

Computational and reasoning abilities in arithmetic: Cross-generational change in China and the United States

DAVID C. GEARY and CARMEN O. HAMSON
University of Missouri, Columbia, Missouri

GUO-PENG CHEN
East China Normal University, Shanghai, China

FAN LIU
China Central Institute for Educational Science, Beijing, China

MARY K. HOARD
University of Missouri, Columbia, Missouri

and

TIMOTHY A. SALTHOUSE
Georgia Institute of Technology, Atlanta, Georgia

A Chinese advantage over Americans was found for economically relevant computational and reasoning abilities in arithmetic for groups of 6th- and 12th-grade students matched or equated on general intelligence. No cross-national difference for computational or reasoning abilities was found for samples of older (60- to 80-year-old) Chinese and American adults equated on general intelligence. The pattern of change in arithmetical competencies across cohorts suggests that the Chinese advantage in 6th and 12th grade is due to a cross-generational decline in competencies in the United States and a cross-generational improvement in China.

The poor mathematical competencies of American students in relation to international standards have been demonstrated many times over the past 30 years and has been shown to be true for nearly all mathematical domains that are taught in school (Beaton et al., 1996; Husén, 1967). Differences between the mathematical performance of U.S. students and students from other countries seem to be most dramatic when comparisons are made between the United States and East Asian nations (Stevenson, Chen, & Lee, 1993). In fact, the mathematical competencies of East Asian students are consistently among the best in the world, whereas the mathematical competencies of American students consistently rank below the international average (Beaton et al., 1996).

Explanations of the stark contrast between the mathematical development of American children and their East

Asian peers tend to focus on (1) an East Asian advantage in general intelligence (IQ; Lynn, 1982; Rushton, 1995) or (2) differences in the mathematics education of U.S. and East Asian children (Stevenson & Stigler, 1992). The present study was designed to directly test the first explanation and indirectly assess the second. At the same time, the arithmetic tests used in this study were chosen because performance on similar tests has been shown to be related to employability, wages, and productivity in many blue collar and white collar occupations (Boissiere, Knight, & Sabot, 1985; Rivera-Batiz, 1992). Any cross-national differences on these measures thus bode potentially important economic consequences for the United States.

In a related study, Stevenson and his colleagues indirectly tested the hypothesis that the cross-national achievement gap in mathematics is related to differences in the IQ of East Asian and American children (Stevenson et al., 1985). Here, groups of Chinese, Japanese, and Anglo-American children were administered a battery of psychometric tests along with achievement tests in reading and mathematics. There were no systematic differences in the overall psychometric test performance of these groups, but the Chinese and Japanese students outperformed their American peers on the mathematics achievement test, suggesting that the achievement gap in mathematics is not related to an East Asian advantage in IQ.

We thank Linda Coutts, Harry Oldweiler, Mary Porter, Becky Gregory, Nicole Harris, Aaron Stratman, Al Reuck, Sheela Vishwanath, Natalija Popovic, Jennifer Lin, Hong Hu, and Shang S. Liu for their assistance with various aspects of the study, and Michael Stadler, Donald Kausler, and three anonymous reviewers for comments on an earlier draft. The research was supported by Grant HD27931 from the NICHD and Grant AG06826 from the NIA. Correspondence should be addressed to D. C. Geary, Department of Psychology, 210 McAlester Hall, University of Missouri, Columbia, MO 65211-2500 (e-mail: psycore@showme.missouri.edu).

A more recent study supported the same conclusion (Geary, Salthouse, Chen, & Fan, 1996). Here, groups of comparably educated Chinese and American college students and older (57- to 85-year-old) adults were administered a battery of psychometric tests, including measures of computational arithmetic (e.g., $45+83+19$), perceptual speed (e.g., speed of reading numbers), and spatial abilities. There were no substantive national differences for performance on the perceptual speed or spatial measures for either age group, suggesting that there were no IQ differences across these Chinese and American adults. For the arithmetic tests, the older Americans outscored their Chinese peers' in complex addition, but there were no significant differences for three other arithmetic measures. The Chinese college students, in contrast, outperformed—by between 1.1 and 2.2 *SD*—the American college students on all four arithmetic tests. Again, the pattern suggests that the poor arithmetical performance of Americans relative to their East Asian cohorts is not due to IQ differences and suggests that the current achievement gap is a relatively recent phenomenon, at least for arithmetic.

The present study provides a more direct assessment of the relation between IQ and the East Asian/American achievement gap in arithmetic than is possible with the psychometric test approach used in these earlier studies. Specifically, groups of 6th graders, 12th graders, and older adults from the United States and China were administered a standardized IQ test and sets of arithmetic computational and reasoning ability tests. We assessed these age groups on the basis of evidence that basic arithmetical abilities have shown a cross-generational decline in the United States (Geary et al., 1996; Schaie, 1996). For computational arithmetic, individuals who started elementary school in the 1920s and 1930s outperform—once age-related changes in processing speed are taken into account—individuals who began their primary education just after World War II (i.e., 1945), who, in turn, outperform individuals who began elementary school in the mid-1960s or later (Schaie, 1996).

If the current achievement gap in arithmetic stems from a cross-generational change in educational outcomes, rather than differences in IQ, then a pattern of East Asian children and adolescents outperforming their IQ-matched American peers should be found, in combination with no cross-national differences for older adults (Geary et al., 1996). Moreover, once IQ, years of schooling, and so on, are controlled, a cross-generational decline in arithmetical competencies should be evident across the U.S. samples; no such decline, or even improving competencies, should be found across the Chinese samples.

METHOD

Participants

The participants were 372 6th graders, 12th graders, or older adults from Shanghai, China; Columbia, Missouri; or Atlanta, Georgia.¹ As shown in Table 1, the IQ scores of the Chinese and American participants differed significantly for the samples of 6th graders and older adults, the former favoring China and the latter favoring the United States ($ps < .05$). Thus IQ was used as a covariate in the analyses of arithmetic

Table 1
Participant Characteristics

Variable	China		United States	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
	Grade 6			
IQ	121	13	116	11
Age	12	0.6	12	0.5
% male	45		63	
<i>n</i>	77		57	
	Grade 12			
IQ	120	11	118	10
Age	17	—	17	—
% male	48		31	
<i>n</i>	80		55	
	Older Adult			
IQ	105	10	113	14
Age	66	4.0	70	4.5
Years of education	11.0	2.2	13.8	2.0
% male	54		32	
<i>n</i>	56		47	

Note—One American 6th grader and 1 Chinese 12th grader did not answer the gender item.

test performance for the 6th-grade and older adult samples. For the older sample, years of education also favored the Americans ($p < .001$) and was therefore included as an additional covariate. There was no significant difference between the IQ scores of the Chinese and American 12th graders ($p > .10$), and thus IQ was not used as a covariate for this sample.

To control for national differences in the number of adolescents attending high school—only about 28% of adolescents attend academic high schools in Shanghai—the American 12th-grade sample was selected to represent the top 25%–30% of the 17-year-old cohort. After taking the dropout rate and enrollment patterns for advanced mathematics courses into consideration, this end was achieved by restricting the American sample to students enrolled in calculus, precalculus, or college algebra courses. All members of this clearly elite group of U.S. students were in their 4th year of high school mathematics, as were all of the Chinese 12th graders.

In both China and the United States, participants were selected from two high schools. School officials in Shanghai described the high schools as “good but not excellent” (these were not key schools). During the year the study was conducted, one of the U.S. high schools was 1 of 149 recipients of a National Excellence in Education Blue Ribbon Award, and the only high school in the United States to be concurrently awarded technology and parent involvement awards. The other U.S. high school was a previous recipient of a National Excellence in Education Blue Ribbon Award. If anything, the U.S. high schools were relatively more elite than their counterparts in China.

In both countries, the 6th graders were selected from two middle schools that send students (via junior high school) to the schools from which the 12th graders were selected. In order to select students who were comparable to the students in the high school samples, teachers in both countries were asked to select the top one third of their math students for potential inclusion in the study.² The sample of older adults was restricted to 60- to 80-year-olds and to individuals with between 9 and 16 years of education, inclusive. The latter restriction was imposed to control for variability in educational levels within and across samples. The age range was based on Schaie's (1996) findings of cohort differences in computational arithmetic and included cohorts for which arithmetical abilities are near the U.S. peak.

Ability Measures

For the computational and IQ tests, the stimuli were identical in the U.S. and Chinese versions, and the translation and back-translation of the instructions followed previous studies (Geary, Fan, & Bow-Thomas, 1992). For the arithmetical reasoning tests—composed of multistep word problems—the items were first translated from English to Chi-

nese by an experienced translator. The initial Chinese versions (one for each test) were then back-translated by another individual who was not familiar with the English version of the tests. Disagreements were resolved through discussion between D.C.G. and the two translators, resulting in a second Chinese version of the tests. These versions were then administered to a Chinese adult and a 6th grader to assess the clarity of the items. On the basis of this administration, a few minor changes were incorporated into a third and final Chinese version of the arithmetical reasoning tests.

Intelligence. *Raven's Progressive Matrices* (Raven, Court, & Raven, 1993) served as the measure of general intelligence. Due to time constraints, the test was started at the item representing an IQ of 80, and participants were given credit for all items up to this level. Participants were then allowed 30 min to complete the remaining items and nearly all did so in the allotted time; IQ scores were based on urban Chinese norms to ensure that the performances of the U.S. and Chinese samples would be directly comparable (Lee & Chen, 1989).

Arithmetical computations. Three tests, each including two forms, were used to assess computational abilities. One of these, the Addition Test, was from the Educational Testing Service (ETS) kit of factor-referenced tests (Ekstrom, French, & Harman, 1976) and the other two, simple subtraction and complex subtraction, were used on the basis of previous research (Geary et al., 1996). Each form of the Addition Test, hereafter referred to as "complex addition," allows 2 min to solve as many complex addition problems as possible (e.g., 19+8+27). For each form of the simple subtraction (e.g., 9-5) and complex subtraction (e.g., 78-9) tests, participants were allowed 1 min to solve as many problems as possible. For all tests, the score was the number of items solved correctly across both forms. Within cohorts, test reliabilities were estimated by means of Cronbach's alpha and ranged between .88 and .97, with a median value of .93.

Arithmetical reasoning. Both arithmetical reasoning tests—Necessary Arithmetic Operations (NAO) and Arithmetic Aptitude (AA)—were from the ETS kit of factor-referenced tests. Again, both forms of each test were administered, with 5 min allowed for each form of NAO and 10 min for each form of AA. As noted, both tests require the solving of multistep arithmetic word problems, but NAO only requires participants to indicate the order in which arithmetic operations (e.g., multiplication then division) need to be used to solve the problem; no computations are needed. Items for AA require the participants to solve multistep word problems. The score is the number of items solved correctly minus a fraction of the number of items solved incorrectly across both forms. Within cohorts, reliability estimates ranged between .65 and .87, with a median value of .75.

Procedure

All testing was done in a quiet room—either at the school site or in a laboratory setting—with groups of participants. Standard procedures were followed. The 12th graders and older adults were tested in a single session that lasted about 1.5 h, and the 6th graders were tested in two 45-min sessions that were, in most cases, separated by 1 week.

RESULTS

The results are presented in two sections. The first presents mean arithmetical test scores for each cohort, and the second presents an assessment of the pattern of cross-generational change in arithmetical abilities in China and the United States.

Cross-National Differences in Arithmetical Abilities

Table 2 shows that the Chinese 6th and 12th graders outperformed their American peers on every arithmetic test ($ps < .05$).³ The last column of Table 2 represents the magnitude of this Chinese advantage in terms of effect size (d)—that is, the difference between the Chinese and American mean scores relative to the SD of the Chinese

sample (Cohen, 1988). Inspection of these results reveals that 8 of the 10 effect sizes are >1.0 and five are >1.5 . The practical significance of these results is shown in Figure 1. Here, the mean score of the American sample is represented in terms of a percentile ranking relative to the Chinese sample. As an example, for the simple subtraction test, the performance of the average American 6th-grade student was at the 2nd percentile of his/her same-grade, IQ-equated Chinese peer. Stated otherwise, 98% of the Chinese 6th graders had scores on the simple subtraction test that were higher than the mean of the American 6th graders.

Overall, the mean scores of the American 6th and 12th graders were at or below the 11th percentile on each of the three computational tests. Moreover, the Chinese 6th graders performed as well as the American 12th graders

Table 2
Mean Arithmetic Test Scores and Standard Deviations Across Cohorts and Nations

Test	China		United States		Effect Size (d)
	M	SD	M	SD	
Grade 6					
Computational tests					
Simple subtraction	117.1	21.2	73.7	21.4	2.05
Complex subtraction	51.4	14.5	24.0	8.6	1.89
Complex addition	36.2	9.5	24.6	7.8	1.22
Reasoning tests					
Necessary arithmetic operations	14.0	3.9	10.0	4.3	1.03
Arithmetic aptitude	10.9	6.1	8.7	4.4	0.36
Grade 12					
Computational tests					
Simple subtraction	138.3	10.8	116.2	23.3	2.05
Complex subtraction	68.1	16.9	37.4	9.3	1.82
Complex addition	47.2	8.6	36.0	8.3	1.30
Reasoning tests					
Necessary arithmetic operations	19.4	4.1	17.3	4.6	0.51
Arithmetic aptitude	22.4	4.7	15.2	5.5	1.53
Older Adults					
Computational tests					
Simple subtraction	92.8	30.8	94.7	20.4	-0.06
Complex subtraction	30.1	17.4	35.4	10.8	-0.30
Complex addition	26.9	12.2	39.0	11.4	-0.99
Reasoning tests					
Necessary arithmetic operations	12.3	3.9	12.4	5.8	-0.03
Arithmetic aptitude	10.5	4.6	10.2	7.2	0.07

Note—The scores for the U.S. 6th graders were adjusted to control for a Chinese advantage in mean IQ scores, and the scores for the Chinese older adults were adjusted to control for a U.S. advantage in mean IQ scores and years of education. No adjustments were made to the scores of the 12th graders, because the mean IQ scores of the U.S. and Chinese 12th graders did not differ significantly ($p > .10$).

The effect size is given by

$$d = \frac{M(\text{China}) - M(\text{U.S.})}{SD(\text{China})}$$

where $M(\text{China})$ and $M(\text{U.S.})$ are the mean scores of the Chinese and U.S. samples, respectively, and $SD(\text{China})$ is the standard deviation of $M(\text{China})$. Positive scores indicate a Chinese advantage, and negative scores indicate a U.S. advantage.

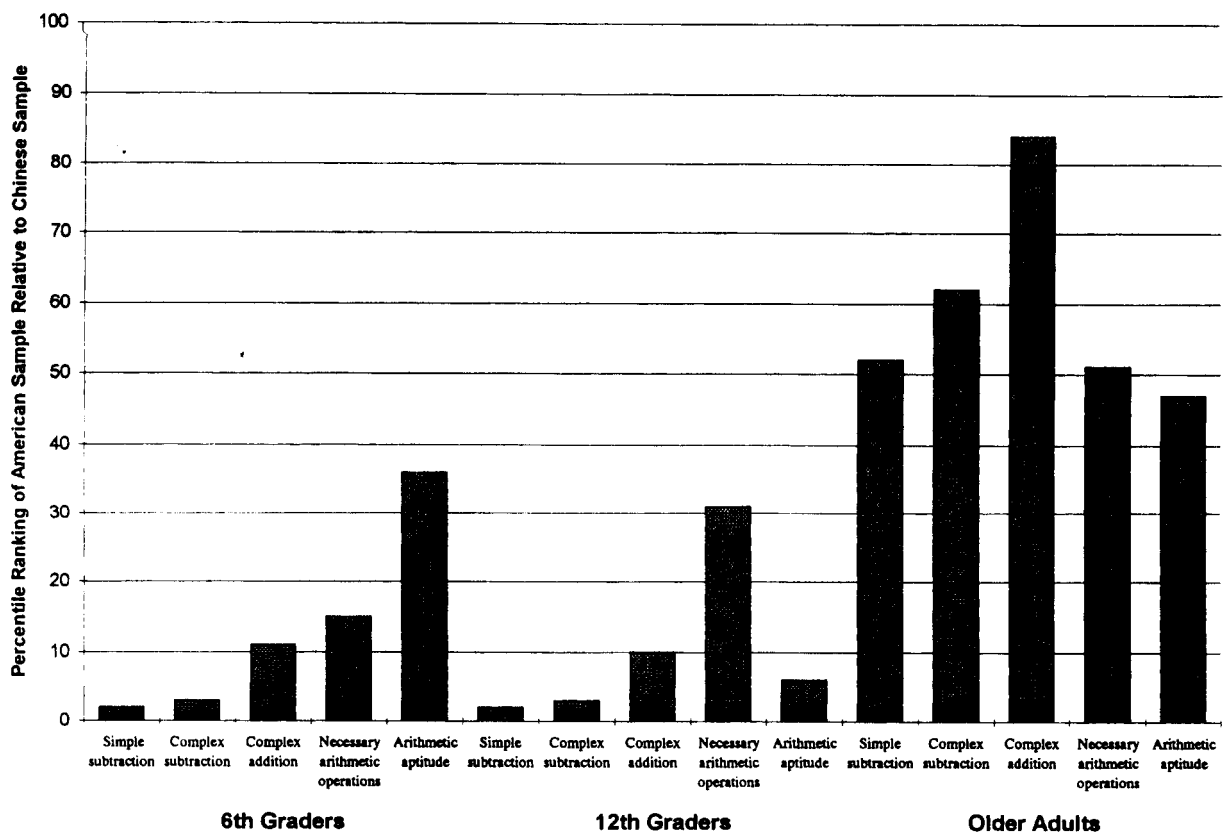


Figure 1. The percentile ranking of the mean U.S. score relative to the performance of the same-age Chinese participants. Scores below 50 indicate a Chinese advantage and scores above 50 an American advantage.

on the simple subtraction and complex addition tests ($ps > .50$) and substantially outscored the American 12th graders on the complex subtraction test ($p < .001$). The mean performance of the American 6th and 12th graders was somewhat better on the reasoning tests, ranging between the 6th and 36th percentiles of the respective Chinese samples.

The pattern of results was very different for the older adults. There were no significant national differences on four of the five tests ($ps > .05$), and the older Americans outperformed their Chinese peers on the complex addition test ($p < .001$), scoring at the 84th percentile of the Chinese sample.

Cross-Generational Change in Arithmetical Abilities

The pattern of across-generational change was estimated by controlling for age, IQ, and educational level on performance on the computational and reasoning measures. First, performance on each of the five arithmetic tests was standardized with a mean of 0 and an SD of 1. Composite computational and reasoning variables were created on the basis of the respective means of the three computational and two reasoning tests; the respective reliabilities of the resulting composites were .89 and

.88. Next, the linear, quadratic, and cubic trends of age, IQ, years of education, and raw scores on the IQ test were partialled from the composite variables. The IQ test raw scores were partialled to control for age-related changes in performance on psychometric tests (Fry & Hale, 1996; Kail, 1991). The resulting residualized standard scores are plotted in Figure 2 and should largely reflect cohort differences in the effects of schooling, independent of IQ, age, years spent in school, and age-related performance on psychometric tests.

Inspection of Figure 2 reveals that within the U.S. sample, the older adults had significantly higher computational and reasoning scores than did the 6th and 12th graders ($ps < .05$), who did not differ from each other on either composite ($ps > .05$). Within the Chinese sample, the older adults had significantly lower computational and reasoning scores than the 6th and 12th graders ($ps < .05$), who did not differ from each other on either composite ($ps > .05$). Across nations, the slight advantage of the older American adults over the older Chinese adults was marginally significant for computational ($p = .085$) but not for reasoning ($p > .50$) abilities, whereas the Chinese advantage for 6th and 12th graders was significant for both computational and reasoning abilities ($ps < .001$).

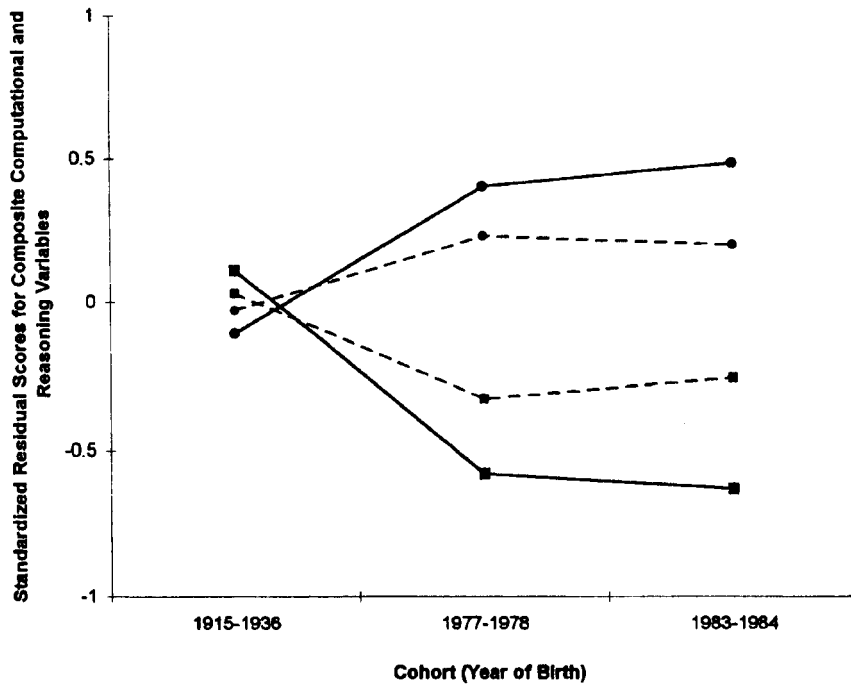


Figure 2. Estimated pattern of cross-generational (1915–1936, older adults; 1977–1978, 12th graders; 1983–1984, 6th graders) change in arithmetical computational (solid line) and reasoning (dashed line) abilities. ■, the United States; ●, China.

DISCUSSION

The results of this study suggest that the current East Asian/American achievement gap in arithmetic is a relatively recent phenomenon that stems from a cross-generational decline in basic competencies in the United States combined with a cross-generational improvement in basic competencies in China. The pattern of results is most consistent with the view that the current achievement gap is largely related to differences in schooling and cultural supports for educational achievement rather than an East Asian advantage in IQ (Stevenson & Stigler, 1992). Although we did not directly test the schooling hypothesis, the pattern of cohort differences noted in Figures 1 and 2 is consistent with the cohort effects described by Schaie (1996) and would appear to be consistent with a cross-generational change in arithmetical competencies in the United States. Any such change is very likely to be the result of changes in schooling or the valuation of schooling, given that performance on the arithmetic tests used in this study is strongly affected by schooling (Geary, 1995).

Alternatively, it might be argued that our findings reflect an improvement in arithmetical competencies throughout the life span in such a way that Americans eventually catch up to East Asians. There are several reasons why this alternative explanation is not likely to be correct. First, performance on psychometric tests almost always declines with adult aging (Salthouse, 1996; Schaie, 1996), suggesting that any catching up in adulthood is very unlikely. Furthermore, a steady improvement in arithmetical abilities in adulthood should result in the same cross-generational pattern in China and the United States, but this was clearly not the case. Finally, Schaie's earlier described cross-generational findings for computational arithmetic are robust—that is, evident in comparisons within and across families. For computational arithmetic, but not other cognitive abilities, grandparents outperform their children, who in turn outperform their children (i.e., the grandparents' grandchildren).

It might also be argued that our results stem from systematic differences between the older samples. Although this explanation cannot be ruled out, there are two reasons why it is not likely. First, the U.S. cohort effects found in this study and our previous study are the same as those described by Schaie (1996), indicating that our samples of older

American adults are not unusual. Second, in our previous study, we found the same pattern of no differences—except in complex addition—in computational arithmetic for different samples of older adults from the United States and China, samples that were matched, as a group, on years of education (Geary et al., 1996). The combination of results suggests that basic arithmetical competencies have declined in the United States and that this decline has contributed to the current achievement gap in mathematics.

REFERENCES

- BEATON, A. E., MULLIS, I. V. S., MARTIN, M. O., GONZALEZ, E. J., KELLY, D. L., & SMITH, T. A. (1996). *Mathematics achievement in the middle school years: IEA's third international mathematics and science study (TIMSS)*. Chestnut Hill, MA: Boston College.
- BOISSIERE, M., KNIGHT, J. B., & SABOT, R. H. (1985). Earnings, schooling, ability, and cognitive skills. *American Economic Review*, *75*, 1016-1030.
- COHEN, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- EKSTROM, R. B., FRENCH, J. W., & HARMAN, H. H. (1976). *Manual for kit of factor-referenced cognitive tests 1976*. Princeton, NJ: Educational Testing Service.
- FRY, A. F., & HALE, S. (1996). Processing speed, working memory, and fluid intelligence: Evidence for a developmental cascade. *Psychological Science*, *7*, 237-241.
- GEARY, D. C. (1995). Reflections of evolution and culture in children's cognition: Implications for mathematics development and instruction. *American Psychologist*, *50*, 24-37.
- GEARY, D. C., FAN, L., & BOW-THOMAS, C. C. (1992). Numerical cognition: Loci of ability differences comparing children from China and the United States. *Psychological Science*, *3*, 180-185.
- GEARY, D. C., SALTHOUSE, T. A., CHEN, G.-P., & FAN, L. (1996). Are East Asian versus American differences in arithmetical ability a recent phenomenon? *Developmental Psychology*, *32*, 254-262.
- HUSÉN, T. (1967). *International study of achievement in mathematics: A comparison of twelve countries* (Vols. 1-2). New York: Wiley.

- KAIL, R. (1991). Developmental change in speed of processing during childhood and adolescence. *Psychological Bulletin*, **109**, 490-501.
- LEE, D., & CHEN, G.-P. (1989). *Raven's Progressive Matrices (CRT): Chinese version*. Shanghai, China: East China Normal University.
- LYNN, R. (1982). IQ in Japan and the United States shows a growing disparity. *Nature*, **297**, 222-223.
- RAVEN, J. C., COURT, J. H., & RAVEN, J. (1993). *Manual for Raven's Progressive Matrices and Vocabulary Scales*. London: H. K. Lewis.
- RIVERA-BATIZ, F. (1992). Quantitative literacy and the likelihood of employment among young adults in the United States. *Journal of Human Resources*, **27**, 313-328.
- RUSHTON, J. P. (1995). *Race, evolution, and behavior: A life history perspective*. New Brunswick, NJ: Transaction.
- SALTHOUSE, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, **103**, 403-428.
- SCHAE, K. W. (1996). *Intellectual development in adulthood: The Seattle longitudinal study*. New York: Cambridge University Press.
- STEVENSON, H. W., CHEN, C., & LEE, S.-Y. (1993). Mathematics achievement of Chinese, Japanese, and American children: Ten years later. *Science*, **259**, 53-58.
- 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.
- STEVENSON, H. W., STIGLER, J. W., LEE, S.-Y., LUCKER, G. W., KITAMURA, S., & HSU, C. C. (1985). Cognitive performance and academic achievement of Japanese, Chinese, and American children. *Child Development*, **56**, 718-734.

NOTES

1. In the U.S. sample, 86.6% were Anglo-American, 3.2% were African American, 5.7% were Asian (educated in the United States), and the remaining participants were Hispanic or identified themselves as being of mixed race.
2. All identified students who wished to participate were included in the study, the only restriction being an IQ score > 99.
3. There were only a few significant sex differences on the arithmetic tests ($ps < .05$). All analyses involving these tests were redone using sex as a covariate and yielded no changes in the pattern of results.

(Manuscript received March 7, 1997;
revision accepted for publication April 14, 1997.)