# Development of Arithmetical Competencies in Chinese and American Children: Influence of Age, Language, and Schooling

David C. Geary; C. Christine Bow-Thomas; Fan Liu; Robert S. Siegler

*Child Development*, Vol. 67, No. 5 (Oct., 1996), 2022-2044.

# Development of Arithmetical Competencies in Chinese and American Children: Influence of Age, Language, and Schooling

**David C. Geary and C. Christine Bow-Thomas**
*University of Missouri at Columbia*

**Fan Liu**
*China Central Institute for Educational Science*

**Robert S. Siegler**
*Carnegie Mellon University*

GEARY, DAVID C.; BOW-THOMAS, C. CHRISTINE; LIU, FAN; and SIEGLER, ROBERT S. *Development of Arithmetical Competencies in Chinese and American Children: Influence of Age, Language, and Schooling.* CHILD DEVELOPMENT, 1996, **67**, 2022–2044. The arithmetical competencies of more than 200 Chinese or American kindergarten, first-, second-, or third-grade children were assessed toward the beginning and toward the end of the U.S. school year. All children were administered a paper-and-pencil test of addition skills, a digit span measure, and an addition strategy assessment. The addition strategy assessment provided information on the types of strategies the children used to solve simple addition problems as well as information on the speed and accuracy of their strategy use. Information on the number of math instruction periods across times of measurement was also obtained for each of the first-, second-, and third-grade children. The pattern of arithmetical development across the academic year and across the Chinese and American children suggests that a mix of cultural and maturational factors influence the emergence of early arithmetical competencies and that the Chinese advantage in early mathematical development is related to a combination of language- and school-related factors.

Children's mathematical development is influenced by a complex mix of cultural and biological factors (Geary, 1995; Gelman, 1990; Siegler & Crowley, 1994). More important, it appears that different cultural and maturational influences affect different aspects of children's developing mathematical competencies. Bisanz, Morrison, and Dunn (1995), for instance, recently demonstrated that different aspects of children's early numerical competencies were differentially related to chronological age and schooling. In particular, performance on number conservation tasks was strongly related to age but not schooling, whereas performance on a mental addition task (i.e., solving problems without the use of paper and pencil) was related to quantity of schooling and age. For the mental addition task, schooling was related to the accuracy of problem solving, whereas age was related to the types of strategies used in problem solving.

The apparent specificity of the relations among age, schooling, and early numerical competencies suggests that a complete understanding of the development of these competencies will require a fine-grained assessment of the relation between component mathematical skills and potential maturational and cultural influences on their development. In this study, the influence of age and two cultural factors, language-related differences in number words and schooling,

We would like to thank Linda Coutts, Dave Brunda, Larry Jones, Jennifer Mueller, Mary O'Brien, Mary Hoard, Becky Gregory, Melissa Moon, and Nuoo-Ting Lin for their assistance with data collection and collation, Yuhong Yao and Zhitang Liu for their assistance with the translation and back-translation of the Chinese materials, and Jeffrey Bisanz, Robert Kail, and two anonymous reviewers for comments on an earlier draft. The study was supported by grant 1R01-HD27931 from the National Institute of Child Health and Human Development and a grant from the Research Board of the University of Missouri, both awarded to the first author.
Correspondence concerning this article should be addressed to David C. Geary, Department of Psychology, 210 McAlester Hall, University of Missouri, Columbia, MO 65211; e-mail: psycorie@mizzou1.missouri.edu.

[*Child Development*, 1996, **67**, 2022–2044.  0009-3920/96/6705-0037$01.00]

on the development of addition skills were examined for groups of kindergarten, first-, second-, and third-grade children from mainland China and the United States. All children were assessed toward the beginning and toward the end of the U.S. academic year and were administered psychometric (e.g., paper-and-pencil tests) and information processing (e.g., speed of problem solving) tasks. The administration of different tasks enabled a fine-grained assessment of the arithmetical competencies of Chinese and American children as well as an assessment of how language and schooling might differentially affect the component skills underlying the development of these competencies. The design of the study also allowed for an assessment of the relation between chronological age and children's emerging arithmetical competencies. In the two sections that follow, the potential effects of language and schooling on mathematical development are summarized. A brief overview of the development of addition skills follows the latter section.

*Language and Mathematical Development*

There appear to be at least two ways in which language might influence mathematical development. The first involves the influence of the structure of number words on emerging numerical competencies (Fuson & Kwon, 1991; Geary, 1994; Miller, Smith, Zhu, & Zhang, 1995; Miura, Okamoto, Kim, Steere, & Fayol, 1993). In East Asian languages, the structure of the base 10 number system is transparently represented in the structure of the number words themselves. For instance, the Chinese number words for 11, 12, and 13, are translated as *ten one, ten two,* and *ten three,* respectively. The fact that 11, for instance, is composed of a single tens value and a single units value is obvious in the associated Chinese number word but not in the English number word *eleven.* The difference in the structure of number words appears to create differences in the ease with which East Asian and American children can learn to count past 10, their understanding of the base 10 system itself, and the types of strategies that might be used to solve simple arithmetic problems (Fuson & Kwon, 1992a; Miller et al., 1995; Miura et al., 1993).

Fuson and Kwon (1992b) found that the structure of Asian language number words supported Korean children's use of decomposition as a strategy for solving simple addition and simple subtraction problems. For Korean children, decomposition is centered on 10, which follows from the base 10 structure of the number words themselves. For instance, to solve the problem 6 + 7, the child might decompose the 7 into a 3 and 4, then add 6 + 4, and finally 10 + 3. At the same time, the structure of number words does not appear to be related to other aspects of children's early numerical competencies, such as understanding the one-to-one relation between number words and counted items (i.e., during counting each object is tagged with a different number word; Miller et al., 1995).

The second way in which language might influence mathematical development is through the speed with which basic number names (e.g., one, two, three) can be pronounced. Speed of number pronunciation influences the number of digits that can be retained in working memory (i.e., digit span), which, in turn, may influence the types of strategies that children can use to solve simple arithmetic problems (Geary, Bow-Thomas, Fan, & Siegler, 1993; Stigler, Lee, & Stevenson, 1986). In particular, when an answer cannot be retrieved, American children tend to count to solve simple arithmetic problems. Counting can occur with the aid of fingers or without them. The use of fingers during counting appears to be a working memory aid that allows the child to keep track of the addends physically, rather than mentally, during the process of counting (Geary, 1990).

The apparent relation between digit span and finger counting might result in an early difference in the mix of strategies used by Chinese and American children to solve simple arithmetic problems. In particular, the short digit spans of American children may lead them to rely on finger counting more often than Chinese children. This is because Chinese number words can be spoken more quickly than English number words, which affords Chinese individuals about a two-digit span advantage over Americans (Stigler et al., 1986). Indeed, when answers to simple addition problems cannot be retrieved from memory, then American kindergartners tend to count on their fingers, whereas Chinese kindergartners tend to count verbally (Geary et al., 1993). This difference in the type of counting did not exist for Chinese and American children who were matched on digit span. Further, for the American sample, frequency of finger counting increased as digit span decreased.

Thus, differences in the structure of Chinese and English number words may influence the relative frequency with which Chinese and American children use decomposition and finger counting as problem-solving strategies. However, different features of these number words appear to influence strategy usage in different ways. In particular, digit span might be selectively related to finger counting but not the use of decomposition or any other strategy. If the structure of Chinese number words does facilitate the use of decomposition, as it appears to in Korean children, then not only should Chinese children use decomposition more frequently than their American peers, decomposition should be primarily centered on 10, as illustrated above.

*Schooling and Mathematical Development*

In the present study, the same-grade Chinese and American children were of the same age and were assessed during the same calendar-year time frames. Based on these controls, inferences could be made about the effects of schooling, broadly defined, and age on the development of early arithmetical competencies. Any influence of age on addition skill development should be reflected in parallel changes in the component skills of Chinese and American children across times of measurement. With an analysis of variance (ANOVA), any such effect should result in a significant main effect for time and a nonsignificant nation × time interaction, with one caution. Across times of measurement, change in age and change in time in school are necessarily the same. As a result, the design is sensitive to those component skills that change with age and are minimally affected by *differences* in the schooling of Chinese and American children.

A significant nation × time interaction, in turn, might reflect differences in the schooling of Chinese and American children or interactions between age and schooling. It cannot be stated with any certainty, however, which features of schooling might contribute to any such interaction, nor can additional influences, such as parental support for mathematics learning be ruled out. Thus, a significant time × nation interaction could reflect differences in the in-school instruction of Chinese and American children, differences in quantity of homework, parental and peer influences, or some interaction between instruction and language-related influences (Geary, 1994; Perry, VanderStoep, & Yu, 1993). Nevertheless, to bolster the argument that in-school differences in instruction contribute to Chinese and American differences in the acquisition of early arithmetical competencies, data were collected on the number of math instruction periods between times of measurement, and the relation between quantity of instruction and the academic-year change in the children's arithmetical competencies was assessed.

*Development of Addition Skills*

In this section, a brief overview of the development of addition skills is presented; detailed descriptions are provided elsewhere (Ashcraft, 1992; Geary, 1994; Siegler, 1986, 1987). Mastery of basic addition skills, that is, for solving single-digit problems (e.g, 5 + 6), is achieved when all basic facts (e.g., the addition table) can be retrieved accurately and quickly from long-term memory (Geary, 1994). Prior to this, when children, and adults, cannot retrieve an answer from memory, they tend to resort to one of three types of backup strategy. The first involves some form of counting, either with fingers or without them (i.e., verbal counting). The most sophisticated counting procedure involves starting the count with the value of the larger addend and then counting the number of times indicated by the value of the smaller addend, as in solving 5 + 3 by counting "5, 6, 7, 8." This procedure is termed *min counting* or *counting on* (Fuson, 1982; Groen & Parkman, 1972). Alternatives to min counting involve counting both addends (*sum counting*) or counting on from the smaller-valued addend (*max counting;* Groen & Parkman, 1972).

The second category of backup strategy involves decomposition, which, as described earlier, tends to be centered on 10 or on tie (e.g., 6 + 7 = 6 + 6 + 1) problems (Geary, 1994). Third, children sometimes use the fingers strategy to solve simple arithmetic problems (Siegler & Shrager, 1984). The fingers strategy involves looking at uplifted fingers, which represent the addends, but not counting them to get the answer. Under these circumstances, the use of fingers appears to prompt direct retrieval.

*Overview*

In all, the study provides a documentation of academic-year changes in the arithmetical competencies of Chinese and American kindergarten through third-grade children. The study also allows inferences to be drawn about the potential influence of language structure, schooling, broadly defined, and age on the development of arithmetical competencies and explores the potential relation between these factors and differences in the arithmetical development of Chinese and American children.

## Method

*Participants*

The participants included 105 (50 males, 55 females) elementary school children from Columbia, Missouri, and 104 (54 males, 50 females) elementary school children from Hangzhou, China. As shown in Table 1, the mean age of the same-grade Chinese and American subjects was comparable at all grade levels. The difference in the number of boys and girls in the American sample was not significant across grade levels, $\chi^2(3) = 5.4$, $p > .10$, nor was the difference in the numbers of boys and girls across the Chinese and American samples, $\chi^2(1) = 0.4$, $p > .50$. All children were selected from the same elementary schools used in our original study of the arithmetical abilities of Chinese and American children (Geary, Fan, & Bow-Thomas, 1992). In this original study, the relative performance of our Chinese and American children on a test of addition skills was comparable to the performance of larger and more representative samples of Chinese and American children for a similar test (Stevenson, Lee, Chen, Lummis, et al., 1990). Thus, the schools used in our original study and in the current study would appear to be well matched, in terms of providing a relatively unbiased assessment of the arithmetic skills of urban American and Chinese children.

*Experimental Tasks*

All children were administered a paper-and-pencil test of addition skills, a digit span measure, and an addition strategy assessment. The task stimuli were identical for the English and Chinese versions of the tasks. The procedures used for the translation and back-translation of task instructions were identical to those described in Geary et al. (1992).

*Paper-and-pencil addition test.*—Two versions of the addition test were constructed, a simple one for the kindergarten children and a more advanced one for the first- through third-graders. Each version included two forms. Both forms of the kindergarten version consisted of 63 items, which ranged from 1 + 1 to 5 + 5, inclusive; each problem was presented two or three times. Both versions of the more advanced test consisted of 81 items, those defined by the pairwise combinations of 1 to 9, inclusive. For each form, the subject was allowed 1 min to solve as many problems as possible. The score was the number of items solved correctly across both forms of the test.

*Digit span measure.*—The same memory span measure was administered to the kindergarten and older children. The stimuli consisted of 14 strings of digits, including two strings for each length of 3–9 digits. Digit strings excluded consecutive numbers and the consecutive presentation of numbers with similar sounding word tags in Chinese. The experimenter read the numbers at a rate of one per second, beginning with the three-digit string and proceeding to progressively larger strings. The subject was re-

TABLE 1

SUBJECT CHARACTERISTICS

| | Age | | Gender | |
|---|---|---|---|---|
| Grade | *M* | SD | Male | Female |
| China: | | | | |
| Kindergarten | 71 | 3.1 | 13 | 13 |
| First | 83 | 2.4 | 13 | 13 |
| Second | 94 | 2.8 | 14 | 12 |
| Third | 105 | 3.9 | 14 | 12 |
| United States: | | | | |
| Kindergarten | 71 | 3.7 | 16 | 9 |
| First | 83 | 4.6 | 15 | 14 |
| Second | 94 | 4.2 | 8 | 11 |
| Third | 104 | 4.7 | 11 | 21 |

NOTE.—Age is in months.

quired to repeat the exact sequence. Testing was stopped when the subject missed both sequences of the same length. The digit span score was the number of digits in the longest digit string correctly recalled.

*Addition strategy assessment.*—Three sets of stimuli were constructed for the addition strategy assessment task, two for the kindergarten children and one for the older children. For the kindergarten children, the stimuli in the first set consisted of 25 single-digit addition problems, those defined by the pairwise combination of the integers 1 to 5, inclusive. At the second time of measurement, the kindergarten children were also administered a second set of nine larger-valued addition problems, which consisted of problems with sums greater than 10 (e.g., 5 + 6, 7 + 8, 7 + 6). The inclusion of these larger-valued problems allowed us to assess the problem-solving competencies of Chinese and American kindergarten children for problems where decomposition might be used as a backup strategy. For the older children, the stimuli consisted of 40 single-digit addition problems, which ranged from 2 + 3 to 9 + 8 (no tie problems, such as 3 + 3, 4 + 4, were included).

The problems were presented, one at a time, at the center of a cathode-ray tube (CRT) controlled by a microcomputer. For each problem, a prompt appeared at the center of the CRT for 1,000 ms, followed by a blank screen for 1,000 ms. Then, the problem appeared on the screen and remained until the child responded by speaking the answer into a voice operated relay connected to the microcomputer. A hardware clock ensured the collection of reaction time (RT) with an accuracy of about ±1 ms. Equal emphasis was placed on speed and accuracy of responding; specifically, "try to get the answer as fast as you can, but try not to make many mistakes" in the English version and "of course, you have to get the answer quickly and correctly" in the Chinese version.

*Mathematics Instruction*

To obtain a rough estimate of the children's exposure to mathematics in school, we calculated both the number and the average duration of math instruction periods between the first and second measurement for each first-, second-, and third-grade child. We were unable to estimate the quantity of math instruction for the American kindergarten children, because there were no formal math instruction periods for these children. Rather, any exposure to mathematics was "embedded" in other activities (e.g., counting out snacks).

In China, mathematics is taught based on a national curriculum that determines the content and length of math classes. Based on this curriculum and teacher reports, the average duration of a math instruction period for the Chinese students was estimated to be 40 min. All of the same-grade Chinese children received their math instruction from the same teacher and at the same time. In the United States, there is no national curriculum. As a result, the duration of math instruction periods can vary considerably from one teacher to the next (Stevenson, Lee, & Stigler, 1986). Thus, the teachers of our American participants were asked to provide information on the length and general content of each math lesson between the two times of measurement. To facilitate this process, folders with lines for each instruction day and places to note the length and content of each math lesson were developed for each teacher. The teacher simply had to fill in the appropriate lines each day. For each teacher, two independent scorers calculated the mean number of minutes in a math lesson (i.e., total minutes of math instruction/number of math instruction periods).

*Procedure*

School begins in the first week of September in China and the United States. For the first measurement (time 1), all data were collected in both countries between November 4 and December 5. For the second measurement (time 2), all data were collected in both countries between April 20 and May 15. All of the Chinese subjects were tested at both times of measurement, and 103 of the 105 American subjects were tested at both times of measurement (two American first graders moved between the first and second assessment).[1] The Chinese children were individually tested in a quiet room in the Engineering Psychology laboratory at Hangzhou University, whereas the American children were individually tested in a quiet room in their own school. For the first assessment, all participants were first administered the digit span measure, then the paper-

[1] The strategy choice data were also lost for one American second grader (for time 1). All analyses were based on all available data for the respective time of measurement.

and-pencil addition test, and finally the addition strategy assessment in a single session that lasted less than 30 min. For the second assessment, all participants were administered the paper-and-pencil addition test followed by the addition strategy assessment. For the addition strategy assessment, the nine larger-valued addition problems were administered to the kindergarten children following presentation of the 25 smaller-valued (i.e., sums <11) addition problems.

For the strategy assessment task, the answer and strategy used to solve the problem were recorded on a trial-by-trial basis by the experimenter and classified as one of the following strategies; counting fingers, fingers, verbal counting, retrieval, or decomposition. Strategy classifications were, for the most part, based on the child's behavior during problem solving (Siegler & Shrager, 1984). If the child was observed moving his or her fingers in sequence during problem solving, then the strategy was classified as counting fingers. If the child counted aloud or softly during problem solving, or if indications of subauditory vocalizations were present (e.g., lip movements), then the strategy was classified as verbal counting. The counting trials were further classified based on where counting began; that is, the trials were classified as min, sum, or max. If the child looked at her or his fingers but did not count them to get an answer, then the trial was classified as fingers.

If the child spoke the answer without counting on fingers, counting verbally, or looking at their fingers, then the strategy was initially classified as retrieval. After each trial, subjects were asked to describe how they arrived at the answer. For trials initially classified as retrieval, if the child described a stepwise process (e.g., $7 + 6 = 6 + 6 = 12$, $12 + 1 = 13$) then the strategy was reclassified as decomposition. Although the distinction between retrieval and decomposition was based solely on the child's description, the finding that mean RTs for decomposition trials were generally twice the mean RTs for retrieval trials (see Tables 3, 4, 5, and 6) suggests that the descriptions were generally accurate. The decomposition trials were further classified based on the form of decomposition used in problem solving, that is, 10 based, tie based, or other (e.g., for $9 + 4$, $9 + 3 = 12$, $12 + 1 = 13$). Tie-based decomposition is theoretically important, because, unlike 10-based decomposition, its use is not dependent on an understanding of the base 10 structure of the number system. Comparisons of the child's description and the experimenter's initial classification indicated agreement between the child and the experimenter on more than 90% of the trials for each of the samples. For those trials on which the child and experimenter disagreed, the strategy was classified based on the child's description, except in cases where overt finger counting or verbal counting (e.g., counting aloud) was observed by the experimenter and the child claimed to have used some other strategy. For these trials, the trial was classified based on the experimenter's observation.

## Results

For ease of discussion, the results are presented in six sections. In the first two, results for performance on the paper-and-pencil addition test and the digit span measure are presented, respectively. Results for performance on the strategy assessment task are presented in the third section. The fourth section presents an analysis of the relation between paper-and-pencil addition test performance and the children's strategy choices, while the fifth section presents an analysis of the relation between digit span and strategy choices. The final section presents results for the math instruction data. Across sections, an alpha level of .05 was adopted to indicate statistical significance, unless otherwise noted.

### *Paper-and-Pencil Addition Test*

In this section, performance on the paper-and-pencil addition test is described separately for the kindergarten and older children, because the addition test differed for these groups.

*Kindergarten children.*—Figure 1 shows the kindergarten children's mean performance on the paper-and-pencil addition test. A repeated-measures ANOVA, with nation as a between-subjects factor and time as a within-subjects factor, revealed significant main effects for nation, $F(1, 49) = 113.26$, and time, $F(1, 49) = 117.23$, and a significant nation × time interaction, $F(1, 49) = 34.13$. Although dependent $t$ tests revealed significant improvements in addition test performance from time 1 to time 2 for both the Chinese and American children, the significant nation × time interaction confirmed greater academic-year gains by the Chinese children relative to their American peers. A separate analysis revealed a nonsignificant

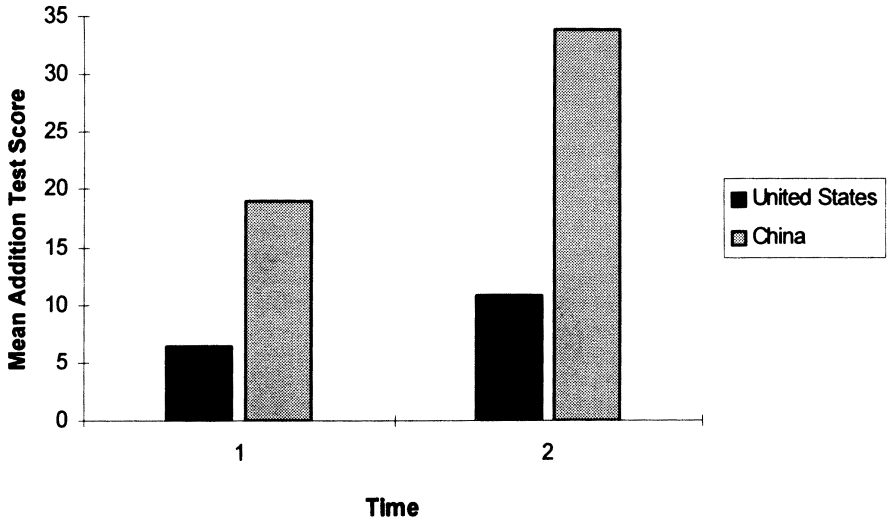


FIG. 1.—Mean addition test scores for the Chinese and American kindergarten children

main effect for gender and a nonsignificant gender × nation interaction.

*First- through third-grade children.*—The addition test performance of the older children was assessed by means of a 2 (nation) × 3 (grade) × 2 (time) mixed-design ANOVA, with nation and grade as between-subjects factors and time as a within-subjects factor. The ANOVA revealed that all main effects and interactions were significant; nation, $F(1, 150) = 772.63$, grade, $F(2, 150) = 192.47$, time, $F(1, 150) = 508.56$, nation × grade, $F(2, 150) = 36.69$, nation × time, $F(1, 150) = 222.21$, grade × time, $F(2, 150) = 16.16$, and nation × grade × time, $F(2, 150) = 24.82$. A separate analysis indicated that the main effect for gender was not significant nor were any of the interactions involving gender.

The significant three-way interaction is shown in Figure 2 and reflects differences in the relative gains, from time 1 to time 2, across the Chinese and American samples and across grade levels. Post hoc orthogonal contrasts for the interaction revealed significant linear trends across grade for the U.S. sample, $F(1, 76) = 4.07$. Specifically, linear increases in addition test performance were found across grade level, but the within-grade gains across time 1 and time 2 decreased linearly across grade level. For the Chinese sample, a linear increase in addition test performance was found across grade level, $F(1, 75) = 10.47$, but the relative gains from time 1 to time 2 within each grade level differed and was best fitted by a quadratic contrast, $F(1, 75) = 54.93$. More important, the overall pattern indicates much greater gains at each grade level across the academic year for the Chinese children relative to their American peers.

### *Digit Span*

The digit span scores were analyzed across all four grade levels, because the same test was administered to the kindergartners and the older children. Mean digit span scores across nation and grade are shown in Figure 3; recall, the test was only administered at time 1. A 2 (nation) × 4 (grade) ANOVA revealed significant main effects for nation, $F(1, 201) = 437.45$, and grade, $F(3, 201) = 10.83$, but a nonsignificant nation × grade interaction. Post hoc comparisons (honestly significant difference test) revealed that the mean scores of the kindergarten and first-grade children did not differ significantly, nor did the mean scores of the second- and third-grade children. All other comparisons were significant.

### *Addition Strategy Assessment*

Strategy choice and RT data are considered separately for the kindergarten and older children, because the experimental stimuli differed for these groups. For the strategy choice data, the percentage of trials on which each subject used each strategy was calculated; the reported group means represent the mean of the individual percentages (the analyses of the strategy data, however, were based on raw frequencies).

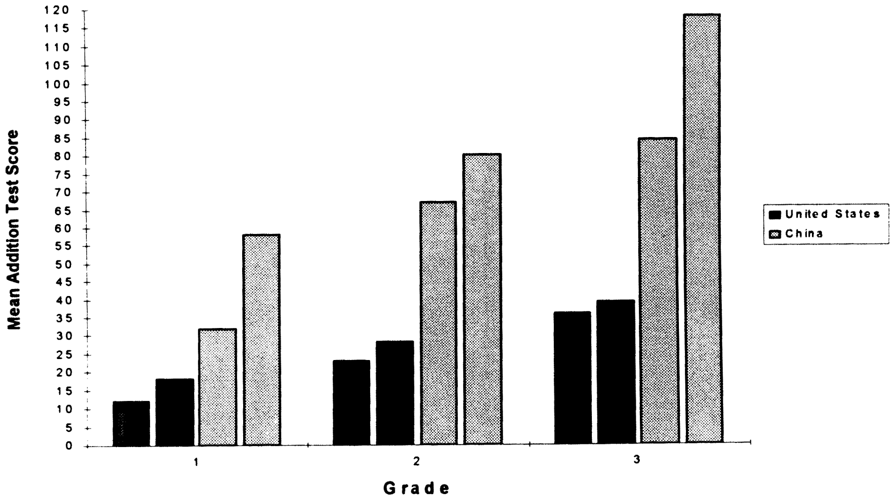


FIG. 2.—Mean addition test scores for the older Chinese and American children. For each set, the leftmost bar represents the mean score for time 1 and the adjacent bar represents the mean score for time 2.

The degrees of freedom differ across some of the analyses, because not all subjects used all strategies. Nevertheless, this procedure seems preferable to averaging data across strategies (Siegler, 1987). To make the results easier to follow, significant between-group effects, which were analyzed by means of ANOVAs, are noted in the tables by an asterisk. Within-group changes in strategy characteristics from time 1 to time 2 were analyzed by means of dependent $t$ tests, and significant differences are underscored for time 2. Due to the length and complexity of these analyses, a summary of the primary findings is presented at the end of this section.

*Kindergarten children.*—Table 2 shows the distribution of strategy choices, error percentages, use of min counting, and RTs for the smaller-valued problems (i.e., problems with sums <11). Here, it can be seen that the Chinese and American kindergarten children used the same types of strategies to solve the addition problems but differed significantly in the distribution of their strategy choices. For time 1, the American children relied more heavily on finger counting, $F(1, 49) = 6.01$, and retrieval, $F(1, 49) = 11.85$, than their Chinese peers, who, in turn, used the fingers, $F(1, 49) = 5.91$, and verbal counting, $F(1, 49) = 36.98$, strategies more frequently than the American children.

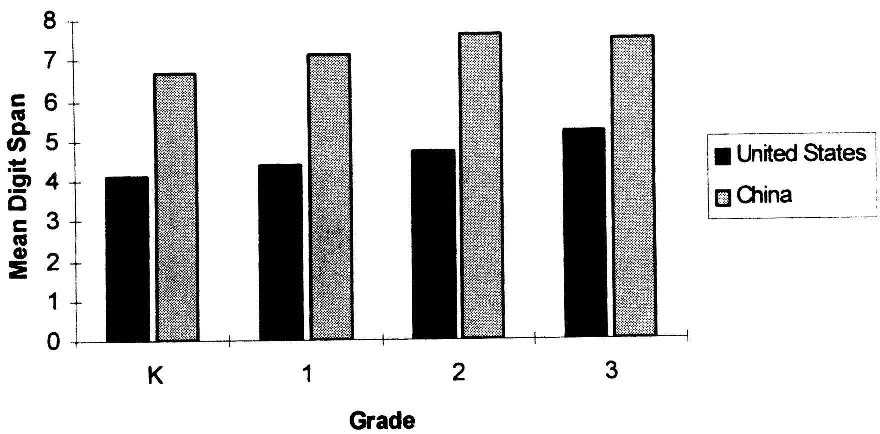


FIG. 3.—Mean digit span scores for the Chinese and American children

TABLE 2

CHARACTERISTICS OF ADDITION STRATEGIES IN KINDERGARTEN FOR PROBLEMS WITH SUMS <11

| STRATEGY | Mean Percentage of Trials | | Mean Percentage of Errors | | Mean Percentage of Min Strategy Trials | | Mean Reaction Time (in Seconds) | |
|---|---|---|---|---|---|---|---|---|
| | China | United States | China | United States | China | United States | China | United States |
| Time 1: | | | | | | | | |
| Counting fingers | 11 | 29* | 8 | 13 | 13 | 6 | 9.3 | 7.6 |
| Fingers | 11 | 0* | 9 | ... | ... | ... | 5.8 | ... |
| Verbal counting | 47 | 12* | 5 | 13 | 69 | 41* | 3.2 | 4.6* |
| Retrieval | 31 | 59* | 1 | 33* | ... | ... | 1.6 | 2.8* |
| Time 2: | | | | | | | | |
| Counting fingers | 0 | 32* | ... | 13 | ... | 15 | ... | 8.9 |
| Fingers | 0 | 3[a] | ... | 3 | ... | ... | ... | 5.4 |
| Verbal counting | 16 | 11 | 2 | 7 | 86 | 49* | 3.7 | 5.3[b] |
| Retrieval | 84 | 54* | 1 | 8* | ... | ... | 1.5[c] | 2.6* |

NOTE.—Mean reaction time excluded error and spoiled trials as well as outliers (i.e., RTs <500 ms or >2.5 SDs for the mean RT for individual subjects and strategies). Significant, $p < .05$, changes across times of measurement are underscored for time 2. Significant, $p < .05$, group differences are noted by an *. Fingers involves uplifting fingers to represent the value of one or both addends but not counting them before stating an answer.

[a] $p = .065$, for the group difference.

[b] $p = .056$, for the group difference.

[c] $p = .073$, for the differences across times of measurement.

The percentage of errors did not differ significantly across groups for either finger counting or verbal counting, but the Chinese children committed significantly fewer retrieval errors than their American peers, $F(1, 48) = 18.10$. Thus, the national difference in the frequency of retrieval was likely due to a lot of guessing on the part of the American children. In fact, a comparison of the number of addition facts that were correctly retrieved revealed no significant difference between the Chinese and American children.

For verbal counting trials, the Chinese children used the min procedure significantly more often than the American children, $F(1, 34) = 6.22$, and had significantly faster mean RTs, $F(1, 32) = 5.12$. The Chinese advantage in mean retrieval RTs was also significant, $F(1, 41) = 32.86$, but the advantage of the American children for finger counting RTs was not.

For time 2, the Chinese children relied primarily on retrieval to solve the presented problems, whereas the American children primarily used a combination of retrieval and finger counting. The difference in the frequency with which the Chinese and American children used the various strategies was significant for counting fingers, $F(1, 49) = 30.60$, and retrieval, $F(1, 49) = 25.96$, but not for fingers or verbal counting. As was found for time 1, the American children committed significantly more retrieval errors at time 2 than their Chinese peers, $F(1, 48) = 12.69$, but the difference in the percentage of verbal counting errors was not significant. As was found for time 1, the American children used the min procedure less frequently than the Chinese children when they used verbal counting, $F(1, 34) = 12.92$. The advantage of the Chinese children for retrieval RTs was again significant, $F(1, 47) = 75.91$, and was marginally significant for verbal counting RTs, $F(1, 32) = 3.92$, $p = .056$.

Across times of measurement, there was a clear shift from the use of counting to the use of retrieval for the Chinese kindergarten children. In contrast, there was little change in the strategy mix across times of measurement for the American kindergarten children. Indeed, for the Chinese children, the change in strategy usage across times of measurement was significant for all four strategies. In contrast, the change was not significant for any of the strategies for the American children. Across groups, the only significant change in error percentage was the decrease in retrieval errors for the American children, $t(22) = -3.01$. Across time 1 and time 2, mean verbal counting RTs increased significantly for both the Chinese, $t(18) = 2.16$, and American, $t(14) = 4.07$, children. The mean finger counting RTs were also significantly higher at time 2 relative to time 1 for the American children, $t(19) = 3.97$. Finally, the decrease in mean retrieval RTs was marginally significant for the Chinese children, $t(25) = -1.87$, $p = .073$, but was not significant for the American children.

Table 3 shows that the Chinese and American kindergarten children used the same types of strategies to solve the larger-valued addition problems, those with sums greater than 10, but differed greatly in their strategy distributions. To solve these problems, the American children used finger counting much more frequently than the Chinese children, $F(1, 48) = 92.85$, who, in turn, used verbal counting, $F(1, 48) = 6.83$, and decomposition, $F(1, 48) = 29.45$, much more frequently than the American children. The Chinese and American children did not differ in the use of retrieval, but the American children committed many more retrieval errors than their Chinese peers, $F(1, 23) = 7.61$. The group difference, which favored the Chinese children, in the percentage of verbal counting errors, $F(1, 12) = 3.57$, $p = .083$, and decomposition errors, $F(1, 22) = 3.52$, $p = .074$, was marginally significant but was not significant for finger counting errors. Finally, the only significant difference in mean RTs was for retrieval. Here, the Chinese children were significantly faster than the American children, $F(1, 16) = 23.42$.

Another important comparison involves the difference in the mix of strategies used for solving smaller- (for time 2) and larger-valued problems. Recall, smaller-valued problems were those with sums $<11$, whereas larger-valued problems were those with sums $>10$. For the Chinese kindergarten children, decomposition was used more frequently to solve larger-valued problems, $t(25) = 6.12$, whereas retrieval was used more frequently to solve smaller-valued problems, $t(25) = -11.97$; differences in the relative use of counting fingers, fingers, and verbal counting were not significant. The most striking difference for the American kindergarten children was the increased use of counting fingers for solving larger-valued problems, $t(23) = 5.61$. Relative to smaller-

TABLE 3

CHARACTERISTICS OF ADDITION STRATEGIES IN KINDERGARTEN FOR PROBLEMS WITH SUMS >10

| STRATEGY | MEAN PERCENTAGE OF TRIALS[a] | | MEAN PERCENTAGE OF ERRORS | | MEAN PERCENTAGE OF MIN STRATEGY TRIALS | | MEAN REACTION TIME (in Seconds) | |
|---|---|---|---|---|---|---|---|---|
| | China | United States | China | United States | China | United States | China | United States |
| Counting fingers | 3 | 76* | 25 | 44 | 25 | 13 | 7.6 | 14.1 |
| Fingers | 7 | 0 | 6 | . . . | . . . | . . . | 13.4 | . . . |
| Verbal counting | 25 | 5* | 3 | 33[b] | 86 | 33* | 10.8 | 15.2 |
| Retrieval | 21 | 17 | 8 | 47* | . . . | . . . | 2.3 | 4.8* |
| Decomposition | 44 | 2* | 3 | 25[c] | . . . | . . . | 6.0 | 7.3 |

NOTE.—Mean reaction times excluded error and spoiled trials as well as outliers. Significant, $p < .05$, group differences are noted by an *. Fingers involves uplifting fingers to represent the value of one or both addends but not counting them before stating an answer.

[a] For 15% of 3% of the trials, for the U.S. and China samples, respectively, the subject indicated that he or she did not know the answer or did not respond to the item. The reported results include only trials for which the problem was solved, correctly or incorrectly.

[b] $p = .083$, for the group difference.

[c] $p = .074$, for the group difference.

valued problems, the decrease in the use of retrieval for solving larger-valued problems was also significant, $t(23) = -5.81$, as was the slight increase in the use of decomposition, $t(23) = 2.10$. Differences in the relative use of fingers and verbal counting were not significant for the American children.

Across samples, only two significant changes in the proportion of errors were found. For the American children, the proportion of counting fingers and retrieval errors increased significantly, comparing smaller to larger problems.

*First- through third-grade children.*—Table 4 shows that the Chinese and American first graders used the same types of strategies to solve the addition problems but, again, differed significantly in their strategy mix. At time 1, the Chinese children used retrieval, $F(1, 53) = 14.00$, and decomposition, $F(1, 53) = 191.41$, much more frequently than the American children, who, in turn, relied on finger counting, $F(1, 53) = 21.54$, and verbal counting, $F(1, 53) = 7.62$, much more frequently than the Chinese children. The American children also committed many more verbal counting, $F(1, 47) = 10.50$, and retrieval, $F(1, 41) = 9.64$, errors than their Chinese peers. For verbal counting trials, there was no practical difference in the Chinese and American children's use of the min procedure (97% and 100%, respectively). Finally, mean RTs were significantly faster for the Chinese children, relative to the American children, for verbal counting, $F(1, 47) = 10.80$, and retrieval, $F(1, 39) = 143.09$.

For time 2, the distribution of strategy choices across the Chinese and American first graders was highly similar to that found in our previous study of first graders (Geary et al., 1992; testing in the two studies was done at roughly the same point in the school year) and again showed large cross-national differences in the strategy distributions. The Chinese children used retrieval significantly more often than the American children, $F(1, 51) = 148.74$, who, in turn, relied on finger counting, $F(1, 51) = 22.40$, and verbal counting, $F(1, 51) = 52.28$, much more frequently than the Chinese children. The American children were more error prone than the Chinese children, for both retrieval, $F(1, 46) = 6.53$, and decomposition, $F(1, 27) = 8.03$, and had significantly higher mean RTs for verbal counting, $F(1, 34) = 15.54$, retrieval, $F(1, 45) = 76.15$, and decomposition, $F(1, 26) = 184.73$.

For the Chinese children, the change in strategy usage across times of measurement was significant for verbal counting, $t(25) = -4.70$, retrieval, $t(25) = 13.04$, and decomposition, $t(25) = -10.74$. For the American children, the only significant changes across times of measurement were increases in the use of decomposition, $t(26) = 3.08$, and min counting for finger counting trials, $t(14) = 3.53$. Across groups, the only significant change in error percentage was for decomposition. Here, the Chinese children committed fewer errors at time 2 than at time 1, $t(18) = -3.13$. Finally, for the American children, mean RTs were significantly higher at time 2 than at time 1 for counting fingers, $t(19) = 2.15$, but significantly lower for retrieval, $t(12) = -4.03$. For the Chinese children, mean RTs were significantly lower at time 2 relative to time 1 for both retrieval, $t(25) = -6.76$, and decomposition, $t(18) = -5.70$.

Tables 5 and 6 show that the Chinese second and third graders relied almost exclusively on retrieval to solve the basic addition problems, whereas the American children continued to use a combination of retrieval and counting (verbal and finger). Given the almost exclusive reliance on retrieval in the Chinese children, the analyses of the strategy characteristics of the second and third graders focused on retrieval trials, that is, on the frequency of retrieval, the frequency of retrieval errors, and on mean retrieval RTs.

A 2 (nation) × 2 (grade) × 2 (time) mixed ANOVA confirmed that the Chinese children used retrieval more than the American children, $F(1, 98) = 210.45$, the third graders retrieved more answers than the second graders, $F(1, 98) = 6.46$, and that retrieval increased from time 1 to time 2, $F(1, 98) = 18.31$. The nation × time effect was the only significant interaction, $F(1, 98) = 7.84$, but was due to a ceiling effect; that is, 100% retrieval at both times of measurement for the Chinese third graders.

The analyses of retrieval errors revealed significant nation, $F(1, 98) = 42.33$, and nation × time effects, $F(1, 98) = 8.25$, but all other main effects and interactions were nonsignificant. These significant results were largely due to floor effects; that is, the low frequency of retrieval errors for both the American and Chinese second and third graders.

The analyses of retrieval RTs revealed that all main effects and interactions were

TABLE 4

CHARACTERISTICS OF ADDITION STRATEGIES IN FIRST GRADE

| STRATEGY | MEAN PERCENTAGE OF TRIALS[a] | | MEAN PERCENTAGE OF ERRORS | | MEAN PERCENTAGE OF MIN STRATEGY TRIALS | | MEAN REACTION TIME (in Seconds)[b] | |
|---|---|---|---|---|---|---|---|---|
| | China | United States | China | United States | China | United States | China | United States |
| Time 1: | | | | | | | | |
| Counting fingers | 0 | 34* | ... | 21 | ... | 69 | ... | 8.1 |
| Fingers | 0 | 2 | ... | 15 | ... | ... | ... | ... |
| Verbal counting | 18 | 42* | 1 | 11* | 97 | 100[c] | 2.9 | 4.9* |
| Retrieval | 43 | 20* | 3 | 22* | ... | ... | 1.5 | 3.6* |
| Decomposition | 36 | 1* | 6 | 0 | ... | ... | 3.6 | ... |
| Time 2: | | | | | | | | |
| Counting fingers | 0 | 22* | ... | 17 | ... | 96 | ... | 8.5 |
| Fingers | 0 | 0 | ... | ... | ... | ... | ... | ... |
| Verbal counting | 3 | 46* | 0 | 7 | 99 | 92 | 2.4 | 4.5* |
| Retrieval | 91 | 28* | 2 | 12* | ... | ... | 1.2 | 2.7* |
| Decomposition | 6 | 4 | 1 | 14* | ... | ... | 2.2 | 4.9* |

NOTE.—Mean reaction times excluded error and spoiled trials as well as outliers. Significant, $p < .05$, changes across times of measurement are underscored for time 2. Significant, $p < .05$, group differences are noted by an *. Fingers involves uplifting fingers to represent the value of one or both addends but not counting them before stating an answer.

[a] Columnar sums may not equal 100 due to rounding. Also, for time 1, for 2% of the trials the Chinese children reported using a combination of two strategies, such as counting and retrieval.

[b] For the U.S. sample, time 1 mean RTs are not reported for fingers and decomposition due to the small number of trials for these strategies.

[c] $p = .055$, for the group difference.

TABLE 5

CHARACTERISTICS OF ADDITION STRATEGIES IN SECOND GRADE

| STRATEGY | MEAN PERCENTAGE OF TRIALS | | MEAN PERCENTAGE OF ERRORS | | MEAN PERCENTAGE OF MIN STRATEGY TRIALS | | MEAN REACTION TIME (in Seconds)[a] | |
|---|---|---|---|---|---|---|---|---|
| | China | United States | China | United States | China | United States | China | United States |
| Time 1: | | | | | | | | |
| Counting fingers | 0 | 35* | ... | 10 | ... | 91 | ... | 5.7 |
| Fingers | 0 | 0 | ... | ... | ... | ... | ... | ... |
| Verbal counting | 2 | 33* | 0 | 6 | 100 | 99 | ... | 4.2 |
| Retrieval | 94 | 31* | 5 | 4 | ... | ... | 1.1 | 2.6* |
| Decomposition | 4 | 1* | 2 | 11 | ... | ... | 2.5 | ... |
| Time 2: | | | | | | | | |
| Counting fingers | 0 | 25 | ... | 12 | ... | 92 | ... | 5.7 |
| Fingers | 0 | 0 | ... | ... | ... | ... | ... | ... |
| Verbal counting | 1[b] | 34 | 0 | 6 | 100 | 98 | ... | 4.1 |
| Retrieval | 98 | 41[c] | 4 | 4 | ... | ... | 1.0 | 2.1* |
| Decomposition | 1 | 1 | 7 | 0 | ... | ... | ... | ... |

NOTE.—Mean reaction times excluded error and spoiled trials as well as outliers. Significant, $p < .05$, changes across times of measurement are underscored for time 2. Significant, $p < .05$, group differences are noted by an *. Fingers involves uplifting fingers to represent the value of one or both addends but not counting them before stating an answer.

[a] For the U.S. sample, mean RTs are not reported for decomposition due to the small number of trials for this strategy. For the same reason, time 2 mean RTs are not reported for verbal counting or decomposition for the Chinese sample.

[b] $p = .065$, for the difference across time of measurement.

[c] $p = .054$, for the difference across time of measurement.

TABLE 6

CHARACTERISTICS OF ADDITION STRATEGIES IN THIRD GRADE

| STRATEGY | MEAN PERCENTAGE OF TRIALS | | MEAN PERCENTAGE OF ERRORS | | MEAN PERCENTAGE OF MIN STRATEGY TRIALS | | MEAN REACTION TIME (in Seconds) | |
|---|---|---|---|---|---|---|---|---|
| | China | United States | China | United States | China | United States | China | United States |
| Time 1: | | | | | | | | |
| Counting fingers .......... | 0 | 23* | ... | 6 | ... | 96 | ... | 4.7 |
| Fingers .......................... | 0 | 0 | ... | ... | ... | ... | ... | ... |
| Verbal counting ............ | 0 | 28* | ... | 2 | ... | 99 | ... | 3.2 |
| Retrieval ....................... | 100 | 45* | 5 | 0* | ... | ... | .8 | 2.0* |
| Decomposition ............. | 0 | 4[a] | ... | 13 | ... | ... | ... | 3.0 |
| Time 2: | | | | | | | | |
| Counting fingers .......... | 0 | 15* | ... | 7 | ... | 93 | ... | 4.8 |
| Fingers .......................... | 0 | 0 | ... | ... | ... | ... | ... | ... |
| Verbal counting ............ | 0 | 24* | ... | 8 | ... | 98 | ... | 3.2 |
| Retrieval ....................... | 100 | 56* | 3 | 2 | ... | ... | .8 | 1.9* |
| Decomposition ............. | 0 | 4* | ... | 0 | ... | ... | ... | ... |

NOTE.—Mean reaction times excluded error and spoiled trials as well as outliers. Significant, $p < .05$, changes across times of measurement are underscored for time 2. Significant, $p < .05$, group differences are noted by an *. Fingers involves uplifting fingers to represent the value of one or both addends but not counting them before stating an answer.

[a] $p = .065$, for the group difference.

significant, except the nation × grade interaction, $F(1, 91) = 268.00, 17.25, 49.02, 29.13, 15.10, 7.09$ for the nation, grade, time, nation × time, grade × time, and nation × grade × time effects, respectively. The three-way interaction is shown in Figure 4 and reflects a larger change in mean RTs across times of measurement for the American children relative to the Chinese children, $F(1, 93) = 16.41$, and larger changes in mean RTs for the second graders relative to the third graders, $F(1, 93) = 5.43$.

*Decomposition.*—In order to determine the form of decomposition used by the Chinese and American children, decomposition trials were individually examined for those samples where decomposition was most frequently used; that is, for Chinese kindergartners when solving the larger-valued problems, Chinese first graders (both times of measurement), American first graders at the first measurement, and American third graders (both times of measurement). As noted earlier, for each decomposition trial, the procedure was scored as being centered on 10, based on the use of tie problems, or other.

Across grade levels, 78% of the Chinese children's decomposition trials were centered on 10, relative to 45% of the American children's decomposition trials, $z = 8.3$. In comparison, 42% of the American children's decomposition trials were based on tie problems, relative to 20% of the Chinese children's decomposition trials, $z = -5.5$. There were also within-nation shifts in the relative use of tie-based or 10-based forms of decomposition across grade level. For the Chinese kindergarten children, 40% and 53% of the decomposition trials were tie based or 10 based, respectively, as compared to 16% and 83%, respectively, for the Chinese first graders (almost identical percentages were found for both times of measurement), $\chi^2(1) = 37.4$. For the American first graders (time 2), 56% and 24% of the decomposition trials were tie based or 10 based, respectively, as compared to 37% and 53%, respectively, for the American third graders (across times of measurement), $\chi^2(1) = 8.1$.

*Summary.*—Across grade levels, the Chinese and American children used the same types of strategies to solve simple addition problems but differed considerably in the distribution of strategy choices and in the academic-year changes in the strategy mix. When they counted, the Chinese children were much more likely to rely on verbal counting, as contrasted with the American children's reliance on finger counting, and, at least in kindergarten, were more likely than the American children to use the min procedure. Differences in the use of decomposition were also striking. The Chinese children used decomposition much more frequently than the American children. In fact, decomposition was the primary backup

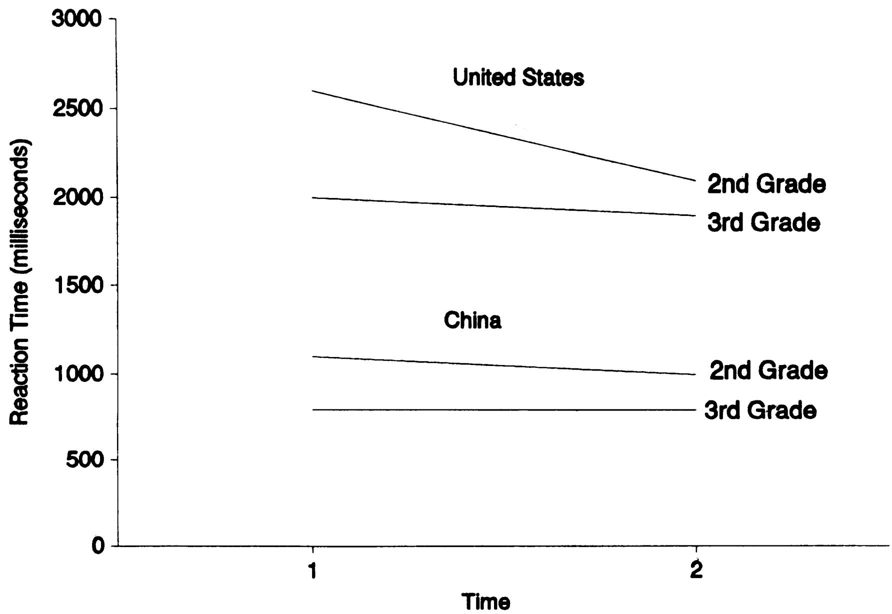


FIG. 4.—Mean retrieval RTs for the Chinese and American second and third grade children

strategy for the Chinese children, whereas counting fingers was the primary backup strategy for the American children. Equally important, 10-based decomposition was used much more frequently by the Chinese children than the American children, and the general reliance on 10-based, as opposed to tie-based, decomposition occurred much earlier in the Chinese children than in the American children.

From time 1 to time 2, changes in the Chinese children's strategy mix were rather dramatic, with, up until the point of 100% retrieval, important increases in the frequency of direct retrieval across times of measurement. Within-grade changes in the American children's strategy mix across times of measurement were much less dramatic and often not statistically significant, or practically significant for that matter.

### *Strategy Choices and Addition Test Performance*

In this section, the relation between strategy choices and addition test performance is presented separately for kindergarten and older children.

*Kindergarten children.*—Geary and Burlingham-Dubree (1989) found that for preschool and kindergarten children, a combination of the frequency of correct retrieval, speed of fact retrieval, and efficient use of backup strategies provided a good representation of individual differences in psychometric test performance in arithmetic. For the current study, these factors were represented, respectively, by the frequency of correct retrieval, mean retrieval trial RTs, and a backup strategy variable. The backup strategy variable represented the sophistication and accuracy of the strategies used for problem solving when the correct answer could not be retrieved. Specifically, if the frequency of backup strategies (i.e., fingers, and verbal and finger counting) was greater than the frequency of retrieval errors, then backup was coded F + min + .5(sum + max), where F refers to the frequency of use of the fingers strategy and min and (sum + max) refer to the frequency with which these procedures were used across the verbal and finger-counting strategies. Otherwise, backup was coded 0 − (frequency of retrieval errors) (see Geary, 1990; Geary & Burlingham-Dubree, 1989). A low score on the backup strategy variable represented either frequent guessing (i.e., retrieval errors) or frequent counting errors, whereas a high score represented the frequent use of the min counting procedure, with either finger or verbal counting.

Across groups, this set of three variables was highly correlated with kindergartners' performance on the addition test at time 1, $F(3, 43) = 36.87$, $R^2 = .72$, and each of the individual variables was uniquely predictive of addition test performance. Basically, the more efficient the use of backup strategies and the more facts that were correctly retrieved the higher the addition test score. Lower mean retrieval RTs (i.e., faster fact retrieval) were also associated with higher addition test scores. These three variables also predicted individual differences in addition test performance within the Chinese, $F(3, 22) = 5.01$, $R^2 = .41$, and American, $F(3, 17) = 10.00$, $R^2 = .64$, samples. Finally, partialing these three variables reduced the advantage of the Chinese children over their American peers on the addition test to nonsignificance. This set of three variables was also highly predictive of addition test performance at time 2, $F(3, 45) = 46.15$, $R^2 = .75$, but nation still explained a significant amount of incremental variance, $\Delta R^2 = .05$, $F(1, 44) = 11.24$.

*First- through third-grade children.*—For the older children the relation between strategy choices and addition test performance was straightforward: Across groups and grade level, the frequency of correct retrieval and mean retrieval trial RTs were the primary source of individual differences in addition test performance. At time 1 and across grade level, these two variables explained 77% (i.e., $R^2$) of the variance in addition test performance; these variables explained 74% and 72% of the variance in addition test performance within the Chinese and American samples, respectively. Finally, partialing these two variables reduced the Chinese advantage on the addition test, across grade levels, to nonsignificance. Again, the same pattern was found for time 2—these variables explained 68% of the variance in addition test performance across groups and grade level—except that they did not completely explain the Chinese advantage on the addition test, $\Delta R^2 = .02$, $F(1, 141) = 9.79$. Nation explained an additional 2% of the performance variance on the addition test above and beyond the 68% explained by the retrieval variables.

### *Digit Span and Strategy Choices*

As noted earlier, for the time 1 performance of our samples of Chinese and American kindergarten children, we showed a relation between digit span performance and the use of the counting fingers strategy (Geary et al., 1993). Basically, shorter digit spans were associated with the frequent use of finger counting. In this section, we extend this analysis to time 2 for the Chinese and American kindergarten children and to first through third grade for the American children; an assessment of the relation between finger counting and digit span was not possible for the older Chinese children, because they did not use the counting fingers strategy (see Tables 4 to 6).

For the kindergarten children, the relation between strategy choices and digit span was examined by means of partial correlations. After partialing the effect of nation, digit span was correlated with the frequency with which the counting fingers, verbal counting, decomposition (for the larger-valued problems), and fingers strategies were used, as well as with the number of facts correctly retrieved. For the counting strategies, digit span was also correlated with the use of the min and sum procedures. For time 1, the use of the counting fingers strategy and the use of the sum procedure for finger counting was significantly and inversely related to digit span; $pr(49) = -.31$ and $-.31$ for overall frequency of finger counting and frequency with which the sum procedure was used with finger counting, respectively. At time 2, digit span was correlated with the use of the sum procedure during finger counting for the larger-valued problems, $pr(48) = -.28$, $p = .055$, but was not correlated with finger counting for the smaller-valued problems. Finally, digit span was not significantly related to the use of fingers, verbal counting, correct retrieval, or decomposition at either time of measurement.

For the first-, second-, and third-grade American children, digit span showed a significant or marginally significant correlation with the frequency of finger counting for five of the six comparisons; no other strategy variable was consistently related to the digit span measure (i.e., no more than a total of two significant or marginally significant correlations). For first graders, counting fingers and digit span were not correlated at time 1, $r(27) = -.01$, but were correlated at time 2, $r(25) = -.37, p = .061$. For second and third graders, digit span was inversely related to finger counting at both times of measurement, $rs = -.30$ to $-.45$, $ps < .10$. Across grade level, digit span and the frequency of finger counting were significantly related at time 1, $r(77) = -.23$, and time 2, $t(76) = -.38$, and remained marginally significant at time 1 and significant at time 2 after partialing grade and age, $pr(75) = -.21$, $p = .073$ and $pr(74) = -.37$, respectively. As a comparison, digit span was not significantly correlated with the frequency of decomposition for the older American children within grade level or across grade level for either time of measurement, $rs$ ranged from $-.09$ to $.27$. Finally, digit span was not significantly related to the use of decomposition in the Chinese kindergartners for solving larger-valued problems, $r(24) = .02$, or for the Chinese first graders, $r(24)s = .35$ and $.10$ for time 1 and time 2, respectively.

### *Mathematics Instruction and Addition Development*

In this section, the national difference in the number of math instruction periods between the first and second measurements is described first, followed by an assessment of the relation between the national difference in the quantity of math instruction and the advantage of the Chinese children on the addition tasks. The focus is on instructional periods because preliminary analyses revealed that the average duration (in minutes) of a math instruction period was not related to the national differences on the addition tasks. The relation between math instruction and addition development focuses on the paper-and-pencil addition test, the frequency of correct retrieval, and the speed of correct retrieval. Performance on the addition test provides a global assessment of basic addition skills, while the frequency and speed of fact retrieval provide more direct measures of the degree to which basic addition skills have been mastered (Geary & Burlingham-Dubree, 1989; Siegler, 1988).

Table 7 shows the mean number of math instruction periods across times of measurement for the Chinese and American first

TABLE 7

MEAN NUMBER OF MATHEMATICS INSTRUCTION PERIODS BETWEEN THE FIRST AND SECOND TIME OF MEASUREMENT

| | China | | United States | |
|---|---|---|---|---|
| Grade | Mean | SD | Mean | SD |
| 1 ................ | 122 | 2 | 98 | 6 |
| 2 ................ | 122 | 3 | 96 | 7 |
| 3 ................ | 124 | 2 | 97 | 4 |

through third graders; even though the same-grade Chinese children received their math instruction from the same teacher, they differed in the number of instruction days because of differences in the length of time between the first and second measurement. There was no overlap at any grade level in the number of math instruction periods received by the Chinese and American children. Across grade levels, the Chinese children received between 24 and 27 more math instruction periods, on average, than their same-grade American peers.

Next, a set of regression equations was computed in order to estimate the extent to which the larger number of math instruction periods for the Chinese children might have contributed to their greater gains in addition test performance relative to their American peers. In the first equation, the time 2 addition test score served as the dependent measure, and the time 1 addition test score, nation (coded 0 for the United States and 1 for China), and grade served as independent measures. The raw regression coefficient of 26 for the nation variable indicated that, after controlling for time 1 performance and grade, the Chinese children solved, on average, 26 more addition problems (in 2 min) at time 2 than the American children (Cohen & Cohen, 1983). The second regression equation was identical to the first except that the number of math instruction periods was added as an independent measure. Controlling for the national difference in math instruction periods significantly reduced the raw coefficient for nation from 26 to 20, $t(153) = 1.97$, $p < .06$, suggesting that the Chinese advantage in math instruction periods contributed to, but did not totally explain, their time 2 performance advantage on the addition test.

The same procedures were used to assess the relation between math instruction and the change in the frequency and speed of correct retrieval across times of measurement. The results indicated that the Chinese advantage in math instruction was not statistically related to the national difference in the change in retrieval frequency across times of measurement. However, this null result needs to be interpreted with some caution, given that the frequency of correct retrieval was near ceiling for the Chinese children by the end of first grade. In contrast, statistically controlling for the national difference in the number of math instruction periods eliminated the Chinese advantage in retrieval speed (i.e., mean retrieval RTs).

## Discussion

The results of this study add to our understanding of the factors that influence children's early arithmetical development and to our understanding of the sources of cross-national achievement differences in elementary mathematics. Most generally, the results suggest that age, schooling, and language can differentially affect the emergence of different components of children's early numerical competencies (Bisanz et al., 1995) and that cross-national achievement differences in elementary mathematics reflect a mix of language and school-related factors, among others (see Geary, 1994; Stevenson & Stigler, 1992). In the first section below, the relation between language and arithmetical development is discussed and is followed by a consideration of the relations among age, schooling, and children's emerging arithmetical competencies. The discussion closes with a brief consideration of wider social and educational issues.

### *Language and Arithmetical Development*

The results of this study support the view that the structure of Asian language and English language number words influences the development of early numerical and arithmetical competencies (Fuson & Kwon, 1992a, 1992b; Miller et al., 1995; Miura et al., 1993). The finding that 10 based decomposition was the primary backup strategy of the Chinese children is consistent with the findings of Fuson and Kwon (1992b) for Korean children and supports their argument that the base 10 structure of Asian language number words facilitates the learning of the 10 based decomposition procedure in Asian children.

Digit span, which is influenced by the speed with which number words can be pronounced, also appears to be a language-related influence on children's strategy choices, in particular the use of finger counting as a backup strategy. Although not all of the correlations between digit span and the frequency of finger counting were significant, the overall pattern was consistent; relatively short digit spans were associated with frequent reliance on finger counting. Thus, individual differences in digit span appeared to be one factor that influenced individual differences in strategy choices within the American sample and might have been one factor associated with relatively early abandonment of finger counting in the Chinese children (see Geary et al., 1993). Moreover, digit span was not related to the use of decomposition for either the Chinese or American children, suggesting that number pronunciation speed and the base 10 structure of number words differentially influence children's strategy choices.

*Age, Schooling, and Arithmetical Development*

In this section, potential school-related and age-related influences on addition test performance, strategy choices, speed of strategy execution, and digit span are considered in turn.

Schooling effects are represented by the relatively larger academic-year gains in the arithmetical skills of the Chinese children in comparison to their same-grade American peers. As noted earlier, it cannot be stated with certainty which features of schooling might have been responsible for any such effects. For our design, schooling likely captures a variety of interacting influences, such as quality of instruction, amount of homework, and parental support of high academic standards (Geary, 1994; Stevenson, Lee, Chen, Stigler, et al., 1990). Nevertheless, the finding that the national difference in the number of math instruction periods across times of measurement explained a portion of the relatively greater academic-year gains in the arithmetical competencies of the Chinese children over the American children suggests that exposure to mathematics in school is one important source of the national difference in mathematical achievement.

Two features of children's early arithmetical development, addition test performance and the strategy mix, appeared to be especially influenced by differences in the schooling (broadly defined) of Chinese and American children. For all four grade levels, the Chinese children's addition test performance improved dramatically across the academic year, in comparison to the rather modest improvements of the American children. Across the academic year, there were also clear within-grade shifts from the use of backup strategies to the use of retrieval for the Chinese children. In comparison, within-grade shifts for the American children were quite small; there was, however, a more substantial shift across grade levels in the American children's tendency to use retrieval.

The argument that changes in children's strategy choices are strongly influenced by schooling appears to be inconsistent with the conclusions of Bisanz et al. (1995). Bisanz et al. found that Canadian children's strategy choices were more strongly related to age than to the quantity of schooling. The design of our study does not allow us to separate age and schooling effects within samples. Thus, it is possible that age is related, in part, to strategy changes (e.g., more retrieval) across the academic year, as argued by Bisanz et al. Nevertheless, the large increases in the use of retrieval in the Chinese children relative to the American children suggest that schooling is a potent influence on children's strategy choices, above and beyond any potential age-related influence.

As noted above, these schooling effects might also reflect the facilitating effect that the structure of Chinese number words has on the acquisition of arithmetical competencies in Chinese children. For instance, the use of 10 based decomposition by the Chinese children likely reflects both language and schooling effects. The structure of Asian number words makes it relatively easy for Asian teachers to teach the base 10 system and any associated base 10 problem-solving strategies (Fuson & Kwon, 1992a). Thus, Chinese teachers are more likely to teach 10 based decomposition than are American teachers and, due to the structure of Asian language number words, Chinese children are more likely to understand the utility of this procedure than are American children.

Moreover, the Chinese children's rapid shift from the use of backup strategies to the use of direct retrieval might have been influenced by the speed with which Chinese number words can be pronounced. Theoretically, direct retrieval results from the formation of associations between addition problems and their correct answers (Siegler, 1986). These associations appear to develop with the use of backup strategies. For instance, after a child has counted many times to solve an addition problem, the answer generated by the count becomes associated with the problem. For this association to be constructed, the problem's addends and the generated answer must be simultaneously active in working memory. The quantity of numbers that can be active in working memory is related to speed of counting. This perspective implies that problems and answers generated by means of counting should become associated in long-term memory with fewer practice trials when the counting is done in languages with relatively short number words.

The use of decomposition might also have influenced the national difference in the frequency of direct retrieval. This is because decomposition can be executed quickly, especially when contrasted with time-consuming finger counting. The fast execution of decomposition should enable all problem features and the generated answer to be simultaneously active in working memory, which, in turn, should facilitate the formation of problem/answer associations.

Changes in mean RTs across times of measurement are more difficult to interpret than are changes in the strategy mix, because many of the children used different strategies to solve the same problem at time 1 and time 2. For instance, the finding that mean counting (finger and verbal) RTs often increased across times of measurement for both the Chinese and American children probably reflects the tendency to use counting to solve fewer smaller-valued problems, such as 3 + 2, at time 2 than at time 1. The reason was that at time 2 answers were more likely to be retrieved for such smaller-valued problems.

For retrieval, there was a clear tendency for mean RTs to decrease across the academic year for both the Chinese and American children. The only exception to this pattern was for the Chinese third graders who appeared to have asymptoted in the speed of retrieval processes early in the academic year. The finding of faster retrieval RTs at time 2 than at time 1 for both the Chinese and American children indicates that age-related improvements in speed of accessing information from semantic memory cannot be ruled out (Kail, 1991). At the same time, the finding that the large Chinese advantage in mean retrieval RTs was eliminated by controlling for the Chinese advantage in number of math instruction periods suggests an important schooling effect on these mean RTs above and beyond any age-related changes.

Finally, the finding of a significant national difference in digit span is consistent with previous findings and presumably reflects the influence of differences in the speed of articulating Chinese and English number words on digit span performance (e.g., Stigler et al., 1986). The finding that the magnitude of the Chinese advantage in digit span did not change with successive years of schooling suggests that changes in digit span are more strongly related to chronological age than to differences in the schooling of Chinese and American children. In keeping with this interpretation is the finding of age-related improvements in the articulatory and memory processes that support performance on memory span tasks (e.g., Cowan et al., 1994). Of course, the change in digit span performance from one grade to the next might also have been related to features of schooling that were similar for the Chinese and American children.

*Social and Educational Implications*

The finding of substantial school-related differences in the emerging arithmetical competencies of Chinese and American children underscores the need for changes in mathematics education in the United States (Stevenson & Stigler, 1992). Although the language-related influences on the early arithmetical development of Chinese and American children appear to give the Chinese children an early advantage in acquiring some arithmetical competencies (Miller et al., 1995), these differences should only influence the early stages of skill acquisition and not the ultimate level of skill that can be achieved in arithmetic (Ackerman, 1988). The finding that American children who received their elementary school education in the 1930s appeared to develop the same level of arithmetical competencies (e.g., 100% retrieval by third grade) as was found for the Chinese children assessed in this study supports this argument (Geary, Salthouse, Chen, & Fan, 1996; Ilg & Ames, 1951). Stated differently, the language-related influences on early mathematical development are not a satisfactory explanation for the advantages that East Asian children have over American children in nearly all mathematical domains (e.g., Stevenson, Chen, & Lee, 1993), though they contribute to selective early differences (Geary, 1994; Miller et al., 1995).

The poor mathematical competencies of American children has wider social consequences. Basic quantitative skills influence employability, productivity, and wages (Bishop, 1989; Boissiere, Knight, & Sabot, 1985; Rivera-Batiz, 1992). Rivera-Batiz found that basic quantitative skills, as indexed by computational skills in arithmetic and the ability to solve arithmetic word problems, strongly influenced the likelihood of employment in the United States, above and beyond the influence of reading abilities and years of education, among other factors. Given this, the gap between the developing mathematical abilities of East Asian children and American children documented in this study and many other studies (e.g., Stevenson et al., 1993) bodes important long-term social and economic consequences for the United States.

## References


Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. *Journal of Experimental Psychology: General*, **117**, 288–318.

Ashcraft, M. H. (1992). Cognitive arithmetic: A review of data and theory. *Cognition*, **44**, 75–106.

Bisanz, J., Morrison, F. J., & Dunn, M. (1995). Effects of age and schooling on the acquisition of elementary quantitative skills. *Developmental Psychology*, **31**, 221–236.

Bishop, J. H. (1989). Is the test score decline responsible for the productivity growth decline? *American Economic Review*, **75**, 178–197.

Boissiere, M., Knight, J. B., & Sabot, R. H. (1985). Earnings, schooling, ability, and cognitive skills. *American Economic Review*, **75**, 1016–1030.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2d ed.). Hillsdale, NJ: Erlbaum.

Cowan, N., Keller, T. A., Hulme, C., Roodenrys, S., McDougall, S., & Rack, J. (1994). Verbal memory span in children: Speech timing clues to the mechanisms underlying age and word length effects. *Journal of Memory and Language*, **33**, 234–250.

Fuson, K. C. (1982). An analysis of the counting-on solution procedure in addition. In T. P. Carpenter, J. M. Moser, & T. A. Romberg (Eds.), *Addition and subtraction: A cognitive perspective* (pp. 67–81). Hillsdale, NJ: Erlbaum.

Fuson, K. C., & Kwon, Y. (1991). Chinese-based regular and European irregular systems of number words: The disadvantages for English-speaking children. In K. Durkin & B. Shire (Eds.), *Language in mathematical education: Research and practice* (pp. 211–226). Milton Keynes, PA: Open University Press.

Fuson, K. C., & Kwon, Y. (1992a). Korean children's understanding of multidigit addition and subtraction. *Child Development*, **63**, 491–506.

Fuson, K. C., & Kwon, Y. (1992b). Korean children's single-digit addition and subtraction: Numbers structured by ten. *Journal for Research in Mathematics Education*, **23**, 148–165.

Geary, D. C. (1990). A componential analysis of an early learning deficit in mathematics. *Journal of Experimental Child Psychology*, **49**, 363–383.

Geary, D. C. (1994). *Children's mathematical development: Research and practical applications*. Washington, DC: American Psychological Association.

Geary, D. C. (1995). Reflections of evolution and culture in children's cognition: Implications for mathematical development and instruction. *American Psychologist*, **50**, 24–37.

Geary, D. C., Bow-Thomas, C. C., Fan, L., & Siegler, R. S. (1993). Even before formal instruction, Chinese children outperform American children in mental addition. *Cognitive Development*, **8**, 517–529.

Geary, D. C., & Burlingham-Dubree, M. (1989). External validation of the strategy choice model for addition. *Journal of Experimental Child Psychology*, **47**, 175–192.

Geary, D. C., Fan, L., & Bow-Thomas, C. C. (1992). Numerical cognition: Loci of ability differences comparing children from China and the United States. *Psychological Science*, **3**, 180–185.

Geary, D. C., Salthouse, T. A., Chen, G. P., & Fan, L. (1996). Are East Asian versus American differences in arithmetical ability a recent phenomenon? *Developmental Psychology*, **32**, 254–262.

Gelman, R. (1990). First principles organize attention to and learning about relevant data: Number and animate-inanimate distinction as examples. *Cognitive Science*, **14**, 79–106.

Groen, G. J., & Parkman, J. M. (1972). A chronometric analysis of simple addition. *Psychological Review*, **79**, 329–343.

Ilg, F., & Ames, L. B. (1951). Developmental trends in arithmetic. *Journal of Genetic Psychology*, **79**, 3–28.

Kail, R. (1991). Developmental change in speed of processing during childhood and adolescence. *Psychological Bulletin*, **109**, 490–501.

Miller, K. F., Smith, C. M., Zhu, J., & Zhang, H. (1995). Preschool origins of cross-national differences in mathematical competence: The role of number-naming systems. *Psychological Science*, **6**, 56–60.

Miura, I. T., Okamoto, Y., Kim, C. C., Steere, M., & Fayol, M. (1993). First graders' cognitive representation of number and understanding of place value: Cross-national comparisons—France, Japan, Korea, Sweden, and the United States. *Journal of Educational Psychology*, **85**, 24–30.

Perry, M., VanderStoep, S. W., & Yu, S. L. (1993). Asking questions in first-grade mathematics classes: Potential influences on mathematical thought. *Journal of Educational Psychology,* **85,** 31–40.

Rivera-Batiz, F. L. (1992). Quantitative literacy and the likelihood of employment among young adults in the United States. *Journal of Human Resources,* **27,** 313–328.

Siegler, R. S. (1986). Unities across domains in children's strategy choices. In M. Perlmutter (Ed.), *Perspectives for intellectual development: The Minnesota Symposia on Child Psychology* (Vol. **19**, pp. 1–48). Hillsdale, NJ: Erlbaum.

Siegler, R. S. (1987). The perils of averaging data over strategies: An example from children's addition. *Journal of Experimental Psychology: General,* **116,** 250–264.

Siegler, R. S. (1988). Individual differences in strategy choices: Good students, not-so-good students, and perfectionists. *Child Development,* **59,** 833–851.

Siegler, R. S., & Crowley, K. (1994). Constraints on learning in nonprivileged domains. *Cognitive Psychology,* **27,** 194–226.

Siegler, R. S., & Shrager, J. (1984). Strategy choice in addition and subtraction: How do children know what to do? In C. Sophian (Ed.), *Origins of cognitive skills* (pp. 229–293). Hillsdale, NJ: Erlbaum.

Stevenson, H. W., Chen, C., & Lee, S. Y. (1993). Mathematical achievement of Chinese, Japanese, and American children: Ten years later. *Science,* **259,** 53–58.

Stevenson, H. W., Lee, S. Y., Chen, C., Lummis, M., Stigler, J., Fan, L., & Ge, F. (1990). Mathematics achievement of children in China and the United States. *Child Development,* **61,** 1053–1066.

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., Lee, S. Y., & Stigler, J. W. (1986). Mathematics achievement of Chinese, Japanese, and American children. *Science,* **231,** 693–699.

Stevenson, H. W., & Stigler, J. W. (1992). *The learning gap: Why our schools are failing and what we can learn from Japanese and Chinese education.* New York: Summit Books.

Stigler, J. W., Lee, S. Y., & Stevenson, H. W. (1986). Digit memory in Chinese and English: Evidence for a temporally limited store. *Cognition,* **23,** 1–20.