MATHEMATICAL COGNITION, 1996, *2* (1), 63–93

The Problem-size Effect in Mental Addition: Developmental and Cross-national Trends

David C. Geary
University of Missouri at Columbia, Columbia, Missouri, USA

Across two experiments, the magnitude of the problem-size effect in mental addition was examined for kindergarten and elementary school children, as well as adults, from mainland China and the United States. In North American samples, the problem-size effect represents the finding that arithmetic problems consisting of larger-valued numbers (e.g. 8+7) take longer to solve and are more error prone than are problems consisting of smaller-valued numbers (e.g. 2+3). This standard finding was found for the kindergarten, elementary school, and adult samples from the United States. For the Chinese children, the problem size effect was evident in kindergarten and at the beginning of first grade. However, the effect had disappeared at the end of first grade and had reversed (i.e. larger-valued addition problems were solved more quickly than smaller-valued problems) by the end of third grade. However, the standard problem-size effect "reappeared" for the Chinese adults. The results are interpreted in terms of theoretical models of the nature of the memory representation for arithmetic facts and in terms of the mechanisms that govern the development of these representations.

In the nearly 25 years since Groen and Parkman's (1972) seminal study of the mental processes underlying the solution of simple addition problems, cognitive arithmetic has emerged as a vibrant area of research. Scientists in this area have mapped the cognitive processes and neurological correlates that govern the mental solution of simple and complex arithmetic problems and have extended these basic findings to more applied issues, such as mathematical anxiety and mathematical disabilities (Ashcraft, 1992, 1995; Ashcraft & Faust, 1994; Ashcraft, Yamashita, & Aram, 1992; Campbell & Clark, 1988; Campbell & Graham, 1985; Dehaene & Cohen, 1991; Geary,

Requests for reprints should be sent to David C. Geary, Department of Psychology, 210 McAlester Hall, University of Missouri, Columbia, MO 65211, USA.

I would like to thank Christy Bow-Thomas, Jennifer Mueller, Mary O'Brien, Liu Fan, and Mary Hoard for their assistance with data collection and collation, and Mark Ashcraft, Brian Butterworth, and Jamie Campbell for comments on an earlier draft. The research 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.

1993; Geary, Widaman, & Little, 1986; LeFevre, Bisanz, & Mrkonjic, 1988; McCloskey, 1992; Miller, Perlmutter, & Keating, 1984; Widaman, Geary, Cormier, & Little, 1989). Yet, a complete explanation of the mech-anisms underlying a very basic phenomenon in mental arithmetic has eluded researchers since its first systematic study by Groen and Parkman (Ashcraft, 1995). This phenomenon, termed the *problem-size effect*, represents a systematic relationship between the magnitude of the numbers in simple arithmetic problems (those with two single integers) and the amount of time needed to solve the problems. Basically, simple arithmetic problems consisting of smaller-valued numbers, such as 2+3, are solved more quickly, and often more accurately, than are problems consisting of larger-valued numbers, such as 9+7 (Ashcraft & Battaglia, 1978; Geary, Frensch, & Wiley, 1993; Miller et al., 1984).

There have been two general classes of explanation for the problem-size effect. The first focuses on the nature of the representation of numerical and arithmetical information in semantic memory, and the second focuses on the different types of strategies used to solve arithmetic problems that include smaller- and larger-valued numbers (Ashcraft & Battaglia, 1978; Gallistel & Gelman, 1992; LeFevre, Sadesky, & Bisanz, in press; Siegler, 1987; Zbrodoff, 1995). The focus of the present paper is on the implications that the problem-size effect has for understanding the nature of the representation of arithmetic facts in semantic memory, and the factors that govern the development of these memory representations.

From the representational perspective, an early position was that the problem-size effect directly mirrors the organisation of arithmetic facts in semantic memory (Ashcraft & Battaglia, 1978; Geary et al., 1986; Miller et al., 1984). A common feature of these models is that arithmetic facts are organised as a two-dimensional "table-like" network, and the retrieval of these facts is governed by the spread of activation through this network. Due to the nature of the organisation of the network, activation spreads more quickly to facts associated with smaller-valued problems than to facts associated with larger-valued problems (Ashcraft & Battaglia, 1978). More recently, it has been argued that arithmetic fact retrieval is governed by multiple codes, specifically physical codes (i.e. arabic or verbal representations of the integers in the problem) and a system of potentially inherent analogue representations of the magnitudes associated with the integers 1 to 9 (Campbell, 1995; Dehaene & Cohen, 1991; Gallistel & Gelman, 1992). In other words, people, and many other animals, are born with a numerical system that can represent the quantities 1 to 9 (Gallistel & Gelman, 1992), although experience will almost certainly lead to the development of magnitude representations greater than 9 (Dehaene, 1992).

Within this numerical system, arithmetic fact retrieval is influenced by the degree of association between the integers in the physical code and the associated answer (termed unitisation by Campbell, 1995), by associations among related codes, and by the value of the underlying magnitude representations (Campbell, 1995; Gallistel & Gelman, 1992). For the magnitude representations, the distinction between successive values becomes fuzzy as the quantity increases (Moyer & Landauer, 1967). For instance, it is easier to discriminate the quantities associated with 1 and 2 than with 8 and 9. From this point of view, the less precise representation of the magnitudes associated with larger values results in longer reaction times (RTs) and more errors for arithmetic problems consisting of larger-valued numbers. This is so because the fuzzy boundaries around larger magnitudes can result in greater interference from adjacent magnitudes during fact retrieval. Moreover, it appears that the number of potential interfering associations among the physical codes (e.g. retrieving 12 for 3+4=?) increases with increases in the size of the integers in the problem (Campbell, 1995). Thus, as the size of the problem addends increases, the number of competing associations increases and the precision of the underlying magnitude representations decreases. The net result is an increase in RTs and error rates as the magnitude of the problem addends increases.

With extensive exposure, however, it is possible that fact retrieval becomes primarily or solely dependent on the physical-code associations between the problem integers and the associated answer, with little direct access to the analogue representations of magnitude, and little interference from competing associations (Campbell, 1995; Dehaene, 1992). This scenario is consistent, in some respects, with the model of arithmetic fact development proposed by Siegler (1986, 1987, 1988). In this model, the problem-size effect reflects both the different types of strategies used to solve smaller- and larger-valued arithmetic problems and the nature of the memory representation for arithmetic problems and the associated answers. During the initial acquisition of arithmetic skills, children have not yet developed associations between arithmetic problems and the correct answer, so they must rely on some form of reconstructive process to solve the problem. For simple addition, these reconstructive processes, which are sometimes called back-up strategies, typically involve some form of counting, either on fingers or verbally, or decomposition (Siegler, 1987; Siegler & Shrager, 1984). Decomposition involves breaking the problem into smaller problems. For instance, to solve the problem 8+5, the child might first retrieve the answer to 5+5, and then count, 11, 12, 13. The answer generated by means of decomposition or counting becomes directly associated with the problem, and after many such associations the problem/answer combination (i.e. the physical code for the problem/answer unit) is represented in semantic memory (Campbell, 1995; Siegler & Jenkins, 1989).

One important feature of Siegler's (1986, 1987, 1988) model is that the problem-size effect essentially reflects differences in the duration and frequency with which different back-up strategies are used to solve smaller- and larger-valued arithmetic problems, rather than reflecting the activational characteristics of a two-dimensional network of stored arithmetic facts, or access to any associated analogue magnitudes. If one has to count to solve an arithmetic problem, the counting takes longer and is more error prone for larger-valued (e.g. 8+7) as compared to smaller-valued (e.g. 2+3) problems. The problem-size effect then results primarily from the averaging of RTs across strategies (LeFevre et al., in press; Siegler, 1987). Once problem/answer associations have been developed for all simple arithmetic problems and assuming equal exposure to these problems, it follows from this model that retrieval speeds should be uniform across all simple arithmetic problems, regardless of the magnitude of the integers in the problem. This is so because the memory representations are specific problem/answer associations, which, once formed in memory, show the same representational and retrieval characteristics regardless of the magnitude of the problem integers. Nevertheless, with unequal exposure to all basic arithmetic problems, a problem-size effect might still be found even when all answers are retrieved (i.e. no back-up strategies are used). Here, the difference in retrieval speeds would reflect differences in the strength of the association between different problems and the associated answers. With extensive practice, however, the problem-size effect should disappear (Siegler, 1988).

Thus, Siegler's (1986) model essentially represents the physical-code feature of the multiple-code models (Campbell, 1995; Dehaene, 1992). In other words, for all of these models associations are formed between the problem integers and the generated answers, but Siegler makes no assumptions about the access of magnitude representations during arithmetic fact retrieval. An unresolved issue then is whether magnitude representations are always accessed during the retrieval of arithmetic facts, or whether arithmetic fact retrieval can be supported solely by physical-code representations. If analogue magnitudes are automatically accessed during the processing of arithmetic problems, then the problem-size effect (for RTs and error rates) should be evident even after extensive practice. If arithmetic fact retrieval is primarily associated with physical-code representations, and if repeated exposure inhibits the potential interfering effects of competing associations, then the problem-size effect should disappear after extensive practice. Any such disappearance of the problem-size effect would not necessarily disprove the multiple-code models. However, such a finding would suggest that any access to analogue magnitudes during fact retrieval and any influence of competing associations are less important contributors to the problem-size effect than frequency of exposure.

One method that might be used to determine whether the problem-size effect can be eliminated with extensive practice is to examine the magnitude of the problem-size effect in children from East Asian nations. Relative to American children, children in East Asian nations receive much more exposure to basic arithmetic and, in fact, to nearly all other mathematical domains (Geary, 1994; Stevenson et al., 1990a; Stevenson & Stigler, 1992). Moreover, the exposure of East Asian children to basic arithmetic facts is likely to be more extensive, at least in terms of repeated and long-term (i.e. spaced) exposure to these facts, than is easily accomplished with experimental manipulations (Zbrodoff, 1995).

In Experiment 1, the problem-size effect for simple addition, when the problems were solved by means of direct retrieval, is examined for kindergarten and elementary-school (grades 1 through 3) children from mainland China and the United States. The data described in this experiment are part of a larger cross-national study of the influence of age, language, and schooling on the arithmetical development of Chinese and American elementary-school children (Geary, Bow-Thomas, Fan, & Siegler, in press). The present report focuses exclusively on the problem-size effect, which is tangential to the goals of the primary study and, therefore, was not addressed in the analyses presented in Geary et al. (in press). In a second experiment, which is unique to this article, the problem-size effect is examined for Chinese and American college students.

EXPERIMENT 1

There are two features of this study that are of relevance to the issue of the nature and development of the problem-size effect in mental addition. First, Chinese children show nearly 100% retrieval by the end of first grade, and 100% retrieval by third grade. The high frequency of retrieval, combined with very fast solution times and low error rates, suggest that the Chinese children have had extensive exposure to the full range of simple addition problems over an extended period of time (i.e. several years). Second, data were collected at the beginning and at the end of the school year and across grade levels in both countries. If the problem-size effect is primarily related to exposure to arithmetic in school, then the magnitude of the effect should systematically decrease for both the Chinese and American children across the school year and across grade levels. As noted earlier, after extensive exposure to addition combinations (i.e. for the Chinese children), Siegler's model and a variation of the multiple-code model that allows for a dissociation between physical codes and magnitude representations predicts uniform access to stored facts, that is, no problem-size effect. In contrast, a multiple-code representation that assumes automatic access to magnitude representations, and that assumes that the interfering effects of competing associations vary across problem size, predicts that even with such extensive exposure to simple addition problems, solution times should still be faster and less error prone for smaller-valued problems relative to larger-valued problems.

Method

Subjects

The subjects included 105 elementary-school children (50 males, 55 females) from Columbia, Missouri, and 104 elementary-school children (54 males, 50 females) from Hangzhou, China. Table 1 shows that the mean ages of the same-grade American and Chinese children were comparable at all grade levels. The difference in the number of boys and girls in the American sample was not significant across grade levels, $\chi^2(3) = 5.4$, $p > 0.10$, nor was the national difference in the numbers of boys and girls, $\chi^2(1) = 0.4$, $p > 0.50$. 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). Here, an individual who was familiar with both Hangzhou and Columbia selected comparable areas, in terms of relative socioeconomic status, from which to choose subjects. In Hangzhou, subjects were selected from a single elementary school that served a working-class district. In Columbia, subjects were selected from two elementary schools that served a working-class population. In this first study, the relative performance of our Chinese and American children on a test of addition skills was comparable to that of larger and more representative samples of Chinese and American children for a similar test (Stevenson et al., 1990b). Thus, the schools used in our original study and the current study would appear to be adequately matched, in terms of providing a relatively unbiased assessment of the arithmetic skills of urban American and urban Chinese children.

TABLE 1
Subject Characteristics

	Age		*Gender*	
Grade	*M*	*SD*	*Male*	*Female*
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
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

Note: Age is given in months.

Experimental Tasks

All children were administered a paper-and-pencil test of addition skills, a digit span measure, and an addition strategy assessment. The 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). Descriptions of the paper-and-pencil addition test and the digit span task are provided in Geary et al. (in press).

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. For the second assessment, the kindergarten children were also administered a second set of nine larger-valued additions problems, that is, problems with sums greater than ten (e.g. 5+6, 7+8, 7+6). Strategies and RTs for these larger-valued problems are not considered here, because the American children only retrieved, on average, answers for 17% of these problems and committed many retrieval errors (47%). As a result, there were not enough correct retrieval trials to provide a meaningful estimate of the problem-size effect for this problem set. For the older children, the stimuli consisted of 40 single-digit addition problems, which ranged in difficulty from 2+3 to 9+8 (no tie problems, such as 3+3 or 4+4, were included).

The problems were presented, one at a time, at the centre of a cathode-ray tube (CRT) controlled by a microcomputer. For each problem, a prompt appeared at the centre of the CRT for 1000 msec, followed by a blank screen for 1000 msec. The problem then appeared on the screen and remained until the child responded. The child responded by speaking the answer into a voice-operated relay that was interfaced with the microcomputer. A hardware clocking mechanism ensured the collection of RTs with an accuracy of about ±1 msec. Equal emphasis was placed on speed and accuracy of responding.

Procedure

School begins during the first week of September in China and in 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 assessed at both times of measurement, and 103 of the 105 American subjects were assessed at both times of measurement (two American subjects moved between the first and second assessment). 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 at their own school. For the first time of measurement, all subjects were first administered the digit span measure, then the paper-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 subjects were administered the paper-and-pencil addition test followed by the addition-strategy assessment.

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.[1] Strategy classifications were, for the most part, based on the child's behaviour 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 sub-auditory vocalisations were present (e.g. lip movements), then the strategy was classified as verbal counting. If the child spoke the answer without counting on fingers or counting verbally, 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+5 = 7+3 = 10, 10+2 = 12) then the strategy was classified as decomposition. 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. As described in the Results section, RT patterns differed across the classified strategies, providing further evidence for the validity of the strategy classifications.

Results

In the first section below, a brief overview of the strategy distributions and analyses of retrieval frequencies and error rates as related to problem size is presented. Detailed descriptions and analyses of the Chinese and American children's overall strategy choices are provided in Geary et al. (in press). The second section presents analyses of the relation between retrieval RTs and problem size, and the third, and final, section presents an assessment of the magnitude of the problem-size effect for individual subjects.

[1]The fingers strategy involves looking at uplifted fingers, which represent the numbers to be added, but not counting them to get the answer. Under these circumstances, the use of fingers appears to prompt direct retrieval (Siegler & Shrager, 1984). The finger counting and verbal counting strategies were also classified (based on observation and subject report) based on the procedure used in counting. These procedures typically involved sum counting (i.e. counting both addends) or min counting (i.e. counting the smaller addend) (see Geary et al., in press).

Strategy choices

Overviews of the children's strategy choices are presented in Tables 2–5 for the kindergarten, first-, second-, and third-graders, respectively. Inspection of the two right-hand columns of each of the tables indicates that for both the Chinese and American children, mean RTs were lowest for retrieval and highest for counting fingers, with intermediate values for decomposition and verbal counting. This pattern is consistent with previous studies of children's strategy choices and provides support for the validity of the strategy classifications (e.g. Geary & Brown, 1991; Siegler, 1987).

The analyses of the retrieval frequencies and retrieval errors are presented separately for the kindergarten and older children, because they were administered different sets of addition problems. Following these analyses, correlations, computed across problems, between error frequencies and variables that are sensitive to the problem-size effect are presented.

Kindergarten. For the kindergarten children, the analyses of retrieval frequency and error rates were by means of 2 (Nation) × 2 (Time) mixed analyses of variance (ANOVAs), with nation as a between-subjects factor and time as a within-subjects factor. For retrieval frequency, the results revealed a non-significant main effect of nation, $F(1, 49) < 1$, $p > 0.25$, but significant time effects, $F(1, 49) = 28.07$, $p < 0.0001$, and Nation × Time effects, $F(1, 49) = 36.67$, $p < 0.0001$. The significant cross-over interaction confirmed that at Time 1 the American kindergarten children used retrieval more

TABLE 2
Characteristics of Addition Strategies in Kindergarten

	Mean % of Trials		*Mean % of Errors*		*Mean RT*[a]	
Strategy	*China*	*US*	*China*	*US*	*China*	*US*
Time 1						
counting fingers	11	29	8	13	9.3	7.6
fingers	11	0	9	—	5.8	—
verbal counting	47	12	5	13	3.2	4.6
retrieval	31	59	1	33	1.6	2.8
Time 2						
counting fingers	0	32	—	13	—	8.9
fingers	0	3	—	3	—	5.4
verbal counting	16	11	2	7	3.7	5.3
retrieval	84	54	1	8	1.5	2.6

[a]Mean reaction times in seconds; excludes error and spoiled trials as well as outliers.

TABLE 3
Characteristics of Addition Strategies in First Grade

	Mean % of Trials[a]		*Mean % of Errors*		*Mean RT*[b]	
Strategy	*China*	*US*	*China*	*US*	*China*	*US*
Time 1						
counting fingers	0	34	—	21	—	8.1
fingers	0	2	—	15	—	—
verbal counting	18	42	1	11	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	—	8.5
fingers	0	0	—	—	—	—
verbal counting	3	46	0	7	2.4	4.5
retrieval	91	28	2	12	1.2	2.7
decomposition	6	4	1	14	2.2	4.9

[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]Mean reaction in times in seconds; excludes error and spoiled trials as well as outliers. For the US sample, Time 1 mean RTs are not reported for fingers and decomposition due to the small number of trials for these strategies.

TABLE 4
Characteristics of Addition Strategies in Second Grade

	Mean % of Trials		*Mean % of Errors*		*Mean RT*[a]	
Strategy	*China*	*US*	*China*	*US*	*China*	*US*
Time 1						
counting fingers	0	35	—	10	—	5.7
fingers	0	0	—	—	—	—
verbal counting	2	33	0	6	—	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	—	5.7
fingers	0	0	—	—	—	—
verbal counting	1	34	0	6	—	4.1
retrieval	98	41	4	4	1.0	2.1
decomposition	1	1	7	0	—	—

[a]Mean reaction times in seconds; excludes error and spoiled trials as well as outliers. For the US 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.

TABLE 5
Characteristics of Addition Strategies in Third Grade

Strategy	Mean % of Trials China	Mean % of Trials US	Mean % of Errors China	Mean % of Errors US	Mean RT[a] China	Mean RT[a] US
Time 1						
counting fingers	0	23	—	6	—	4.7
fingers	0	0	—	—	—	—
verbal counting	0	28	—	2	—	3.2
retrieval	100	45	5	0	0.8	2.0
decomposition	0	4	—	13	—	3.0
Time 2						
counting fingers	0	15	—	7	—	4.8
fingers	0	0	—	—	—	—
verbal counting	0	24	—	8	—	3.2
retrieval	100	56	3	2	0.8	1.9
decomposition	0	4	—	0	—	—

[a]Mean reaction times in seconds; excludes error and spoiled trials as well as outliers.

frequently than their Chinese peers, $F(1, 49) = 11.85$, $p < 0.005$, whereas at Time 2 the Chinese kindergarten children used retrieval more frequently than their American peers, $F(1, 49) = 25.96$, $p < 0.0001$.

The analysis of retrieval errors confirmed that the Chinese kindergarten children committed fewer errors than the American kindergarten children, $F(1, 49) = 18.92$, $p < 0.0001$, and that the overall frequency of retrieval errors decreased from Time 1 to Time 2, $F(1, 49) = 9.15$, $p < 0.005$. Both of these main effects were qualified, however, by a significant Nation × Time interaction, $F(1, 49) = 9.15$, $p < 0.005$. The interaction reflected a significant decrease in retrieval errors across time of measurement for the American children, $t(22) = -3.01$, $p < 0.01$, and, due to very few retrieval errors (i.e. a floor effect), no change across time of measurement for the Chinese children, $t(25) = -0.87$, $p > 0.25$.

First through Third Grade. For the older children, the analyses of retrieval frequency and error rates were by means of 3 (Grade) × 2 (Nation) × 2 (Time) mixed ANOVAs, with grade and nation as between-subjects factors and time as a within-subjects factor. For retrieval frequency, the results revealed that, except for the Grade × Nation interaction, $F(2, 149) = 2.62$, $p > 0.05$, all of the main effects and interactions were significant, ps < 0.005; $F(2, 149) = 34.97$, $F(1, 149) = 278.61$, $F(1, 149) = 87.94$, $F(2, 149) = 25.25$, $F(1, 149) = 8.71$, $F(2, 149) = 36.45$ for the grade, nation, time, Grade × Time, Nation × Time, and Grade × Nation × Time effects, respectively. Overall, retrieval frequency was higher for the Chinese children than the American children and increased across grade level and time of measurement.

The three-way interaction reflected a significant increase in retrieval frequency across time of measurement for the Chinese first-graders, $t(25) = 13.04$, $p < 0.0001$, and no increase in retrieval frequency for the American first-graders, $t(26) = 1.37$, $p > 0.10$, combined with the opposite pattern for the third-graders. For the third-graders, the increase in retrieval frequency was significant for the American children, $t(31) = 3.05$, $p < 0.005$, but, due to 100% retrieval (i.e. a ceiling effect) at both times of measurement, was not significant for the Chinese children, $t(25) = 1.44$, $p > 0.10$. Both the Chinese, $t(25) = 3.30$, $p < 0.005$, and the American, $t(17) = 2.07$, $p = 0.054$, second-graders showed more retrieval at Time 2 than at Time 1.

The analysis of error patterns indicated that overall error rates did not differ across grade level, $F(2, 149) < 1$, or time of measurement, $F(1, 149) = 2.25$, $p > 0.10$, but the Chinese children committed significantly fewer retrieval errors than their American peers, $F(1, 149) = 5.53$, $p < 0.02$. The Grade × Time, $F(2, 149) < 1$, and Nation × Time, $F(1, 149) < 1$, interactions were not significant, $ps > 0.50$, but the Grade × Nation, $F(2, 149) = 10.42$, and Grade × Nation × Time, $F(2, 149) = 5.19$, interactions were, $ps < 0.01$. The three-way interaction was the result of a significant decrease in retrieval errors from Time 1 to Time 2 for the Chinese third-graders, $t(25) = -3.71$, $p < 0.001$, a significant increase (due to a floor effect at Time 1) in retrieval errors from Time 1 to Time 2 for the American third-graders, $t(29) = 2.05$, $p < 0.05$, combined with no statistically significant changes in the frequency of retrieval errors across time of measurement for the Chinese or American first- and second-graders, $ps > 0.25$.

Retrieval Errors and Problem Size. Table 6 presents the correlations between the frequency of errors committed (across subjects) for each problem and variables that index the size of the problem, that is, variables that have been shown to be sensitive to the problem-size effect in previous research (e.g. Ashcraft & Battaglia, 1978; Campbell & Graham, 1985; Geary et al., 1986; Groen & Parkman, 1972; Siegler & Shrager, 1984). Associative strength (AS) is $1 - p$(correct retrieval), that is, 1 minus the probability that a child will correctly retrieve the answer when the child is not allowed to use back-up strategies (from Siegler, 1986). A positive correlation between AS and error frequencies indicates agreement between the probability of incorrect retrieval, from Siegler's research, and the frequency of retrieval errors in the current study. Min refers to the smaller of the two addends (e.g. 3, in 5+3), while max refers to the larger of the two addends. The three remaining variables are defined as follows: sum = a+b; product = a×b; $\text{sum}^2 = (a+b)^2$.

Consistent with previous studies of North American individuals (Campbell & Graham, 1985; Miller et al., 1984; Siegler & Shrager, 1984), the top section

TABLE 6
Correlations between Frequency of Retrieval Errors and Problem Size Variables

	Grade							
	Kindergarten		*First*		*Second*		*Third*	
Variable	*Time 1*	*Time 2*	*Time 1*	*Time 2*	*Time 1*	*Time 2*	*Time 1*	*Time 2*
United States								
AS	0.24	0.25	0.59*	0.40***	−0.19	0.49**	—	0.23
min	0.33	0.32	0.75*	0.55*	−0.07	0.59*	—	0.18
max	0.41***	0.34	0.37***	0.11	−0.28	0.16	—	0.27
sum	0.44***	0.39****	0.65*	0.39***	−0.20	0.44**	—	0.27
product	0.39****	0.37****	0.70*	0.44**	−0.17	0.50*	—	0.24
sum^2	0.39****	0.36****	0.65*	0.38***	−0.21	0.44**	—	0.26
China								
AS	−0.07	−0.17	−0.11	0.05	−0.29****	−0.15	0.11	0.18
min	0.04	−0.06	−0.14	−0.01	−0.40***	−0.15	0.03	0.07
max	0.27	0.27	−0.05	0.01	−0.27	−0.23	0.04	0.14
sum	0.18	0.12	−0.11	0.00	−0.39***	−0.23	0.04	0.12
product	0.12	0.02	−0.14	−0.01	−0.39***	−0.19	0.05	0.08
sum^2	0.15	0.07	−0.13	−0.01	−0.39***	−0.21	0.04	0.08

Note: AS = associative strength. Min refers to the smaller of the two addends and max refers to the larger of the two addends.
$*p < 0.001$; $**p < 0.01$; $***p < 0.05$; $****p < 0.07$.

of Table 6 shows that for the American children the frequency of retrieval errors increased as the size of the problem increased. The only exception to this pattern was for the Time 1 assessment of the second-graders, where all of the correlations were negative, but non-significant. Inspection of the bottom section of Table 6 indicates a different pattern for the Chinese children. Here, there was no significant relationship between the indicators of problem size and the frequency of retrieval errors, except for the Time 1 assessment of the second-graders. For the Chinese second-graders, at Time 1, the frequency of retrieval errors tended to decrease as problem size increased.

Error Patterns. Campbell (Campbell & Graham, 1985; Campbell, 1995) has noted that retrieval errors are not random but, in addition to varying with the size of the problem, reveal a pattern showing that such errors are typically correct answers to related problems, presumably the result of multiple associations between the addends and potential answers. For addition, the most common class of retrieval error involves retrieving an answer that is ±1 or ±2 from the correct sum. Campbell (1995) referred to such errors as table-related because they almost always involve retrieving an answer that is correct for another addition problem, that is, an answer that is correct for one or more problems in the addition table, but incorrect for the presented problem (such as 9+2=10). Two other common sources of retrieval errors for addition include operation confusions (e.g. 9+2=18) and naming errors (Campbell, 1995; Miller & Paredes, 1990). Naming errors involve restating one of the addends as the answer (e.g. 9+2=9) or stating the augend as the tens value of the answer and the addend as the units value of the answer (e.g. 9+2=92).

Retrieval errors for the first-, second-, and third-graders were classified into the above three categories plus a fourth category, "other", which included all errors that were not classifiable into the first three categories; error patterns were not examined for the kindergarten children, because of the low rate of retrieval errors for the Chinese kindergarten children. Table 7 presents a summary of the resulting categorisations. Consistent with the error patterns found for adults (Campbell, 1995), the majority of the children's retrieval errors were table-related across all three grade levels and for both the American and Chinese children. Operation (typically confusing multiplication with addition) and naming errors were found for both the American and Chinese children, although operation errors occurred at an earlier grade in the Chinese sample than in the US sample, and naming errors were more common for the Chinese children than for the American children.

Reaction Times

Two sets of analyses of retrieval RTs were conducted. The first set of analyses used an estimate of the problem-size effect for individual subjects as the dependent measure. The second set of analyses were based on correlations

TABLE 7
Percentages of Error Types

	Grade					
	First		*Second*		*Third*	
Type	*Time 1*	*Time 2*	*Time 1*	*Time 2*	*Time 1*	*Time 2*
United States						
table-related	67	61	100	78	—	71
operation	0	0	0	0	—	21
naming	5	0	0	0	—	0
other	27	39	0	22	—	7
China						
table-related	67	61	80	82	83	77
operation	17	0	0	13	9	0
naming	8	17	2	5	8	8
other	8	22	18	0	0	15

between mean retrieval RTs (averaged across subjects) and the problem-size variables described above (e.g. AS, min, max, etc). All analyses focused on correct retrieval trials, but excluded outliers. Outliers were defined as any RT less than 500 msec, or any RT ±2.5 *SD*s from the mean of the correct retrieval RTs for each subject. In all, 6.3% of the correct retrieval RTs were identified and eliminated as outliers.

Kindergarten. In order to assess the problem-size effect for individual subjects, the 25 addition problems administered to the kindergarten children were divided into smaller- (n = 10) and larger- (n = 15) valued problems. Smaller-valued problems were those with sums less than or equal to five, while larger-valued problems had sums between six and ten. The problem-size effect can be represented by the difference in mean RTs for smaller- and larger-valued problems (Ashcraft, 1992). The use of smaller- and larger-valued sums seemed to be a straightforward method to obtain estimates of the problem-size effect for individual subjects and is empirically justified based on the correlations between the sum variable and mean retrieval RTs described below (see Table 9).

Mean RTs for smaller- and larger-valued problems across nation, grade, and time of measurement are shown in Table 8 and, for the kindergarten children, were analysed by means of a 2 (Nation) × 2 (Time) × 2 (Size) mixed ANOVA, with nation as a between-subjects factor and time and size as within-subjects factors. The results confirmed that the Chinese kindergarten children had faster overall retrieval times than their American peers, $F(1, 41) = 72.23$, $p < 0.0001$, and that smaller-valued problems were solved more quickly than larger-valued problems, $F(1, 41) = 41.76$, $p < 0.0001$. A significant Nation × Size interaction, however, indicated that the overall problem-size effect was smaller for the Chinese children in comparison to the American children, $F(1, 41) = 7.06$, $p < 0.02$. The main effect for time and all interactions involving time were non-significant, ps > 0.05; $F(1, 41) = 2.91$ for the time effect, $F(1, 41) = 2.81$ for the Time × Size effect, and $F(1, 41) < 1$ for the Nation × Time and the Nation × Time × Size effects.

The difference between the means for smaller- and larger-valued problems (i.e. the mean of the difference between smaller- and larger-valued problems calculated for each subject) are displayed in Fig. 1; note that these values are not identical to the difference between the means presented in Table 8 (which would be based on a difference between group means, not individual estimates). Values greater than zero indicate longer mean RTs for larger-valued than smaller-valued problems, that is, a positive problem-size effect. In keeping with previous studies of the relation between RTs and problem size (e.g. Ashcraft & Battaglia, 1978), the left section of Fig. 1 shows positive and significant, ts > 3.00, ps < 0.005, problem-size effects for both the American and Chinese kindergarten children at both times of measurement.

TABLE 8
Mean RTs for Smaller- and Larger-Valued Problems across Grade and Nation

	Kindergarten				*First*				*Second*				*Third*			
	Small		*Large*		*Small*		*Large*		*Small*		*Large*		*Small*		*Large*	
Measurement	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
United States																
Time 1	2488	1054	3245	1195	3489	900	3858	1471	2459	600	3159	977	1893	532	2278	795
Time 2	2417	682	2868	749	2556	864	3095	1057	2063	477	2367	693	1718	514	2104	656
China																
Time 1	1432	159	1805	551	1414	301	1603	221	1048	194	1077	227	819	116	847	162
Time 2	1439	193	1560	264	1160	173	1145	179	975	136	971	168	835	115	808	119

Note: Tabled values in milliseconds.

First through Third Grade. In order to assess the problem-size effect in the older children, the 40 addition problems administered to these children were also divided into smaller- (n = 19) and larger- (n = 21) valued problems. Smaller-valued problems were those with sums less than or equal to ten, while larger-valued problems were those with sums greater than ten. Again, the problem-size effect can be estimated by the difference in mean RTs comparing smaller- and larger-valued problems; the mean retrieval RTs across the smaller- and larger-valued problems are shown in Table 8.

For the older children, retrieval RTs were analysed by means of a 3 (Grade) × 2 (Nation) × 2 (Time) × 2 (Size) mixed ANOVA, with nation and grade as between-subjects factors and time and size as within-subjects factors. The results confirmed that the Chinese children had faster overall RTs than the American children, $F(1, 113) = 340.44$, $p < 0.0001$, and that overall RTs decreased across grade level, $F(2, 113) = 34.09$, $p < 0.0001$, and time of measurement, $F(1, 113) = 111.88$, $p < 0.0001$. A significant main effect for size indicated that smaller-valued problems were solved more quickly than larger-valued problems, $F(1, 113) = 78.00$, $p < 0.0001$. The significant main effects were qualified, however, by significant Grade × Nation, $F(2, 113) = 5.52$,

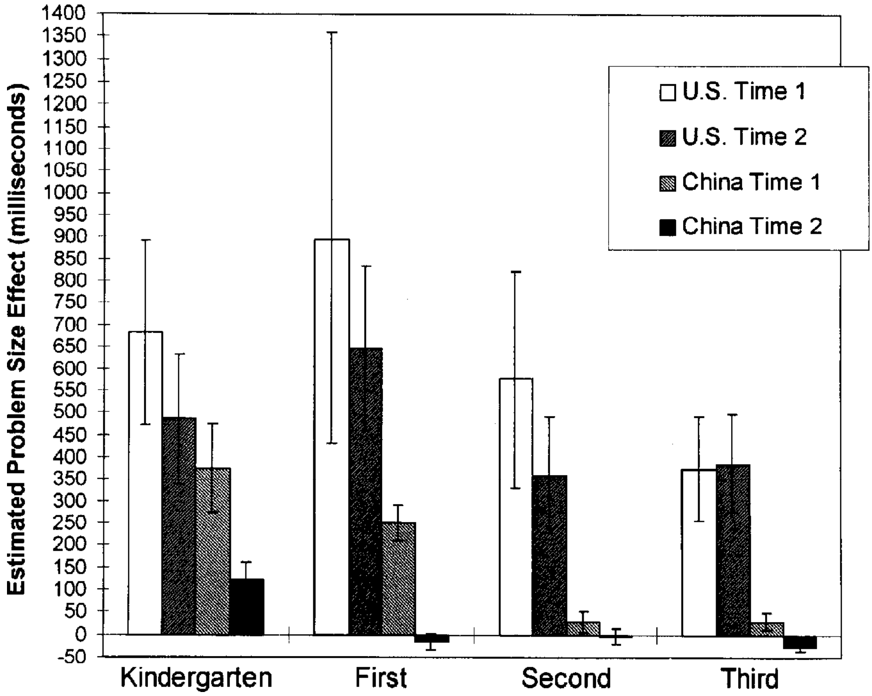

FIG. 1. Estimated problem-size effects across grade level, time of measurement, and nation. The error bars represent standard errors.

$p < 0.01$, Grade × Time, $F(2, 113) = 16.67$, $p < 0.0001$, Nation × Time, $F(1, 113) = 36.88$, $p < 0.0001$, Grade × Nation × Time, $F(2, 113) = 5.57$, $p < 0.005$, Grade × Size, $F(2, 113) = 5.99$, $p < 0.005$, Nation × Size, $F(1, 113) = 57.07$, $p < 0.0001$, and Time × Size, $F(1, 113) = 5.05$, $p < 0.05$, interactions. None of the remaining three-way interactions nor the four-way interaction were significant, ps > 0.10.

The significant interactions reflect the disappearance and reversal of the problem-size effect in the Chinese children, as shown in Fig. 1. For the American children, the overall (across time of measurement) problem-size effect was significant at all three grade levels, ps < 0.05. The problem-size effect was also significant, ts > 2.00, ps < 0.05, at each time of measurement, except for the Time 1 assessment of the first-graders, which, due to a large standard error, was only marginally significant, $t(7) = 1.93$, $p < 0.10$. For the Chinese children, in contrast, the problem-size effect was only significant at Time 1 in first grade, $t(22) = 6.30$, $p < 0.001$, and at Time 2 in third grade, $t(25) = -2.60$, $p < 0.05$. However, for the Chinese third-graders the problem-size effect was reversed. In other words, larger-valued problems were solved significantly faster than smaller-valued problems. For the Chinese children, the problem-size effect was not significant at Time 2 in first grade, at either time of measurement in second grade, or at Time 1 in third grade (ps > 0.10).

Retrieval RTs and Problem Size. Table 9 presents the correlations between mean retrieval RTs and the problem-size variables. Inspection of the top section of Table 9 shows that for the American children retrieval RTs were positively correlated with all of the problem-size variables, although some of the correlations do not differ significantly from zero. Thus, for the American subjects, retrieval RTs increased as the size of the problem addends increased, in keeping with the findings of many other studies (e.g. Ashcraft & Battaglia, 1978). The correlations presented in the bottom section of Table 9 show the same pattern for the Chinese kindergarten children and the Chinese first-graders at Time 1. However, there is no relation between mean retrieval RTs and any of the problem-size variables for the Chinese first-graders at Time 2, the Chinese second-graders, or the Chinese third-graders at Time 1. Moreover, in keeping with the pattern shown in Fig. 1, for the Chinese third-graders, the Time 2 retrieval RTs are *negatively* correlated with the problem-size variables.

The finding of a negative correlation between mean RTs and the problem-size variables for the Time 2 measurement of the Chinese third-graders does not appear to be due to a speed–accuracy trade-off, because error frequencies and mean RTs were only marginally correlated, $r(38) = 0.28$, $p = 0.08$. Moreover, the correlation was positive, indicating that error frequencies increased with increases in RTs, not decreased as would be expected with a speed–accuracy trade-off.[2]

TABLE 9
Correlations between Mean Retrieval RTs and Problem Size Variables

	Grade							
	Kindergarten		*First*		*Second*		*Third*	
Variable	*Time 1*	*Time 2*	*Time 1*	*Time 2*	*Time 1*	*Time 2*	*Time 1*	*Time 2*
United States								
AS	0.66*	0.28	0.24	0.61*	0.16	0.31***	0.26	0.45**
min	0.49***	0.46***	0.36***	0.60*	0.24	0.23	0.19	0.40**
max	0.21	0.56**	0.25	0.43**	0.10	0.38***	0.37***	0.48**
sum	0.41****	0.60**	0.37***	0.61*	0.20	0.36***	0.33***	0.52*
product	0.46***	0.55**	0.38***	0.64*	0.18	0.32***	0.30****	0.48**
sum^2	0.40****	0.55**	0.35***	0.62*	0.16	0.35***	0.34***	0.50*
China								
AS	0.64**	0.12	0.51**	−0.08	0.13	−0.10	0.11	−0.20
min	0.68*	0.26	0.42**	−0.21	0.04	−0.25	–0.03	−0.28****
max	0.43***	0.58**	0.43**	−0.10	0.04	−0.18	0.23	−0.36***
sum	0.64**	0.49***	0.50**	−0.18	0.04	−0.25	0.11	−0.38***
product	0.70*	0.39****	0.46**	−0.17	0.07	−0.22	0.09	−0.29****
sum^2	0.66*	0.43***	0.47**	−0.17	0.05	−0.23	0.13	−0.33***

Note: AS = associative strength. Min is the smaller of the two addends and max is the larger of the two addends.
*$p < 0.001$; **$p < 0.01$; ***$p < 0.05$; ****$p < 0.08$.

Assessment of Individual Problem Size Effects

In this section, the magnitude of the problem-size effect is examined for individual subjects. Specifically, subjects were categorised based on the magnitude of the problem-size effect for their retrieval RT data. The first category (< 0) included subjects with negative problem-size effects, that is, subjects who had shorter mean RTs for larger-valued problems than for smaller-valued problems. The three remaining categories included subjects with positive problem-size effects of various magnitudes (between 0 and 500 msec, between 501 and 1000 msec, and above 1000 msec), that is, subjects who had shorter mean RTs to smaller-valued problems than for larger-valued problems. The category "missing" refers to the number of subjects for whom a problem-size effect could not be estimated, those who did not have correct retrieval RTs for both smaller- and larger-valued problems. Table 10 shows the distribution of subjects across the four problem-size categories.

[2] There were no significant negative correlations between mean RTs and error frequencies, $ps > 0.05$, indicating that there were no speed–accuracy trade-offs for any of the groups.

Chi-square tests indicated that the distribution of subjects across the four categories differed significantly across the Chinese and American children at all grade levels and both times of measurement, *p*s < 0.01, except for Time 1 in kindergarten, $p > 0.10$. More important, examination of the bottom section of Table 10 indicates that in kindergarten and for the first assessment in first grade, the majority of the Chinese children showed positive problem-size effects. However, by the second assessment in first grade, at least half of the Chinese children showed negative problem-size effects, and at Time 2 in third grade the majority (69%) of the Chinese children showed negative problem-size effects. Finally, the magnitude of the problem-size effect was not correlated with error frequencies for any of the assessments *p*s > 0.15. In other words, Chinese children with negative problem-size effects were no more error-prone than their peers with positive problem-size effects, suggesting that the negative problem-size effects were not due to speed–accuracy trade-offs.

Discussion

The relation between error frequencies, retrieval RTs, and problem size found for the American children, except for the Time 1 assessment of the second-graders, mirrored the standard pattern found in other samples of North

TABLE 10
Magnitude of the Problem Size Effect for Individual Subjects

	Magnitude of the Problem Size Effect[a]									
	< 0		*0–500*		*501–1000*		*> 1000*		*Missing*	
Grade	*Time 1*	*Time 2*	*Time 1*	*Time 2*	*Time 1*	*Time 2*	*Time 1*	*Time 2*	*Time 1*	*Time 2*
United States										
kindergarten	3	7	5	4	2	8	7	3	8	3
first	2	3	2	5	1	2	3	5	21	14
second	4	2	0	4	3	4	4	1	8	8
third	5	6	12	13	5	6	4	5	6	2
China										
kindergarten	4	6	16	19	3	1	3	0	0	0
first	1	15	20	11	2	0	0	0	3	0
second	14	15	12	11	0	0	0	0	0	0
third	13	18	13	8	0	0	0	0	0	0

Note: The tabled values represent the number of subjects falling in each category. Missing refers to subjects for whom a problem-size effect could not be calculated, that is, for subjects who did not have correct retrieval RTs for both smaller- and larger-valued problems.

[a]In milliseconds.

American subjects (e.g. Ashcraft & Battaglia, 1978; Campbell, 1995; LeFevre et al., in press; Miller et al., 1984). Specifically, in keeping with nearly all, if not all, previous studies, error frequencies and retrieval RTs increased as the size of the problem addends increased—the standard problem-size effect. Also in keeping with previous studies, the majority of the American and Chinese children's retrieval errors were table-related (i.e. ±1 or ±2 from the correct sum; Campbell, 1995). These error patterns are consistent with the argument that the problem addends have multiple associations to a number of addition facts, including the correct sum, and that these multiple associations can interfere with fact retrieval (Campbell, 1995; Campbell & Clark, 1988; Siegler & Shrager, 1984).

As noted earlier, Campbell (1995) has argued that the number of potentially interfering associations increases as the size of the addends increases and, therefore, is one factor contributing to the problem-size effect. However, unlike the American children, the frequency of retrieval errors did not increase with increases in problem size for the Chinese children. The dissociation between error patterns and the relation between problem size and error frequencies suggests that, at least after extensive practice, competing associations do not necessarily contribute to the problem-size effect, or do not necessarily produce a positive problem-size effect (see Experiment 2). Moreover, the finding that the magnitude of the problem-size effect for retrieval RTs systematically disappeared from kindergarten to third grade, and was in fact reversed by the end of third grade, for the Chinese children does not support the view that fact retrieval is associated with the automatic access of the magnitude representations associated with the problem addends. In other words, the pattern shown in Experiment 1 suggests that extensive practice of basic addition combinations might lead to a dissociation between physical-code representations and any associated magnitude representations, such that fact retrieval in such individuals appears to be primarily governed by the answer/problem associations, which in turn appear to develop based on frequency of exposure (Ashcraft, 1992; Siegler, 1986).

Two results suggest that problem/answer associations and any associated problem-size effect are strongly influenced by the frequency of exposure. First, in addition to showing strong national differences, the magnitude of the problem-size effect decreased systematically across the school year, except for the American third-graders who showed no change from Time 1 to Time 2.[3] This pattern is consistent with the view that the problem-size effect is influenced by schooling, as contrasted with more biological factors, such as maturation (Kail, 1991).

[3]The American third-graders also showed little across-the-year improvement for performance on a pencil-and-paper test of addition skills, suggesting little exposure to basic addition for these children during third grade (Geary et al., in press).

Second, a tally of the frequency with which simple addition problems (the basic 100 problems, excluding 0+0) were presented in the Chinese first-, second-, and third-graders' workbooks (which are used to practise solving arithmetic problems) indicated that the frequency of problem presentation and the sum of the problem were significantly and positively correlated at all three grade levels, $r(97) = 0.55$, 0.32, and 0.49, respectively, $ps < 0.005$.[4] Moreover, the cumulative frequency, across grades 1 to 3, of problem presentation was strongly and positively correlated with the sum of the problem, $r(97) = 0.68$, $p < 0.0001$. This pattern is the exact opposite of the pattern found in American mathematics textbooks. For instance, Hamann and Ashcraft (1986) found that the cumulative frequency of problem presentation (kindergarten to third grade) was negatively correlated with the sum of the problem, $r = -0.82$. In other words, larger-valued problems are presented more frequently than smaller-valued problems in these Chinese workbooks, but are presented less frequently in American textbooks.

The emphasis on solving larger-valued problems in the Chinese workbooks might explain the reversal of the problem-size effect in third grade. In fact, the cumulative frequency of problem presentation in the Chinese workbooks was significantly correlated with mean retrieval RTs for the Chinese third-graders, at Time 2, $r(38) = -0.34$, $p < 0.05$, but was not correlated with mean retrieval RTs for the American third-graders (Time 2), $r(38) = 0.04$, $p > 0.50$. In contrast, the cumulative frequency of problem presentation in American mathematics textbooks was significantly correlated with the American third-graders' (Time 2) retrieval RTs, $r(38) = -0.49$, $p < 0.005$; the value for the Chinese third-graders was 0.40, $p < 0.05$. The negative correlations between the Chinese third-graders' retrieval RTs and problem frequency in the Chinese workbooks and American third-graders' retrieval RTs and problem frequency in American textbooks indicate that answers to more frequently presented problems were retrieved more quickly than were answers to less frequently presented problems, for both Chinese and American children.

In contrast to the Chinese emphasis on solving larger-valued problem in first to third grade, in Chinese kindergartens the emphasis is on solving addition problems with sums up to and including ten, that is, the smaller-valued

[4]For each grade level, the tallies were taken from two workbooks, one from the first semester of school and the other from the second semester. Tallies included all simple addition problems (e.g. 6+7) and all unit-value combinations for more complex problems, such as 46+23 (6+3 was coded for this problem). The correlations excluded the 0+0 combination because of the high frequency with which problems such as 50+40 and 30+90, were presented in the second- and third-grade workbooks. The emphasis on solving such problems distorted the relation between problem size and problem frequency. In fact, the simple 0+0 problem was presented only once across the second- and third-grade workbooks.

problems for the first- to third-graders. This emphasis in Chinese kindergartens on learning the basic facts associated with smaller-valued problems might explain the problem-size effect at the beginning of first grade, for the Chinese children, as well as the emphasis on solving larger-valued problems in first through third grade.

EXPERIMENT 2

The goal of Experiment 2 was to determine whether the problem-size effect "reappears" in Chinese adults (i.e. college students). A test of this possibility is important because it might be argued that the intense exposure to arithmetic that Chinese children receive, combined with the emphasis on solving larger-valued problems, biased the results of Experiment 1 by obscuring the potential influence of access to magnitude representations and competing associations on the problem-size effect (Campbell, 1995; Gallistel & Gelman, 1992). In fact, it could be argued that if frequency of exposure were the only factor influencing the problem-size effect, then the negative problem-size effect found for the Chinese third-graders (at Time 2) should have been larger than it actually was, given the strong emphasis on solving problems with larger-valued addends. Thus, it is possible that frequency of exposure and magnitude representation/ interference effects have relatively independent and additive effects, which, for the Chinese children, resulted in a net negative problem-size effect by the end of third grade. If so, then without continual exposure to basic addition problems, which is not part of the mathematics curriculum in Chinese universities, the problem-size effect might reemerge in adulthood.

Method

The subjects included 35 undergraduates (23 males, 12 females) from the University of Missouri at Columbia, and 26 undergraduates (13 males, 13 females) from Hangzhou University, Hangzhou, China. The American subjects were recruited from psychology courses and received partial credit for participating in the experiment. The Chinese subjects were recruited through the Psychology Department at Hangzhou University, and they received a small donation in their name to a general student fund for participating in the experiment. The mean age of the Chinese and American subjects was 20 years.

Experimental Task and Procedure

Using the same strategy-assessment procedures, all subjects were administered the same set of addition problems that was administered to the older children in Experiment 1.

Results

For ease of presentation, the results are presented in three sections. The first describes the strategy mix for the Chinese and American adults, the second presents an examination of error patterns for the Chinese adults, and the third presents analyses of the relation between problem size and retrieval RTs.

Strategy Choices

Table 11 shows that the Chinese adults retrieved answers to 100% of the simple addition problems. In contrast, the American adults retrieved answers to only 73% of the problems and had to rely on some form of back-up strategy to solve the remaining problems. The national difference in the frequency of retrieval was significant, $F(1, 59) = 38.36$, $p < 0.0001$, as was the difference in the percentage of retrieval errors, $F(1, 59) = 8.38$, $p < 0.01$.

For the American sample, the frequency with which back-up strategies were used for problem solving increased with increases in the sum of the problem, $r(38) = 0.79$, $p < 0.0001$. In fact, for these subjects, back-up strategies were used, on average, to solve 2.5 of the 19 smaller-valued addition problems, and 14 of the 21 larger-valued addition problems, $F(1, 38) = 111.53$, $p < 0.0001$. In other words, the American college students used counting and decomposition primarily to solve the larger-valued addition problems.

Error Patterns

Error patterns were not analysed for the American adults, due to their low frequency. For the Chinese adults, the frequency of retrieval errors was positively and significantly correlated with all of the problem-size variables (i.e. AS, min, max, etc.), $r(38)$s = 0.33 to 0.48, ps < 0.05. In keeping with the error patterns found in previous research, and for the Chinese and American

TABLE 11
Characteristics of Addition Strategies: Adults

	Mean % of Trials		*Mean % of Errors*		*Mean RT*[a]	
Strategy	*China*	*US*	*China*	*US*	*China*	*US*
verbal counting	0	9	—	11	—	1716
decomposition	0	17	—	6	—	1101
retrieval	100	73	4	1	605	739

Note: The columnar sum for the mean percentage of trials does not sum to 100 in the US sample, due to rounding.

[a]Reaction times in milliseconds; excludes error and outlier RTs.

children in Experiment 1, 85% of the Chinese adults' retrieval errors were table-related. An additional 7.5% were naming errors, and the final 7.5% were unclassified; there were no operation confusion errors.

Reaction Times

In keeping with the analyses for the older children in Experiment 1, the problem-size effect was determined by examining RT patterns for smaller- and larger-valued problems, which were defined in a manner identical to that described for the older children. Again, the analyses only included correct retrieval trials, after eliminating outliers (1.9% of correct retrieval trials).

The retrieval RTs were analysed by means of a 2 (Nation) × 2 (Size) mixed ANOVA, with nation as a between-subjects factor and size as a within-subjects factor. The results confirmed that the Chinese adults had faster overall RTs than their American peers, $F(1, 56) = 21.72$, $p < 0.0001$, and that smaller-valued problems were solved more quickly than larger-valued problems, $F(1, 56) = 32.15$, $p < 0.0001$. These main effects were qualified, however, by a significant Nation × Size interaction, $F(1, 56) = 13.14$, $p < 0.001$. Within-nation analyses indicated that the problem-size effect was significant for both the Chinese, $t(25) = 2.12$, $p < 0.05$, and American, $t(31) = 5.73$, $p < 0.0001$, subjects, but was about 4.5 times larger in the American sample (mean = 110 msec) than in the Chinese sample (mean = 24 msec), as shown in Fig. 2.

Zero-order correlations indicated a significant and positive relation between mean retrieval RTs and all of the problem-size variables for the American sample, $r(38)$s = 0.38 to 0.55, ps < 0.05. For the Chinese sample, mean retrieval RTs were significantly and positively correlated with the product, $r(38) = 0.35$, $p < 0.05$, and sum^2, $r(38) = 0.34$, $p < 0.05$, variables.[5] In order to obtain an additional estimate of the magnitude of the problem-size effect, the product variable was used to represent retrieval RTs. The resulting b coefficients of 1.02 and 2.33 for the Chinese and American samples, respectively, were significantly different from 0, ps < 0.05, and significantly different from each other, $t(38) = -2.98$, $p < 0.05$. Moreover, the b coefficient for the American sample did not differ significantly, $t(38) = -1.33$, $p > 0.10$, from the 3.2 estimate (i.e. the b coefficient for the product variable) obtained for a separate sample of American undergraduates who were administered the same problem set (Geary & Wiley, 1991). However, this 3.2 value was significantly higher than the 1.02 estimate obtained for the Chinese undergraduates, $t(38) = -4.97$, $p < 0.01$.

[5]The frequency of retrieval errors and mean retrieval RTs were not significantly correlated for the Chinese sample, $r(38) = 0.26$, $p > 0.10$, suggesting no speed–accuracy trade-off.

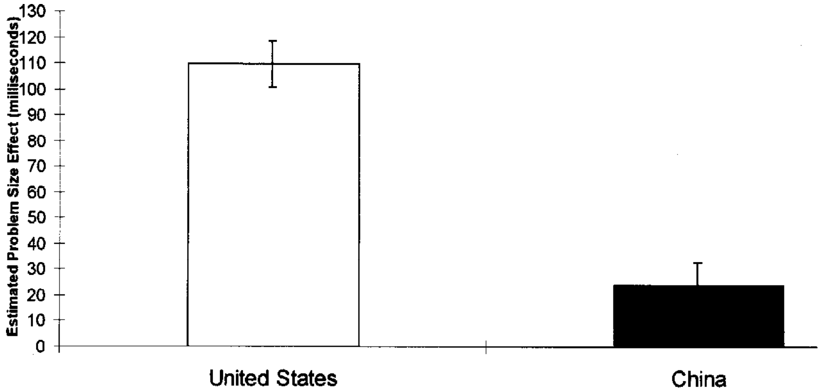

FIG. 2. Estimated problem-size effects for the American and Chinese adults. The error bars represent standard errors.

Discussion

In keeping with other studies of the mix of strategies used by North American college students to solve simple addition problems, the American students in this study used a combination of counting and decomposition to solve 27% of the presented problems (Geary & Wiley, 1991; LeFevre et al., in press). The finding that back-up strategies were used to solve, on average, 2 out of 3 (i.e. 14 of 21) of the larger-valued addition problems but only about 1 out of 9 of the smaller-valued addition problems suggests that the American subjects have had much more exposure to smaller-valued addition problems than to larger-valued addition problems, consistent with differences in the frequency with which smaller- and larger-valued problems are presented in American mathematics textbooks (e.g. Ashcraft & Christy, 1995; Hamann & Ashcraft, 1986). Within this context, the problem-size effect for retrieval for the American subjects is consistent with the argument that the problem-size effect is influenced by the difference in the frequency with which these subjects have been exposed to smaller- and larger-valued problems in school (Ashcraft, 1992; Siegler, 1986, 1988). Indeed, retrieval RTs for the American adults were significantly correlated with the cumulative frequency of problem presentation in kindergarten to third-grade mathematics textbooks (Hamann & Ashcraft, 1986), $r(38) = -0.51$, $p < 0.0001$.

If a simple frequency of exposure were the only factor influencing the problem-size effect, then it might be argued, based on the results of Experiment 1, that the Chinese adults should have had a negative problem-size effect. However, a significant and positive problem-size effect was found for the Chinese adults' retrieval errors and RTs, although the magnitude of the effect for the RTs was rather small (24 msec). Nevertheless, the Chinese adults showed the standard, though attenuated, problem-size effect that is almost always found in samples of North American subjects (e.g. Ashcraft & Battaglia, 1978; Campbell, 1995; Miller et al., 1984). The "reemergence" of the positive problem-size effect in Chinese adults suggests that frequency of exposure is not likely to be the only factor influencing the magnitude of the effect. In fact, retrieval RTs for the Chinese adults were not significantly correlated with frequency of problem exposure in the Chinese workbooks (from Experiment 1), $r(38) = 0.16$, $p > 0.25$. Thus, it appears that other factors related to the size of the addends in the problem, such as access to magnitude representations and/or interference, contribute to the problem-size effect (Campbell, 1995).

Indeed, based on the positive correlation between problem size and the frequency of retrieval errors, combined with the finding that most of these errors were table-related, suggests that the interfering effects of competing associations probably contributed to the problem-size effect for the Chinese adults (Campbell, 1995). In fact, when the frequency of retrieval errors is partialled from the relationship between retrieval RTs and the product variable, the magnitude of the problem-size effect is reduced ($b = 0.81$ vs. 1.02) but is still marginally significant, $t(38) = 1.69$, $p < 0.10$. Thus, if the frequency of retrieval errors provides an estimate of the interfering effects of multiple associations, then an additional influence on the problem-size effect, such as access to magnitude representations, is plausible.

GENERAL DISCUSSION

The results of these experiments suggest that the problem-size effect is influenced by several factors, including frequency of exposure and factors more directly associated with the size of the problem addends (Ashcraft, 1992; Campbell, 1995; Dehaene, 1992; Siegler, 1986). For children, the magnitude of the problem-size effect appears to be strongly influenced by the frequency of exposure to arithmetic in school (Hamann & Ashcraft, 1986). The finding of large national differences in the magnitude of the problem-size effect combined with the finding that the magnitude of the effect tended to diminish from the beginning to the end of the school year in both American and Chinese children supports the schooling argument. Furthermore, for both American and Chinese third-grade children (at Time 2), the speed of fact retrieval was directly related to the cumulative frequency with which the problems were presented, respectively, in American mathematics textbooks and Chinese workbooks. Specifically, answers to frequently presented problems were retrieved more quickly than answers to infrequently presented problems, regardless of whether the frequently presented problems consisted of larger- (Chinese workbooks) or smaller- (American textbooks) valued addends.

Finally, for the Chinese children, the finding of a positive problem-size effect at the beginning of kindergarten combined with the gradual disappearance and eventual reversal of the effect by the end of third grade also implicate schooling, and presumably the frequency of problem exposure in school, as an important influence on the magnitude of the problem-size effect. The reversal of the problem-size effect by the end of third grade suggests that extensive exposure to the basic addition combinations might lead to a dissociation between problem/answer units in the physical code and any underlying magnitude representations (Dehaene & Cohen, 1991). Alternatively, the influences of exposure, access to magnitude representations, and interference due to multiple associations might be independent. For the latter perspective, the speed and accuracy of fact retrieval would be influenced by all three factors (exposure, magnitude access, and interference), the effects of which would summate (Campbell, 1995).

From this point of view, increases in frequency of exposure would result in increases in the speed of fact retrieval (or unitization; Campbell, 1995), while the influence of magnitude representations would remain relatively constant. Exposure would influence the development of multiple and competing associations to the problem addends, but high levels of exposure to the correct problem/answer association would result in an eventual inhibition of competing associations (Campbell, 1995; Siegler, 1986; Zbrodoff, 1995). If so, then the reversed problem-size effect for the Chinese third-graders would reflect the net result of all of these influences, with a heavy weighting on the frequency of exposure. Indeed, if the magnitude of the problem-size effect were simply the result of the frequency of exposure, then it seems likely that the magnitude of the negative problem-size effect found in Chinese children at the end of third grade would have been much larger than the observed 27-msec effect.

This multiple-influence model of the problem-size effect is consistent with previous research (Campbell, 1995; LeFevre et al., 1995; Zbrodoff, 1995) and would also seem to accommodate the "reappearance" of the problem-size effect, for both RTs and error frequencies, in the Chinese adults. Although the magnitude of the effect for adults, at least for RTs, was smaller than is typically found in North American samples (Ashcraft & Battaglia, 1978; Campbell, 1995; Geary & Wiley, 1991; Miller et al., 1984), it was still the standard positive problem-size effect. More important, the failure to find a relation between the frequency of problem exposure in the Chinese workbooks and adult retrieval RTs suggests that the problem-size effect for retrieval errors and RTs for the Chinese adults is due largely to factors other than exposure. In other words, the attenuated problem-size effect in the Chinese adults might provide an estimate of the influence of access to magnitude representations and/or competing associations on the problem-size effect, above and beyond the influence of exposure.

If so, then the problem-size effect for the American adults in Experiment 2 was likely the result of all three factors, but the effect of exposure appears to have been largely eliminated for the Chinese adults. Given this, and based on the values presented in Figure 2, 78% {(110 − 24)/110} of the problem-size effect for the American adults might be attributed to exposure and the remaining 22% (24/110) attributed to other factors, such as the effect of competing associations. Alternatively, if the values of the regression coefficients for the product variable are used as an estimate of the problem-size effect, then 56% {(2.33 − 1.02)/2.33} of the problem-size effect in the American adults might be attributed to exposure, and 44% to other factors.

In closing, the current studies suggest that the standard problem-size effect in mental arithmetic (e.g. Ashcraft & Battaglia, 1978) is likely to be the result of multiple factors, including differences in the frequency of exposure to smaller- and larger-valued problems, the interfering effects of competing associations, and, possibly, the influence of access to magnitude representations (Ashcraft, 1992; Campbell, 1995; Gallistel & Gelman, 1992). The two latter factors probably contribute to the problem-size effect because the number of competing associations typically varies with problem size (Campbell, 1995), as does the precision of the magnitude representations (Moyer & Landauer, 1967). Regardless of these three influences, differences in the frequency of exposure to smaller- and larger-valued problems appear to be the primary contributors to the magnitude of the problem-size effect.

Manuscript received 31 March 1995
Revised manuscript received 31 August 1995

REFERENCES

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

Ashcraft, M.H. (1995). Cognitive psychology and simple arithmetic: A review and summary of new directions. *Mathematical Cognition, 1,* 3–34.

Ashcraft, M.H., & Battaglia, J. (1978). Cognitive arithmetic: Evidence for retrieval and decision processes in mental addition. *Journal of Experimental Psychology: Human Learning and Memory*, *4*, 527–538.

Ashcraft, M.H., & Christy, K.S. (1995). The frequency of arithmetic facts in elementary texts: Addition and multiplication in Grades 1–6. *Journal for Research in Mathematics Education*, *26*, 396–421.

Ashcraft, M.H., & Faust, M.W. (1994). Mathematics anxiety and mental arithmetic: An exploratory investigation. *Cognition and Emotion*, *8*, 97–125.

Ashcraft, M.H., Yamashita, T.S., & Aram, D.M. (1992). Mathematics performance in left and right brain-lesioned children. *Brain and Cognition*, *19*, 208–252.

Campbell, J.I.D. (1995). Mechanisms of simple addition and multiplication: A modified network-interference theory and simulation. *Mathematical Cognition*, *1*, 121–164.

Campbell, J.I.D., & Clark, J.M. (1988). An encoding-complex view of cognitive number processing: Comment on McCloskey, Sokol, and Goodman (1986). *Journal of Experimental Psychology: General*, *117*, 204–214.

Campbell, J.I.D., & Graham, D.J. (1985). Mental multiplication skill: Structure, process, and acquisition. *Canadian Journal of Psychology*, *39*, 338–366.

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition*, *44*, 1–42.

Dehaene, S., & Cohen, L. (1991). Two mental calculation systems: A case study of severe acalculia with preserved approximation. *Neuropsychologia* , *29*, 1045–1074.

Gallistel, C.R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition*, *44*, 43–74.

Geary, D.C. (1993). Mathematical disabilities: Cognitive, neuropsychological, and genetic components. *Psychological Bulletin*, *114*, 345–362.

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

Geary, D.C., Bow-Thomas, C.C., Fan, L., & Siegler, R.S. (in press). Development of arithmetical competencies in Chinese and American children: Influence of age, language, and schooling. *Child Development*.

Geary, D.C., & Brown, S.C. (1991). Cognitive addition: Strategy choice and speed-of-processing differences in gifted, normal, and mathematically disabled children. *Developmental Psychology*, *27*, 398–406.

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., Frensch, P.A., & Wiley, J.G. (1993). Simple and complex mental subtraction: Strategy choice and speed-of-processing differences in younger and older adults. *Psychology and Aging*, *8*, 242–256.

Geary, D.C., Widaman, K.F., & Little, T.D. (1986). Cognitive addition and multiplication: Evidence for a single memory network. *Memory & Cognition*, *14*, 478–487.

Geary, D.C., & Wiley, J.G. (1991). Cognitive addition: Strategy choice and speed-of-processing differences in young and elderly adults. *Psychology and Aging*, *6*, 474–483.

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

Hamann, M.S., & Ashcraft, M.H. (1986). Textbook presentations of the basic addition facts. *Cognition and Instruction*, *3*, 173–192.

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

LeFevre, J.A., Bisanz, J., & Mrkonjic, L. (1988). Cognitive arithmetic: Evidence for obligatory activation of arithmetic facts. *Memory & Cognition*, *16*, 45–53.

LeFevre, J.A., Sadesky, G.S., & Bisanz, J. (in press). Selection of procedures in mental addition: Reassessing the problem-size effect in adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

McCloskey, M. (1992). Cognitive mechanisms in numerical processing: Evidence from acquired dyscalculia. *Cognition*, *44*, 107–157.

Miller, K.F., & Paredes, D.R. (1990). Starting to add worse: Effects of learning to multiply on children's addition. *Cognition*, *37*, 213–242.

Miller, K., Perlmutter, M., & Keating, D. (1984). Cognitive arithmetic: Comparison of operations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*, 46–60.

Moyer, R.S., & Landauer, T.K. (1967). Time required for judgement of numerical inequality. *Nature*, *215*, 1519–1520.

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: Lawrence Erlbaum Associates, Inc.

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). Strategy choice procedures and the development of multiplication skill. *Journal of Experimental Psychology: General*, *117*, 258–275.

Siegler, R.S., & Jenkins, E.A. (1989). *How children discover new strategies*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

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: Lawrence Erlbaum Associates, Inc.

Stevenson, H.W., Lee, S.Y., Chen, C., Lummis, M., Stigler, J., Fan, L., & Ge, F. (1990a). 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. (1990b). Contexts of achievement: A study of American, Chinese, and Japanese children. *Monogaphs of the Society for Research in Child Development*, *55* (1–2, Serial No. 221).

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.

Widaman, K.F., Geary, D.C., Cormier, P., & Little, T.D. (1989). A componential model for mental addition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 898–919.

Zbrodoff, N.J. (1995). Why is 9 + 7 harder than 2 + 3? Strength and interference as explanations of the problem-size effect. *Memory & Cognition*, *23*, 689–700.