extent to which the students' verbalizations indicate high cognitive engagement is very much dependent on the extent to which the teachers promote such thinking. Although measuring student responses to questions would also be informative, teacher questions provide one way of assessing the kinds of classroom events that prompt students' cognitive activity.

The question types examined in the present research are potentially generalizable beyond first-grade classrooms and beyond mathematics. If future research uses this classification scheme with older students, the differences between the question types may be even more striking. The kinds of questions that can be asked in first-grade classrooms are restricted by the cognitive limitations of children of that age. Still, we detected differences in question types fairly readily. As students get older, teachers might be more likely to ask

higher level questions. Although the level of knowledge required to answer questions of all types may be increased, the distinctions between the question types may remain. We encourage future research that would test the generalizability of our approach.

The question classification scheme is potentially applicable to other problem-solving disciplines. We suspect that the sorts of distinctions among question types that we found in first-grade addition and subtraction lessons (e.g., placing a problem in a concrete context vs. a symbolic context) also exist in disciplines such as physics and chemistry, in which solving problems is a major component of the coursework. In general, some of the question types (in particular, problem-solving strategies and conceptual knowledge questions) reflect relatively complexly structured knowledge, of the sort that people use when they are experts in a field. We suspect that the question classification scheme presented in this study has potential for use in other classroom research, as a tool for investigating relationships between instruction and cognition.

It should be noted that we are not making any causal inferences about the effectiveness of certain question types on academic achievement, although the data and analyses permit generating causal hypotheses. We recognize the correlational nature of the study and the important role that other variables play in explaining achievement disparities between the United States and Asian countries. Thus, we caution against making a causal argument until such time when the hypotheses generated from our analyses of the observations are tested in a more thorough and rigorous manner.

We hope that U.S. teachers might benefit from the knowledge obtained from the analyses conducted here. For example, U.S. teachers might be encouraged to ask questions that require conceptual knowledge, especially in first-grade mathematics lessons. Introducing teachers to the potential benefits of asking certain types of questions may help them construct lessons that make optimal use of questions.

In conclusion, it appears that Asian students are challenged more by the questions their teachers ask them than are U.S. students. We believe that asking higher level questions may lead to the formation of the relatively rich and broad conceptual knowledge about mathematics demonstrated by Asian children.

References

- Benware, C. A., & Deci, E. L. (1984). Quality of learning with an active versus a passive motivational set. *American Educational Research Journal*, 21, 755-765.
- Bloom, B. (1956). *Taxonomy of educational objectives*. New York: Longman.
- Brown, A., & Kane, M. J. (1988). Preschool children can learn to transfer: Learning to learn and learning from example. *Cognitive Psychology*, 20, 493-523.
- Crosswhite, F. J., Dossey, J. A., Swafford, J. O., McKnight, C. C., & Cooney, T. J. (1985). *Second international mathematics study summary report for the United States*. Champaign, IL: Stipes.
- Graham, T. (1991, July). *Mathematical explanations in Japanese, Taiwanese, Chinese, and American classrooms*. Paper presented at the biennial meeting of the International Society for the Study of Behavioural Development, Minneapolis, MN.
- Husen, T. (1967). *International study of achievement in mathematics*. New York: Wiley.
- Martin, V. L., & Pressley, M. (1991). Elaborative-interrogation effects depend on the nature of the question. *Journal of Educational Psychology*, 83, 113-119.
- Perry, M. (1989, April). Explanations of mathematical concepts in Japanese, Chinese, and U.S. first- and fifth-grade classrooms. In S.-Y. Lee (Chair), *Effective schools, effective homes: A comparison of the achievement of American, Chinese, and Japanese children*. Symposium conducted at the biennial meeting of the Society for Research in Child Development, Kansas City, MO.
- Perry, M. (1992). *Explanations of mathematical concepts in Japanese, Chinese, and U.S. classrooms*. Manuscript submitted for publication.
- Pressley, M., Symons, S., McDaniel, M. A., Snyder, B. L., & Turnure, J. E. (1988). Elaborative interrogation facilitates acquisition of confusing facts. *Journal of Educational Psychology*, 80, 268-278.
- Redfield, D. L., & Rousseau, E. W. (1981). A meta-analysis of experimental research on teacher questioning behavior. *Review of Educational Research*, 51, 237-245.
- Robitaille, D. F., & Garden, R. A. (1989). *The IEA study of mathematics: II. Contexts and outcomes of school mathematics*. Elmsford, NY: Pergamon Press.
- Samson, G. E., Strykowski, B., Weinstein, T., & Walberg, H. J. (1987). The effects of teacher questioning levels on student achievement: A quantitative synthesis. *Journal of Educational Research*, 80, 290-295.
- Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory*, 4, 592-604.
- Stevenson, H. W., & Bartsch, K. (1992). An analysis of Japanese and American textbooks in mathematics. In R. Leestma & H. Walberg (Eds.), *Japanese education* (pp. 103-133). Ann Arbor, MI: Center for Japanese Studies.
- Stevenson, H. W., Lee, S.-Y., & Stigler, J. W. (1986). Mathematics achievement of Chinese, Japanese, and American children. *Science*, 231, 693-699.
- Stevenson, H. W., Lee, S.-Y., Chen, C., Stigler, J. W., Hsu, C. C., & Kitamura, S. (1990). Contexts of achievement: A study of American, Chinese, and Japanese children. *Monographs of the Society for Research in Child Development*, 55(1-2, Serial No. 221).
- Stevenson, H. W., & Stigler, J. W. (1992). *The learning gap*. New York: Summit Books.
- Stigler, J. W., Fuson, K. C., Ham, M., & Kim, M. S. (1986). An analysis of addition and subtraction word problems in American and Soviet elementary mathematics textbooks. *Cognition and Instruction*, 3, 153-171.
- Stigler, J. W., Lee, S.-Y., Lucker, G. W., & Stevenson, H. W. (1982). Curriculum and achievement in mathematics: A study of elementary school children in Japan, Taiwan, and the United States. *Journal of Educational Psychology*, 74, 315-322.
- Stigler, J. W., Lee, S.-Y., & Stevenson, H. W. (1987). Mathematics classrooms in Japan, Taiwan, and the United States. *Child Development*, 58, 1272-1285.
- Stigler, J. W., Lee, S.-Y., & Stevenson, H. W. (1990). *Mathematical knowledge of Japanese, Chinese, and American elementary school children*. Reston, VA: National Council of Teachers of Mathematics.
- Stigler, J. W., & Perry, M. (1988a). Cross-cultural studies of mathematics teaching and learning: Recent findings and new directions. In D. A. Grouws & T. J. Cooney (Eds.), *Effective mathematics teaching* (pp. 194-223). Hillsdale, NJ: Erlbaum.

- Stigler, J. W., & Perry, M. (1988b). Mathematics learning in Japanese, Chinese, and American classrooms. In G. B. Saxe & M. Gearhart (Eds.), *Children's mathematics* (pp. 27–54). (New Directions for Child Development, Series No. 41.) San Francisco: Jossey-Bass.
- Winer, B. J. (1971). *Statistical principles in experimental design* (2nd ed.). New York: McGraw-Hill.
- Winne, P. H. (1979). Experiments relating teachers' use of higher cognitive questions to student achievement. *Review of Educational Research*, 49, 13–50.

Received September 19, 1991
 Revision received June 12, 1992
 Accepted June 17, 1992 ■

Instructions to Authors

Authors should prepare manuscripts according to the *Publication Manual of the American Psychological Association* (3rd ed.). All manuscripts must include an abstract containing a maximum of 960 characters and spaces (which is approximately 120 words) typed on a separate sheet of paper. Typing instructions (all copy must be double-spaced) and instructions on preparing tables, figures, references, metrics, and abstracts appear in the *Publication Manual*. All manuscripts are subject to editing for sexist language. Manuscript length per se is not an issue, although length should be related to the manuscript's "information value." For further details about appropriate manuscript length and content, authors are referred to the Editorials in the March 1991 issue (Vol. 83, No. 1, pp. 5–7), the March 1992 issue (Vol. 84, No. 1, pp. 3–5), and the March 1993 issue (Vol. 85, No.1, pp. 3–6) of the *Journal*.

APA policy prohibits an author from submitting the same manuscript for concurrent consideration by two or more publications. In addition, it is a violation of APA Ethical Principles to publish "as original data, data that have been previously published" (Standard 6.24). As this journal is a primary journal that publishes original material only, APA policy prohibits as well publication of any manuscript that has already been published in whole or substantial part elsewhere. Authors have an obligation to consult journal editors concerning prior publication of any data upon which their article depends. In addition, APA Ethical Principles specify that "after research results are published, psychologists do not withhold the data on which their conclusions are based from other competent professionals who seek to verify the substantive claims through reanalysis and who intend to use such data only for that purpose, provided that the confidentiality of the participants can be protected and unless legal rights concerning proprietary data preclude their release" (Standard 6.25). APA expects authors submitting to this journal to adhere to these standards. Specifically, authors of manuscripts submitted to APA journals are expected to have available their data throughout the editorial review process and for at least 5 years after the date of publication.

Authors will be required to state in writing that they have complied with APA ethical standards in the treatment of their sample, human or animal, or to describe the details of treatment. A copy of the APA Ethical Principles may be obtained by writing the APA Ethics Office, 750 First Street, NE, Washington, DC 20002-4242.

Because the *Journal* has a masked review policy, authors submitting manuscripts are requested to include with each copy of the manuscript a cover sheet that shows the title of the manuscript, the authors' names and institutional affiliations, the date the manuscript is submitted, and footnotes identifying the authors or their affiliations. The first page of the manuscript should omit the authors' names and affiliations but should include the title of the manuscript and the date it is submitted. Every effort should be made by the authors to see that the manuscript itself contains no clues to their identities.

Authors should submit manuscripts in quadruplicate. All copies should be clear, readable, and on paper of good quality. A dot matrix or unusual typeface is acceptable only if it is clear and legible. Manuscripts not meeting readability and APA *Publication Manual* specifications will be returned for repair before being reviewed. In addition to addresses and phone numbers, authors should supply electronic mail addresses and fax numbers, if available, for potential use by the editorial office and later by the production office. Authors should keep a copy of the manuscript to guard against loss. Mail manuscripts to the Editor, Joel R. Levin, Department of Educational Psychology, University of Wisconsin, 1025 West Johnson Street, Madison, Wisconsin 53706-1796.
