Supplementary Information for Halberda, J., Mazzocco, M., & Feigenson, L. (2008). Individual differences in nonverbal number acuity correlate with maths achievement. *Nature*.

Supplementary Methods

Subjects

Subjects were drawn from an ongoing longitudinal study of mathematics achievement [15]. The initial sample (n=249) was recruited in Kindergarten from a large suburban public school district. Within this district, seven target schools were selected on the basis of having relatively low indices of student mobility and low indices of free or reduced price lunch enrolment (as a filter for socioeconomic status). Thus participating schools were representative of a diverse range of socioeconomic areas within the district with the exception of the lowest. This deliberate omission served to diminish the likelihood of poverty-associated poor maths achievement in the longitudinal study of mathematics ability [28].

All Kindergartners attending the participating schools were invited to enroll, except for children with mental retardation or limited English proficiency. A total of 57% of eligible subjects enrolled, resulting in 249 students (120 boys) in the initial sample. Subjects were tested annually, with decreasing enrolment over time due to attrition. Eighty students from the original cohort who were still being tracked in the 9th Grade agreed to participate in the present study. Of these 80, 7 subjects did not have complete data from Years 1 to 7, and so were omitted from our analyses. For 9 subjects, performance in the numerical discrimination task was too variable and the psychophysics model did not settle on a least squares fit for the parameter *w*. Data from these subjects were excluded from the analyses. The final sample included 64 children (32 boys), with a mean age of 14 years 10 months (range 14 years 3 months to 15 years 9 months), who were enrolled in 8th (n=5) or 9th Grade (n=59) at the time of testing. Most of the subjects were right handed (89%), and most were Caucasian (89%). Preliminary analyses revealed that the 64 subjects in the present study were representative of the larger sample of 249 [29].

Procedure

Each subject was tested individually during each year of the longitudinal study by one or two female examiners. During Kindergarten, all testing was completed in the subject's school, away from the classroom in a room occupied by only the child and the examiner(s). During Years 1 to 7 of the longitudinal study the testing battery was divided among two or three testing sessions ranging from 45 to 90 minutes each, with shorter sessions used in primary versus middle school. The numerical discrimination task was administered during a school or lab visit as a portion of a single 60 to 75 minute session in Year 10 only (i.e., 8th-9th Grade). During all years of the study, individual tests were presented in a fixed order except under circumstances beyond the examiner's control (e.g., school fire drill, illness).

Relevant Measures administered during Years 1 to 7

Not all measures were administered each year as determined by developmental appropriateness of a given test (e.g., the *Test of Early Mathematics Ability*) or time constraints for the year in question. The major objective of the longitudinal study was to examine potential cognitive correlates of mathematics achievement, so the primary variables of interest included general measures of mathematics ability, mathematics achievement, reading achievement, phonological retrieval, working memory, and spatial ability. These tasks, described below, included a combination of published standardized tests and experimental measures. Unless otherwise noted, performance on the standardized tests is converted to an age-referenced score based on a mean of 100, SD=15.

Standardized Measures of Mathematics

The *Test of Early Mathematical Ability* – Second Edition (TEMA-2) is a standardized test normed for use with children from 4 to 8 years of age [17]. We administered the TEMA-2 during the first four years of the study. Depending on grade level at testing, the TEMA-2 requires counting aloud, number line concepts, exact numerical computation, and mastery of maths facts. Items on the TEMA-2 tap both formal maths skills that are related to explicit school instruction (e.g., addition algorithms) and informal skills typically not addressed by the school curriculum, such as cardinality. After Grade 3 the TEMA-2 was not included because it is designed for use with younger children.

The *Woodcock Johnson – Revised* is a widely used standardized test of academic achievement normed for use with subjects age 5 years and older [18]. The Mathematics Calculation subtest (WJ-Rcalc) is a paper and pencil task comprised of problems presented in order of increasing difficulty. The WJ-Rcalc requires exact computation using addition, subtraction, multiplication, and/or addition, with whole numbers, rational numbers, and variables. Test problems administered in Kindergarten to 6th Grade range from one digit addition to algebra. This subtest was administered every year from Kindergarten through 6th Grade, except at 2nd Grade when it was omitted due to time constraints.

General Intelligence Measure

The *Wechsler Abbreviated Scale of Intelligence*, Full Scale Intelligence (WASI-full) is a widely used standardized test of cognitive ability [20] that is normed for subjects ages 6 to 89 years. We administered the WASI during 3rd Grade only, when students were 8 years old. Correlations between early and later IQ scores are less stable before 7 years of age [30] so the WASI was not administered before 3rd Grade. Also, it was not administered during subsequent years of the study due to time constraints, and in view of the demonstrated stability of Wechsler scores over time [31]. The WASI includes two verbal subtests (Vocabulary and Similarities) and two nonverbal subtests (Block Design and Matrix Reasoning), each pair of which is used to derive an age-referenced domain IQ score for Verbal or Performance ability, respectively (WASI-verbal, WASI-performance). In turn, each of these domain scores contributes to the full-scale IQ score (WASI-full). We used WASI-full as our measure of estimated general intelligence and included separate WASI-verbal and WASI-performance scores in our "omnibus" correlation analysis (Table 3 of Main Text).

Task Demands Measure

We used the colour subtest of the *Rapid Automatized Naming* test (RAN-colour) as a measure of task demands [19]. This lexical retrieval task requires subjects to rapidly name the colour of a series of squares. During an un-timed warm-up trial, subjects named the colours of five solid-coloured 12 mm squares (red, black, green, yellow, and blue) printed on one white 8 x 11 inch sheet of paper. For the test trial, subjects saw another 8 x 11 inch sheet of white paper containing 5 rows of 10 coloured squares. Subjects named the colours in serial order as fast as possible, while the experimenter recorded overall response time (RT) with a hand-held stopwatch. A shorter RT indicated better performance.

Additional Measures

General Cognitive Ability

We included measures of Performance IQ and Verbal IQ derived from the *Wechsler Abbreviated Scale of Intelligence* (WASI-performance, WASI-verbal) [20] administered in third grade. The methods are given above under the subheading "General Intelligence Measure."

Executive Functions

Cognitive Switching. We used the one-attribute switching task from the *Contingency Naming Test* (CNT) to assess cognitive flexibility and performance efficiency under moderate working memory demands. The CNT, described elsewhere in detail [15,32], requires naming the colour or shape of a series of stimuli according to different rules. The stimuli were approximately 1-inch shapes, outlined in black with solid-colour interiors. Each stimulus also contained a smaller, inner shape (e.g., a solid pink 1" high triangle in which a 1/4" square appears). Level A rules present no working memory demands because they require merely naming the colour (Rule A1) or the larger outer shape (Rule A2) of each stimulus. These tasks serve as a warm up to the switching task introduced by Level B rules. Level B rules present working memory and inhibitory control demands because they require switching between naming the stimuli by colour or by shape, depending on one attribute (whether the inner and outer shapes of the stimulus match or not) or two attributes. In our earlier work, we found that all 3rd Graders in the longitudinal study successfully completed the practice trial for the one attribute task (CNT-B3), and that 3rd Grade performance on this task correlated with 3rd Grade TEMA-2 and WJ-R-calc scores [15]. We also found that several 3rd Graders failed the two-attribute task. Therefore, only the one-attribute task, CNT-B3, was included as a measure of executive function in the present study.

For each CNT task, subjects saw one row of 9 stimuli on an 8" x 11" sheet of white paper, which they verbally labelled during an un-timed warm-up trial. Once it was established that the rule had been learned, subjects were presented with three rows of 9 stimuli on an 8" x 11" sheet of white paper. Subjects named the stimuli as quickly as possible while the experimenter recorded reaction time (RT) using a hand-held stopwatch. The variable of interest is performance efficiency, measured by assessing speed-accuracy trade-off via the following formula:

$$\text{Efficiency} = [(1/RT)/ \sqrt{(\text{errors} + 1)}\,] * 100$$

Visual Working Memory. We used the *Memory Puzzle Test* (MemPuzl) [33] to measure working memory for spatial location following repeated exposure to a fixed visual array. This task followed presentation of the figure ground subtest (a subtest of the DTVP-2, see below). During the figure ground subtest subjects saw 11 pages of stimuli, one page at a time, each of which included a target array of 10 unique figures presented in a boxed 2 X 5 format. The identical array of figures appeared on the bottom half of all 11 pages. Immediately after all 11 pages were presented, the Memory Puzzle Task was administered as follows: The subject was given a cut-out copy of the boxed array, minus the individual shapes, and was also handed the ten individual figures that had been individually cut out and laminated. The subject's task was to reproduce the correct location of the figures, with a maximum "correct location" score of 10 points recorded by the examiner.

Standardized Measures of Visual Spatial Skills

The *Developmental Test of Visual Perception Second Edition* (DTVP-2) [34] includes four motor-reduced subtests, each of which was administered during the first five years of the study. It is normed for use with children ages 4 to 10 years. During the *Figure-Ground* subtest (DTVPfg) subjects identified familiar individual shapes from designs constructed with overlapping and/or embedded shapes. During the *Form Constancy* subtest (DTVPfc), subjects identified two matching shapes (from five or more shapes) differing in size, colour, and/or orientation. During the *Visual Closure* subtest (DTVPvc), subjects identified which of several degraded shapes matched a visually intact target shape. During the *Position in Space* subtest (DTVPps) subjects matched one of four or more figures to a target figure on the basis of spatial orientation. For each subtest we relied on the total number of correct responses. In our earlier work we found that performance scores on select subtests correlate with concurrent maths performance during elementary school [33].

The *Beery-Buktenica Developmental Test of Visual-Motor Integration Fourth Edition* (VMI) [35] is a widely used paper-and-pencil task used to assess children's ability to coordinate visual perception and motor planning. It is normed for use with subjects 2 years of age through adulthood. During the VMI subjects copied individual geometrical figures presented in order of increasing complexity. We relied on the total number of correct responses. Subjects were allowed only one attempt per drawing, and were not given the opportunity to erase or otherwise correct their drawing.

Reading Related Measures

The *Woodcock Johnson – Revised*, reading subtests (WJ-Rlwid, WJ-Rwa) [18] are widely used standardized tests of achievement. The Letter-Word Identification subtest (WJ-R lwid) involves single-word reading or, for younger children, letter symbol recognition. The Word Attack subtest (WJ-Rwa) assesses subjects' ability to decode unfamiliar words and read them aloud (e.g., non-words such as "flep"). For both of these subtests, raw scores were converted into age-referenced standard scores.

The test of *Rapid Automatized Naming* (RAN-number, RAN-letter) [19] is a measure of single word retrieval fluency. The method is identical to the RAN-colour subtest described above in the section "Task Demands Measure." The RAN-number involves naming single digits (e.g., 2, 6, 9, 4, and 7), and the RAN-letter involves naming single letters (e.g., a, p, d, o, s). The variable of interest is response time (RT) to complete the task, with a shorter RT indicating better performance.

Conclusions

Correlations between ANS acuity and all these measures appear in Table 3 of the Main Text. A number of the measures described above have previously been found to correlate with math achievement (e.g., RAN-color, WASI, Executive Functions and Working Memory), while others control for performance and general intelligence factors in our task measuring ANS acuity. That ANS acuity correlates with math achievement even with all of these measures controlled for demonstrates both the robustness of this correlation, and that ANS acuity captures a unique segment of the variance in math achievement throughout the school age years.

Supplementary Notes

28. Leventhal, T., Brooks-Gunn, J. (2004). A Randomized Study of Neighborhood Effects on Low-Income Children's Educational Outcomes.__*Developmental Psychology* 40, 488-507.]

29. Separating the sample into the 64 subjects who participated in the present analyses and the 185 subjects who did not, chi square analyses revealed no difference in the relative portion of males to females in these two groups of participants, nor in the relative proportion of left- vs. right-handed individuals (all *p*s > .74). There was no age difference between the groups at time of testing for any year of the longitudinal assessment (*p* ranged from .18 to .97). We also compared the two groups on relevant measures (TEMA-2, WJ-Rcalc, WASI-full, and RAN-colour) from Year 4 of the longitudinal study (3^{rd} Grade). Although the difference in IQ (WASI-full) approached significance, both groups had a mean score in the average range. Subjects included in the present study (n=64) had a slightly higher mean IQ score (111.8, SD=14.6, range = 78-142) than did the 185 subjects who did not participate (mean score = 107.8, SD=13.4, range = 77-142), $t = 1.903$, $p = .053$. The high average mean score in both groups results, at least in part, from the exclusion of low functioning subjects from study recruitment. These individuals were excluded by targeting only students enrolled in regular classrooms. We found no differences between the two groups on TEMA-2 ($p = .07$), WJ-Rcalc ($p = .20$) and RAN-colours ($p = .62$) in Year 4 of the longitudinal study.

30. Sattler, J. M. (2001). *Assessment of Children: Cognitive Applications (4th ed.).* San Diego: Jerome M. Sattler, Publisher, Inc.

31. Canivez, G. L., & Watkins, M. W. (1998). Long-term stability of the Wechsler Intelligence Scale for Children--Third Edition. *Psychological Assessment*, 10, 285-291.

32. Anderson, V. A., Anderson, P., Northam, E., & Taylor, H. G. (2000). Standardisation of the contingency naming test (CNT) for school-aged children: A measure of reactive flexibility. *Clinical Neuropsychological Assessment*, 20, 385-406

33. Mazzocco, M. M. M., Bhatia, N. S., & Lesniak-Karpiak, K. (2006) . Visuospatial skills and their association with math performance in girls with fragile X or Turner syndrome. *Child Neuropsychology, 12*, 87-110.

34. Hammill, D.D., Pearson, N.A., and Voress, J.K. (1993). *Examiner's Manual. Developmental Test of Visual Perception Second Edition.* Austin, TX: Pro-Ed.

35. Beery, K.E. (1997). *Administration Scoring and Teaching Manual. The Beery-Buktenica Developmental Test of Visual-Motor Integration Fourth Edition*. Modern Curriculum Press.