Developmental Psychology
1996, Vol. 32, No. 2, 254–262

0012-1649/96/$3.00

Are East Asian Versus American Differences in Arithmetical Ability a Recent Phenomenon?

David C. Geary
University of Missouri—Columbia

Timothy A. Salthouse
Georgia Institute of Technology

Guo-Peng Chen
East China Normal University

Liu Fan
China Central Institute for Educational Sciences

Younger and older American and Chinese adults were administered arithmetic, perceptual speed, and spatial orientation tests. For the perceptual speed and spatial orientation tests, the younger adults showed substantial performance advantages over the older adults in both the United States and China. For the arithmetic tests, the younger Chinese adults outperformed the older Chinese adults, but the groups of younger and older American adults had comparable arithmetical abilities. Cross-national comparisons indicated that the younger Chinese adults outperformed the younger American adults on the arithmetic tests, but not on the perceptual speed and spatial orientation tests. The performance of the older American and older Chinese adults was comparable for all of the ability measures. The overall pattern suggests that the advantage of Chinese adults over American adults in complex arithmetic might be a relatively recent phenomenon.

The first systematic cross-national study of mathematical abilities was conducted in 1964 (Husén, 1967). The results of this study showed that American adolescents were among the most poorly educated students in mathematics in the industrialized world. In the ensuing 30 years, this basic finding has been replicated many times and has been shown to be true for nearly all mathematical domains, from arithmetic to complex mathematics (Crosswhite, Dossey, Swafford, McKnight, & Cooney, 1985; Lapointe, Mead, & Askew, 1992). Differences between the mathematical development of American children and children from other nations are often times most extreme when comparisons are made between the United States and East Asian nations (i.e., China, Korea, Taiwan, & Japan; e.g., Stevenson, Chen, & Lee, 1993; Stevenson et al., 1990).[1] The consistency of academic achievement differences between children from the United States and children from East Asian nations has led some scientists to argue that these differences stem largely from a difference in the general intelligence of Asian and American individuals (e.g., Lynn, 1982; Rushton, 1992).

In this study, we present evidence that the advantage of East Asian individuals over their American peers in arithmetical abilities probably did not exist 60 years ago, and that the more recent East Asian advantage in arithmetic, and perhaps other mathematical domains, likely reflects secular changes within the United States rather than an East Asian advantage in intelligence. The hypothesis that the advantage of East Asian individuals over their American peers in arithmetic is a relatively recent phenomenon is based on two lines of research: cross-generational changes in the arithmetical development of American children and differences in the arithmetical abilities of younger and older American adults.

One dimension that distinguishes the arithmetical development of American children from their East Asian cohorts is the sophistication of the strategies used in problem solving (Fuson & Kwon, 1992a; Geary, Fan, & Bow-Thomas, 1992). For the solving of simple arithmetic problems, such as 8 + 7, when an answer cannot be remembered, American children tend to count mentally or on their fingers (Geary et al., 1992; Siegler, 1987). Children from East Asian nations, in contrast, typically use a form of decomposition, in which the digits are decomposed into smaller sets which, in turn, are added together (Fuson & Kwon, 1992a; Geary et al., 1992). For instance, to solve 8 + 7, the child might decompose the 7 into a 2 and a 5, and then retrieve the answers to 8 + 2, and 10 + 5. The use of decomposition appears to be dependent on a basic understanding of sets, and, as such, should probably be considered a conceptually more sophisticated problem-solving approach than counting (Fuson & Kwon, 1992a). East Asian children not only

David C. Geary, Department of Psychology, University of Missouri—Columbia; Timothy A. Salthouse, School of Psychology, Georgia Institute of Technology; Guo-Peng Chen, Department of Psychology, East China Normal University, Shanghai, China; Liu Fan, Department of Psychology, China Central Institute for Educational Sciences, Beijing, China.

The research was supported by Grant 1R01-HD27931 from the National Institute of Child Health and Human Development and Grant AG06826 from the National Institute on Aging.

We thank Melissa Moon and Ravindran Sabapathy for their assistance with data collection and collation, Zhitang Liu for his assistance with the translations, and Mark Ashcraft, Donald Kausler, and Robert Siegler for comments on earlier drafts of this article.

Correspondence concerning this article should be addressed to David C. Geary, Department of Psychology, 210 McAlester Hall, University of Missouri, Columbia, Missouri 65211.

[1] The mathematical achievement of children from China, Korea, Taiwan, and Japan is generally comparable (Lapointe et al., 1992; Stevenson et al., 1993).

use developmentally advanced problem-solving approaches in comparison with their contemporary American peers, they also memorize the basic arithmetic facts (e.g., the addition and subtraction tables) at a much younger age (e.g., Geary et al., 1992). In fact, recent studies suggest that only about half of contemporary North American college students appear to know all of the basic addition and subtraction facts (Geary, Frensch, & Wiley, 1993; Geary & Wiley, 1991; LeFevre, Sadesky, & Bisanz, in press).

Cross-national differences in the frequency with which decomposition is used as a problem-solving strategy are related, in part, to differences in the structure of Asian-language and English number words (Fuson & Kwon, 1992a). In Asian languages, words for teen values are stated ten one, ten two, ten three, and so on, which facilitates the child's conceptual understanding of the base-10 structure of the number system as well as strategy development (Fuson & Kwon, 1992b; Geary, 1994; Geary, Bow-Thomas, Fan, & Siegler, 1993; Miller, Smith, Zhu, & Zhang, 1995; Miller & Stigler, 1987; Miura, Okamoto, Kim, Steere, & Fayol, 1993). This does not appear to be the whole story, however.[2] Descriptive studies of the arithmetical development of American children who received their elementary-school education in the 1930s indicate that they frequently used decomposition as a problem-solving strategy and had typically mastered the basic arithmetic facts by the end of third grade (Gesell & Ilg, 1946; Ilg & Ames, 1951). In this respect, the arithmetical development of American children 60 years ago was very similar to developmental sequences found in contemporary East Asian children and greatly exceeded that of contemporary American children. The cross-generational changes in the arithmetical development of American children suggest that the advantage of contemporary East Asian children over their American peers in arithmetic reflects decreased arithmetical abilities in American children and not an exceptional facility with arithmetic in East Asian children.

In keeping with the argument of a cross-generational decline in the basic arithmetical abilities of Americans is the finding of cohort effects for performance on written arithmetic tests (i.e., tests that define the Numerical Facility factor; Schaie, 1983, 1989). Basically, adults who received their elementary school education in the 1940s or later perform more poorly on tests of arithmetical ability than adults who received their elementary school education in the 1920s and 1930s (Schaie, 1993). Moreover, the basic arithmetical skills of American adults has continued to decline for those who received their primary education after the 1940s (Schaie, 1983).

Results from several, though not all (Salthouse & Coon, 1994), cognitive aging studies could also be interpreted as support for the argument that arithmetic abilities have declined within the United States (Allen, Ashcraft, & Weber, 1992; Geary & Wiley, 1991; Geary et al., 1993). For instance, Geary et al. contrasted groups of younger (i.e., college students) and older ($M = 72$ years) adults on the strategies, and speed of executing the underlying processes, used to solve simple (e.g., $8 - 4$) and complex (e.g., $96 - 7$) subtraction problems. The older adults not only knew more facts, on average, than the younger adults, they also executed at least one basic process, trading or borrowing, much more quickly (as in $73 - 6$). The results of the Geary et al. (1993) study and several other cognitive arithmetic studies of younger and older adults (e.g., Allen et al., 1992; Geary & Wiley, 1991) stand in sharp contrast to standard findings in the cognitive aging literature. In nearly all other cognitive domains, younger adults outperform older adults on psychometric ability tests, as well as on tasks that assess more basic processes, such as speed of processing or working memory (Salthouse, 1991, 1992, 1994). The contrast between the relative performance of younger and older American adults in arithmetic in comparison with other ability domains (e.g., spatial ability; e.g., Salthouse & Mitchell, 1990) provides a potentially useful background for contrasting age-related performance differences on ability measures across the United States and East Asian nations.

If the relatively well-developed arithmetical abilities of older American adults in comparison with younger American adults reflect a secular trend within the United States, with declining abilities across successive generations, then the standard finding of younger adults outperforming older adults on cognitive measures should be found for arithmetic in other nations. In this case, an advantage of younger adults over older adults should be found for arithmetic tests as well as for tests of other cognitive abilities. In the United States, a contrasting pattern of older adults doing at least as well as younger ones on arithmetic tests but less well in other domains should be found. On the other hand, if the strong arithmetical abilities of older American adults reflect the results of cumulative experience with arithmetic, then the arithmetical abilities of younger and older adults should be comparable in other nations as well. In this case, the pattern of younger adults outperforming older adults in cognitive domains other than arithmetic, in combination with older adults performing at least as well in arithmetic, should be found cross-nationally.

We assessed these possibilities by administering a battery of psychometric ability tests to groups of younger and older adults in mainland China and the United States. Specifically, we administered measures of numerical facility, perceptual speed, and spatial orientation. By administering tests that span three ability domains, we were able to determine if age-related performance differences emerged across nations, as well as across domains, and whether different patterns emerged for the numerical facility, in comparison with the perceptual speed and spatial orientation, measures. Stevenson and his colleagues (Stevenson et al., 1985) found no spatial ability differences for first- or fifth-grade American and Chinese (i.e., Taiwan) children. For the perceptual speed tests, the scores of the American and Chinese fifth graders were comparable, but the American first graders outperformed their same-age Chinese peers. The overall pattern suggested no systematic differences in the overall intelligence of Chinese, Japanese, and American children, even though the Chinese and Japanese children substantially outperformed the American children on a mathematics achievement test. By administering perceptual speed and spatial orientation tests, as well as arithmetic tests, we could attempt to replicate the Stevenson et al. finding for adults and test the hypothesis that the

[2] Presumably, differences in the structure of number words in English and Asian languages would result in national differences in the initial acquisition of number and arithmetic concepts, but they should not result in national differences once basic arithmetic has been mastered (Ackerman, 1988).

advantage of East Asian individuals over Americans in basic mathematics stems from an East Asian advantage in intelligence (e.g., Lynn, 1982). If Chinese adults are more intelligent than American adults, then both the younger and the older Chinese adults should outperform their same-age American peers on all, or nearly all, of the ability measures.

Method

Participants

The participants were 66 younger (42 female, 24 male) and 44 older (36 female, 8 male; age range = 57 to 85 years) adults from the United States, and 40 younger (20 female, 20 male) and 40 older (20 female, 20 male; age range = 60 to 77 years) adults from China. The 66 younger American adults were recruited from an undergraduate psychology course at the University of Missouri—Columbia and received partial course credit for participating in the experiment. The older American adults were from Atlanta, Georgia, and were recruited through a senior citizens group, which received a donation in their name for their participation. In China, the younger adults were recruited from undergraduate courses (primarily in the social sciences) at East China Normal University, Shanghai, China, and the older adults were recruited through advertisements or referral from Shanghai, China. Both the younger and older Chinese adults were paid a small fee for participating in the experiment.

Before being administered the psychometric tests, described below, all participants were asked to provide information on gender, age, years of education, and to report their health status on a 1 (*excellent*) to 5 (*bad*) Likert-type scale. Descriptive statistics for age, education, and reported health are presented in Table 1. A 2 (nation) × 2 (cohort: younger vs. older) analysis of variance (ANOVA) revealed nonsignificant nation, cohort, and Nation × Cohort effects for years of education ($ps > .10$). For age, significant nation, $F(1, 186) = 42.17$, $p < .001$, and Nation × Cohort, $F(1, 186) = 28.02$, $p < .0001$, effects were found. Post hoc comparisons (Tukey's honestly significant difference; HSD) indicated no significant mean age difference for the younger groups ($p > .05$), but the older American adults were significantly older than their Chinese peers ($p < .05$).

For reported health status, significant nation, $F(1, 171) = 143.77$, $p < .001$, and cohort, $F(1, 171) = 19.48$, $p < .001$, effects were found, but the Nation × Cohort interaction was not significant, $F(1, 171) = 3.77$, $p > .05$. Overall, younger adults reported better health than older adults, although the mean differences were not large (0.2 in the U.S. sample and 0.7 in the Chinese sample), and the American adults reported better health than the Chinese adults. The national difference in reported health status might be due to the tendency of Americans to endorse extreme scores (1 in this case) on Likert-type scales more frequently than do Asians (Chen, Lee, & Stevenson, 1995), rather than an actual difference in physical health. In keeping with this position was the finding that reported health status was not significantly correlated with performance on any of the ability tests in any of the samples ($ps > .05$). The two correlations that approached conventional significance levels ($ps < .10$ and $> .05$) were found for the samples of younger ($r = .33$, between health status and performance on the Card Rotations Test) and older ($r = .30$, between health status and performance on the Identical Pictures Test) Chinese adults. In both cases, the worse the reported health, the better the performance on the ability test.

Ability Measures

All participants were administered a battery of psychometric tests that spanned the Numerical Facility, Perceptual Speed, and Spatial Orientation ability factors (Ekstrom, French, & Harman, 1976). Both forms for each test were administered, and the total score for each test was the sum of both forms.

Numerical facility. Four arithmetic tests were used to represent the numerical facility construct. Two of these, the Addition Test and Addition and Subtraction Correction, were from the Educational Testing Service (ETS) kit of factor-referenced tests (Ekstrom et al., 1976), and the other two tests, Simple Subtraction and Complex Subtraction, were used based on a previous cognitive aging study (i.e., Geary et al., 1993). Each form of the Addition Test allows 2 min to solve as many complex addition problems (e.g., 19 + 8 + 27) as possible. Addition and Subtraction Correction requires participants to indicate whether the presented answer for addition (e.g., 11 + 23 = 34 C I) or subtraction (e.g., 35 − 10 = 20 C I) problems is correct (C) or incorrect (I). Each form of Addition and Subtraction Correction allows 2 min. For each form of the Simple Subtraction (e.g., 8 − 4) and Complex Subtraction (e.g., 78 − 9) tests, participants were allowed 1 min to solve as many problems as possible.

Each form of the Addition Test consists of 60 vertically placed three-addend problems. Forty-eight of these consist of double-digit addends (e.g., 12 + 42 + 53), one problem consists of three single-digit addends (i.e., 8 + 3 + 7), and the remaining problems consist of a mix of single- and double-digit addends (e.g., 19 + 8 + 27). Addition and Subtraction Correction consists of a mixture of 60 horizontally presented double-digit addends or double-digit minuends and subtrahends, as illustrated above. The Simple Subtraction test consists of 72 single-digit minuends and subtrahends with nonnegative answers (e.g., 5 − 2). The items for the Simple Subtraction test include the 36 problems defined by the combination of the integers 1 to 9 (excluding tie problems, e.g., 1 − 1); each problem is presented twice. The Complex Subtraction test consists of 56 problems with double-digit minuends and single-digit subtrahends (e.g., 57 − 3); 28 of these problems require a borrow or trade (e.g., 35 − 9).

Perceptual speed. Two tests from the ETS battery were used to represent the perceptual speed construct: the Number Comparison Test and the Identical Pictures Test (Ekstrom et al., 1976). For each form of

Table 1
Descriptive Statistics for the U.S. and Chinese Samples

	United States				China			
	Younger		Older		Younger		Older	
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Age	20.1	3.7	72.7	5.7	19.4	1.2	65.6	4.3
Years of education	14.0	0.6	14.7	2.3	14.3	0.8	14.4	2.5
Health status	1.5	0.5	1.7	0.7	2.4	0.6	3.1	0.6

Note. Health status was rated on a 1 (*excellent*) to 5 (*bad*) Likert-type scale.

both tests, participants were allowed 1.5 min to solve as many problems as possible. For the Number Comparison Test, pairs of digit strings, ranging from 3 to 13 digits, are presented. The task is to indicate whether the digit strings differ. The Identical Pictures Test requires participants to indicate which of five alternative pictures is identical to a comparison picture.

Spatial orientation. Three tests were used to represent the spatial orientation construct: the Card Rotations Test and the Cube Comparisons Test, both from the ETS battery, as well as the Mental Rotation Test (Ekstrom et al., 1976; Vandenberg & Kuse, 1978). For each form of all three tests, participants were allowed 3 min to solve as many problems as possible. The Card Rotations Test requires the rotation of figures in two-dimensional space, whereas the Mental Rotation Test requires the rotation of figures in three-dimensional space. The Cube Comparisons Test requires the rotation of drawings of cubes, some of which need to be rotated in two-dimensional space, others in three-dimensional space.

Procedure

Translation. For all measures, the test stimuli were identical in the English and Chinese versions. To ensure comparability, an experienced translator first translated the English instructions into Chinese. The Chinese version was then back-translated into English by another experienced translator who was not familiar with the English instructions. Discrepancies between the original English instructions and the back-translated instructions were then discussed between David C. Geary and the two translators, which resulted in a second Chinese version of the instructions. To ensure that these instructions were clear, we then administered the Chinese versions of all ability measures to two individuals who were not familiar with either the English or Chinese versions of the tests. Both individuals indicated that the instructions were readily understandable.

Administration. All ability measures were administered, either individually or in small groups, under standard instructions for both the U.S. and Chinese samples. All participants were first administered the numerical facility tests, followed in turn by the perceptual speed and spatial orientation measures. The entire testing session lasted about 50 min.

Results

For ease of presentation, the results are discussed in three general sections. In the first, the factor structure of the psychometric battery is assessed for the American and Chinese samples. National and cohort (i.e., younger vs. older) differences on the ability tests are described in the second section, and issues concerning differences in the intelligence of the samples and sample selectivity are addressed in the final section.

Factor Structure

The factor structure of the ability tests was assessed by means of a confirmatory factor analysis (Jöreskog, 1969), using the CALIS program available through SAS (SAS Institute, 1990). For each sample, partial covariance matrices were obtained by partialing the linear, quadratic, and cubic age trends from performance on the ability tests. A three-factor—Numerical Facility, Perceptual Speed, and Spatial Orientation—model was then fitted to the covariance matrices. For these analyses, the indicators for each factor were identical to those described in the Method section (e.g., the Perceptual Speed factor was defined by the Number Comparison Test and the Identical Pictures Test). The factors were allowed to correlate, and all nondefining factor loadings were fixed at 0. The adequacy of such models can be assessed by means of a goodness-of-fit index (GFI), with values greater than .90 typically considered acceptable (Bentler & Bonett, 1980).

For the initial models, the GFI values were .88 and .89 (actually .899) for the U.S. and Chinese samples, respectively. The estimation of two correlated residuals (i.e., allowing the unique variance for two sets of ability tests to correlate) for the U.S. sample and one correlated residual for the Chinese sample yielded acceptable GFI values for both the U.S. (GFI = .91) and Chinese (GFI = .92) samples.[3] The factor patterns were highly similar across the two samples, with a coefficient of congruence (Gorsuch, 1983) of .99 for the Numerical Facility factor, and .95 and .96, respectively, for the Perceptual Speed and Spatial Orientation factors. Thus, with very slight modifications, the expected three-factor structure of the ability tests was confirmed for both the U.S. and Chinese samples.

Ability Test Performance

Mean ability test scores across nation and cohort are shown in Table 2. To facilitate the discussion of the results, we present the analyses separately for each of the ability constructs. For each of the respective sets of ability tests, national and cohort differences were initially assessed by means of a multivariate analysis of variance (MANOVA). Significant MANOVAs were followed up by 2 (nation) × 2 (cohort) ANOVAs; simple means were contrasted by means of the HSD procedure.

Numerical facility. The MANOVA for the four arithmetic tests revealed significant nation, $F(4, 178) = 30.65$, $p < .001$; cohort, $F(4, 178) = 45.67$, $p < .001$; and Nation × Cohort, $F(4, 178) = 18.98$, $p < .001$, effects. The ANOVAs revealed that the nation, cohort, and Nation × Cohort effects were significant ($ps < .05$) for the Simple Subtraction, Complex Subtraction, and Addition and Subtraction Correction tests. For the Addition Test, the Nation × Cohort interaction was significant ($p < .001$), but the main effects for nation and cohort were not ($ps > .20$).[4]

For the Addition test, comparisons of simple means indicated that the older American adults and the younger Chinese adults had comparable mean scores ($p > .05$), which were significantly higher than the mean scores for the younger American adults and the older Chinese adults ($ps < .05$). The mean score of the younger American adults was, in turn, significantly higher than that of

[3] In the U.S. sample, the correlated residuals were between the Identical Pictures Test and the Card Rotations Test (.42) and between the Addition Test and the Mental Rotation Test (−.43). For the Chinese sample, the correlated residual was between Addition and Subtraction Correction and the Number Comparison Test (.42). Tables of the standardized factor loadings, factor correlations, and correlated residuals are available from David C. Geary on request.

[4] A MANOVA confirmed that there were no gender differences or Cohort × Gender interactions for the numerical facility tests for either the U.S. or Chinese samples ($ps > .10$). For the perceptual speed and spatial orientation tests, there were no Gender × Cohort interactions in either sample ($ps > .10$). For the Chinese sample, the only significant ($p < .05$) gender difference, which favored men, was for the Mental Rotation Test. For the U.S. sample, men outperformed women on the Mental Rotation Test and the Cube Comparisons Test ($ps < .05$), whereas women outperformed men on the Number Comparison Test ($p < .01$).

Table 2
Mean Ability Test Scores Across Group and Nation

	United States				China			
	Younger		Older		Younger		Older	
Test	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Numerical Facility								
Addition	36.7	8.7	45.6	14.4	46.6	9.1	33.1	16.1
Simple subtraction	118.3	19.0	99.1	22.7	134.2	13.3	96.2	31.6
Complex subtraction	35.1	9.5	35.8	10.7	58.6	10.7	36.9	17.3
Addition/subtraction	67.1	15.2	56.9	14.2	99.4	14.8	57.9	24.8
Perceptual Speed								
Number comparison	54.1	9.1	41.3	11.6	55.4	12.1	38.4	17.4
Identical pictures	73.4	12.7	40.9	7.5	63.3	11.8	46.6	12.6
Spatial Orientation								
Card rotations	115.9	25.5	59.8	29.8	94.7	27.9	58.3	42.1
Cube comparisons	16.8	9.7	4.6	5.4	13.7	8.7	3.6	6.2
Mental rotation test	15.3	9.6	4.1	6.4	13.9	9.9	3.3	5.1

the older Chinese adults ($p < .05$). For the Simple Subtraction test and Addition and Subtraction Correction, the younger Chinese adults outperformed all other groups ($ps < .05$). The performance of the older American and older Chinese groups was comparable ($ps > .05$) but was significantly lower than that of the younger American adults ($ps < .05$). For the Complex Subtraction test, the performance of the two older groups and the group of younger American adults was comparable ($ps > .05$) but was significantly lower than that of the younger Chinese adults ($ps < .05$). Finally, partialing age in the comparison of the two older groups to adjust for the significant age difference did not alter the results for any of the arithmetic tests.

In all, the younger Chinese adults outperformed their American peers substantially on all four arithmetic tests. The advantage of the younger Chinese adults over the younger American adults was about 1 standard deviation for the Addition Test and the Simple Subtraction test, and more than 2 standard deviations for the Complex Subtraction and the Addition and Subtraction Correction tests. In contrast, for three of the four arithmetic tests, there was no significant national difference for the older groups. The one test that did show a significant difference across the older groups, that is, the Addition Test, favored the American participants. In fact, for the Addition Test, the mean performance of the older American adults was comparable with that of the younger Chinese adults, and about 0.75 standard deviations above the mean of the younger American adults.

Another way to highlight and summarize the overall pattern is by means of standard scores. Here, for each of the four arithmetic tests, the overall mean was set at 0 and the standard deviation at 1. The resulting *z* scores were then summed to create an overall performance score for the numerical facility construct. On the basis of the cognitive aging literature, this procedure might be expected to yield mean scores above the overall mean for the younger groups, and scores below the overall mean for the older groups. Indeed, the left section of Figure 1 shows exactly this pattern for the Chinese participants; that is, the mean performance of the younger Chinese adults is above the overall mean, whereas the mean performance of the older Chinese adults is below the overall mean. In contrast, the mean performance of both the younger American adults and the older American adults is below the overall mean.

Perceptual speed. For the perceptual speed tests, the initial MANOVA revealed a nonsignificant main effect for nation, $F(2, 181) < 1$, $p > .25$, but significant cohort, $F(2, 181) = 102.23$, $p < .001$, and Nation × Cohort, $F(2, 181) = 16.39$, $p < .001$, effects. The ANOVAs confirmed that the advantage of the younger adults over the older adults was significant for both the Number Comparison Test and the Identical Pictures Test ($ps < .001$). Comparisons of simple means indicated that these age-related differences were significant in both the U.S. and Chinese samples ($ps < .05$). The Nation × Cohort interaction was not significant for the Number Comparison Test, $F(1, 184) = 1.21$, $p > .25$, but was significant for the Identical Pictures Test, $F(1, 182) = 21.13$, $p < .001$. With this cross-over interaction, the younger American adults outperformed their Chinese peers ($p < .05$), whereas the older Chinese adults outperformed their American peers ($p < .05$). However, partialing age from performance on the Identical Pictures Test eliminated the advantage of the older Chinese adults over the older American adults, $F(1, 78) = 1.13$, $p > .25$.

The overall pattern is summarized in the middle section of Figure 1. In keeping with the expected cognitive-aging pattern, it can be seen that for both the U.S. and Chinese samples, the mean for the younger adults is above the overall mean, whereas the mean for the older adults is below the overall mean.

Spatial orientation. For the spatial tests, the MANOVA revealed a nonsignificant main effect for nation, $F(3, 184) = 2.29$, $p > .05$, and a nonsignificant Nation × Cohort interaction, $F(3, 184) = 1.74$, $p > .10$. The main effect for cohort was significant, however, $F(3, 184) = 48.98$, $p < .001$. For both the American and Chinese samples, the advantage of the younger adults over the older adults was significant for all three spatial tests ($ps < .05$). The overall pattern is summarized in the right section of Figure 1 and, again, shows that for both the U.S. and Chinese samples, the mean for the younger adults is above the overall mean, and the mean for the older adults is below the overall mean.

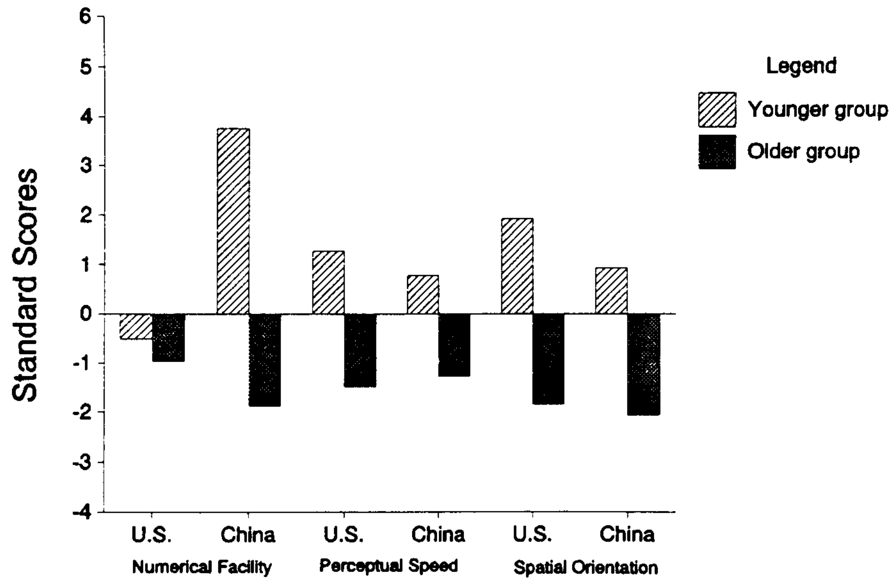

Figure 1. Standard scores across cohort and nation for the numerical facility, perceptual speed, and spatial orientation ability domains.

Intelligence and Selectivity

Even though the pattern of results shown in Figure 1 is consistent with the position that there has been a relatively recent decline in the arithmetical abilities of Americans, two related issues need to be addressed before such an argument can be made with any certainty. The first concerns whether the advantage of the younger Chinese adults over the younger American adults in arithmetic could be attributable to differences in the intelligence of the Chinese and American samples, and the second concerns whether the younger Chinese adults are simply a more highly selected group than the younger American adults.

Intelligence. If the advantage of the younger Chinese adults over their same-age American peers for the arithmetic tests was related to an East Asian advantage in intelligence, then both the younger and older Chinese adults should have shown an advantage over their same-age American peers on all, or nearly all, of the ability measures. This was clearly not the case. Except for the finding of younger Chinese adults outperforming younger American adults on the arithmetic tests, comparisons for the remaining ability tests, as well as for the groups of older adults, revealed either no national differences or a slight American advantage, in keeping with the findings of Stevenson et al. (1985) for comparisons of Chinese and American children.

Nevertheless, to further assess the possibility of ability differences across our Chinese and American samples, we analyzed the overall pattern of performance on the psychometric tests by means of a mixed-design ANOVA. Here, the z scores for the three ability domains (i.e., the mean of the z scores for the individual tests in each domain) served as a within-subjects variable (ability), and nation and cohort served as between-subjects variables. With this analysis, there were two crucial tests: the Nation × Ability interaction and the Nation × Ability × Cohort interaction. The finding of a nonsignificant Nation × Ability interaction, $F(2, 185) < 1, p > .50$, is consistent with the argument that there were no systematic ability differences across the Chinese and American samples. The significant Nation × Ability × Cohort effect, $F(2, 185) = 20.05, p < .0001$, indicated a selective age-related difference across the Chinese and American samples for at least one of the ability domains.

Follow-up ANOVAs within each ability domain indicated that the main effect for nation was not significant for any of the ability domains: numerical facility, $F(1, 186) < 1, p > .50$; perceptual speed, $F(1, 186) = 1.97, p > .15$; and spatial orientation, $F(1, 186) = 1.76, p > .15$. The Nation × Cohort effect was not significant for the perceptual speed, $F(1, 186) = 2.98, p > .05$, or spatial orientation, $F(1, 186) = 1.19, p > .25$, scores, but was highly significant for the numerical facility scores, $F(1, 186) = 20.05, p < .0001$. For numerical facility, the overall performance of the younger and older American groups did not differ ($p > .05$), but the younger Chinese adults outperformed the older Chinese adults ($p < .05$). The overall pattern confirms a selective age-related performance difference across the Chinese and American samples for the numerical facility tests. Moreover, as noted earlier, the pattern of no overall national difference in ability test scores across the three domains is consistent with the view that the East Asian and American samples used in this study were comparable in terms of overall levels of intelligence.

Sample selectivity. The issue of sample selectivity is particularly important for the younger samples, because contemporary Chinese universities are more selective, on average, than are contemporary universities in the United States; this is less of an issue for our older samples given that college attendance was relatively unusual in both countries earlier in this century. Even if the younger samples are comparable in terms of general abilities, it is possible that the large advantage of the younger Chinese adults over the younger American adults in arithmetic was somehow related to the selection criteria (e.g., types of

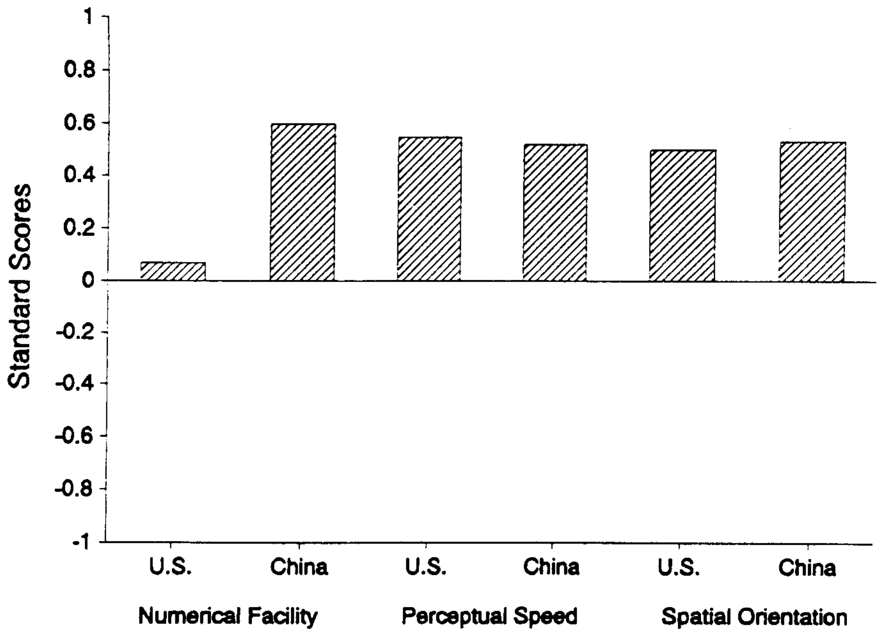

Figure 2. Within-nation standard score comparisons of the relative performance of the younger and older groups for the numerical facility, perceptual speed, and spatial orientation ability domains.

tests) associated with admission to the two respective universities.[5] We have no way of directly testing this possibility.

However, within-nation comparisons of the relative advantage of younger adults over older adults across the three ability domains allow us to indirectly test this possibility. Because the performance of the groups of older Chinese and older American adults was comparable across all three ability domains, and because these groups were comparable in terms of mean levels of education, their performance can provide a benchmark against which the relative performance of their same-nation younger peers can be assessed. If the younger Chinese sample were selected such that they had unusually high arithmetical abilities, then the younger Chinese adults should show a larger advantage over the older Chinese adults on the arithmetic tests, relative to the perceptual speed and spatial orientation tests. If the samples are roughly equivalent, then the relative advantage of the younger adults over same-nation older adults should be similar across the Chinese and American samples and across the three ability domains. In keeping with the central hypothesis of this study, the only exception should be for the comparison of the numerical facility scores for the groups of younger and older Americans.

Figure 2 shows the within-nation standard score comparisons across the younger and older groups for the numerical facility, perceptual speed, and spatial orientation constructs. Scores above the mean of 0 indicate an advantage of younger adults over older adults in terms of standard deviation units. For the Chinese samples, the advantage of the younger adults over the older adults was highly similar across all three ability domains and ranged from 0.52 to 0.60 standard deviations. Similarly, the advantage of the younger American adults over the older American adults was 0.55 and 0.50 standard deviations for the perceptual speed and spatial orientation scores, respectively. The only comparison that does not fit with the overall pattern is the relatively comparable numerical facility scores of the younger and older American adults (the mean difference was 0.07 standard deviations). The overall pattern seems to be consistent with the view that the greater selectivity of Chinese universities in comparison with American universities has not resulted in unusually high arithmetical abilities in our sample of younger Chinese adults.[6] If anything, the overall pattern is consistent with the view that the arithmetical abilities of the younger American adults is unusually low. This is not to say that the samples we have compared in this study are representative of the general populations of the United States and China; they almost certainly are not. The point is that when we contrast groups of comparably, and relatively well, educated younger and older Chinese and American adults, the "value," in terms of fostering the acquisition of basic arithmetic skills, of education in the United States appears to have changed over the past 60 years.

Discussion

By providing a comparison of the psychometric abilities of younger and older American and Chinese adults, the present study provided a unique perspective on cross-national differences in basic mathematical abilities, and it suggests that the often cited advantage of East Asian individuals over individuals in the United States in mathematics (e.g., Stevenson et al., 1990; Stevenson et al., 1993) might not have existed, at least for arithmetic, 60 years ago. The two most important outcomes of

[5] Students at the University of Missouri—Columbia are a relatively select group. Of all of the students taking the American College Test (ACT), their mean ACT score of 25 is at the 83rd percentile nationally.

[6] Any differences in the selectivity of the two universities would likely be most evident for complex mathematics, not the basic skills assessed in this study.