Intelligence 39 (2011) 89–99

Contents lists available at ScienceDirect

Intelligence

A simple syllogism-solving test: Empirical findings and implications for *g* research

Chizuru Shikishima [a,*,1], Shinji Yamagata [a], Kai Hiraishi [b], Yutaro Sugimoto [a], Kou Murayama [c], Juko Ando [d]

[a] *Keio Advanced Research Centers, Keio University, Tokyo, Japan*
[b] *Kokoro Research Center, Kyoto University, Kyoto, Japan*
[c] *Department of Psychology, University of Munich, Munich, Germany*
[d] *Faculty of Letters, Keio University, Tokyo, Japan*

ARTICLE INFO

Article history:
Received 19 October 2010
Received in revised form 18 January 2011
Accepted 19 January 2011
Available online 3 March 2011

Keywords:
Syllogism-solving
g
Intelligence test
Twin study
Behavioral genetics

ABSTRACT

It has been reported that the ability to solve syllogisms is highly *g*-loaded. In the present study, using a self-administered shortened version of a syllogism-solving test, the *BAROCO Short*, we examined whether robust findings generated by previous research regarding IQ scores were also applicable to *BAROCO Short* scores. Five syllogism-solving problems were included in a questionnaire as part of a postal survey conducted by the Keio Twin Research Center. Data were collected from 487 pairs of twins (1021 individuals) who were Japanese junior high or high school students (ages 13–18) and from 536 mothers and 431 fathers. Four findings related to IQ were replicated: 1) The mean level increased gradually during adolescence, stayed unchanged from the 30s to the early 50s, and subsequently declined after the late 50s. 2) The scores for both children and parents were predicted by the socioeconomic status of the family. 3) The genetic effect increased, although the shared environmental effect decreased during progression from adolescence to adulthood. 4) Children's scores were genetically correlated with school achievement. These findings further substantiate the close association between syllogistic reasoning ability and *g*.

1. Introduction

If all humans are mortal,
and all Greeks are humans,
then all Greeks are mortal.

Aristotle, describing this classic syllogism in Prior Analytics in 350 B.C., shed light on an aspect of human mental ability, namely, logical deductive reasoning. Syllogisms are a form of argument relating three terms that consist of two premises and a conclusion. In ancient and medieval Europe, the ability expressed in syllogism solving was considered to be at the heart of human logical thinking (Bochenski, 1970; Kneale & Kneale, 1962). Before the development of the arithmetical methods necessary for quantitative science, the syllogism was a required tool for man "as a means to understanding in whatever field of human intellectual endeavor he had chosen" (Wetherick, 1989, p. 111). If that is true, comprehension and use of syllogisms can be expected to reflect human general intelligence, *g* (Spearman, 1904).

1.1. Syllogism solving and general intelligence

In fact, several studies since the 1960s have shown that individual differences in performance on syllogism solving are associated with individual differences in intelligence. Classic intelligence tests such as the California Test for Mental

* Corresponding author at: 1-40-30 Naritahigashi, Suginami-ku, Tokyo 166–0015, Japan. Tel./fax: +81 3 5307 5722.
E-mail address: shikishima@kts.keio.ac.jp (C. Shikishima).
[1] Organization: Keio University, 2-15-45 Mita, Minato-ku, Tokyo 108–8345, Japan.

0160-2896/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.intell.2011.01.002

Maturity, which postulate syllogistic reasoning ability as a factor of human intelligence, contain syllogism-solving problems in their subtests (Beaumariage, 1960). Frandsen and Holder (1969), using university students, pointed to a relationship between performance on syllogism solving and performance on tests of spatial ability as measured by the Differential Aptitudes Test ($r = .56$). Guyote and Sternberg (1981) reported significant correlations ($r = .35$ to .60) for adults between performance on syllogism-solving problems and a spatial-abstract ability factor extracted from subtests of the Differential Aptitudes Test. Bickersteth and Das (1981) revealed that performance on a syllogism-solving test was better for children with a high score as compared with a low score on the Raven Progressive Matrix. Bara, Bucciarelli, and Johnson-Laird (1995) reported that 31% of the variance in performance on syllogism solving was explained by the score on the perception of identities test, and a further 8% of the variance was explained by working memory test score in children aged 9–10, adolescents, and adults. Stanovich and West (1998, 2000) showed that syllogism-solving performance in undergraduate students was associated with Scholastic Achievement Test scores ($r = .41$–.47), as well as composite cognitive ability scores, including the Raven Matrices and reading comprehension tests ($r = .50$). These findings imply that logical deductive reasoning ability, as measured by syllogism solving, is substantially correlated with general ability, *g*.

Shikishima, Hiraishi, et al. (2009) reported that logical deductive reasoning ability as measured by 100 syllogism-solving problems indicated *g*, with verbal and spatial abilities measured using an intelligence test. Empirical evidence based on adult twin datasets was provided showing that the common pathway model, a multivariate genetic model with a common latent factor (i.e., *g* as a single higher-order construct) fit better than the independent pathway model, a multivariate genetic model without a common latent factor, when three distinct mental abilities – verbal, spatial, and logical deductive reasoning – were jointly analyzed. The factor loading from *g* for performance on syllogism solving was as high as .73, which was equivalent to or even higher than verbal (.72) and spatial (.69) abilities. The findings clearly demonstrated, from psychological and behavioral genetic points of view, that syllogism solving is closely related to *g*.

However, previous studies claiming an association between syllogism solving and intelligence are limited to describing the magnitude, or numerical value, of the extent to which the score on syllogism solving is quantitatively correlated with the score on another measure of *g*. Therefore, it is still premature to acknowledge that syllogism solving is a marker for *g*. To further clarify the close association of the two, it is necessary to provide empirical evidence that substantive results obtained from syllogism solving correspond to those obtained for *g*.

1.2. The present study

Given the background described thus far, the present study explored to what extent performance on syllogism solving would reproduce findings generated by previous research regarding *g*. Evidence confirming the profound association between syllogism solving and *g* allowed us to predict that scores on syllogism solving would reflect results similar to typically observed as characteristics of *g*. If so, the argument, based on the degree of correlations between the two, that syllogism solving is an indicator for *g* would be further strengthened and become more persuasive.

To examine the validity of the prediction, we developed a simplified syllogism-solving test, the *BAROCO Short,* constructed from a significantly shortened version of the 100 problems of syllogism solving, the *BAROCO* II.3 (Shikishima, Hiraishi, et al., 2009; see Ando et al., 2006, for test description and development history). The five problems that showed the highest correlation with IQ were selected from the *BAROCO* II.3 and constituted the *BAROCO Short*.

One advantage of syllogism-solving tests is that the correct answer to syllogism-solving problems cannot easily be found. That is, unsupervised respondents cannot cheat as easily as on verbal or numerical tests where correct answers are easily accessible at home. Therefore, syllogism-solving tests can be self-administered outside a controlled setting.

Another advantage of syllogism-solving tests is that effective data can be obtained from participants of a wide age range using the same set of problems, which allows us to compare performance directly between generations. Bara et al. (1995) administered the same syllogism-solving problems to children aged 9–10, adolescents, and adults. They confirmed that children as young as age 9 were able to draw valid conclusions in solving categorical syllogisms at a level better than chance.

These advantages unique to syllogism-solving tests, as well as the significant reduction in the number of problems, enabled the measurement to be conducted by postal survey. Hence, quantitative data could be relatively easily collected from a comparatively large sample ranging from children to adults.

We tested the following four hypotheses consisting of robust findings regarding IQ scores.

Hypothesis 1. The mean score changes systematically according to age.

In the context of G_f–G_c theory, Horn and Hofer (1992) found that the peak of fluid ability, G_f, is reached in a person's 20s, although crystallized ability, G_c, continues to develop up to one's 40s. Kaufman (2001) examined this differential age-related developmental curve between G_f and G_c by conducting cross-sectional and longitudinal analyses on Performance IQ and Verbal IQ of the WAIS III in individuals aged 16–89. After controlling for the cohort effect of educational attainment, results showed that Verbal IQ (G_c) was maintained throughout the adult life span and declined only during the 80s. However, Performance IQ (G_f) peaked in young adulthood, then declined steadily and dramatically throughout the adult life span.

If the *BAROCO Short* measures *g*, a curve representing the mean level change with age was expected to appear. However, indicating the integration of G_f and G_c, the curve would follow neither the one for G_f nor G_c, but would rather exhibit a trajectory combining the two. This prediction was based on the finding that the ability to solve 100 syllogism problems correlated equivalently with verbal IQ and spatial

IQ ($r = .54$ and .50; Shikishima, Hiraishi, et al., 2009). In the present study, performance on the *BAROCO Short* was directly compared in a wide age range from adolescence through adulthood.

Hypothesis 2. The score can be predicted by socioeconomic status (SES).

A robust positive correlation between SES and IQ has been reported in research in the West (Herrnstein & Murray, 1994; Heckman & Krueger, 2003). In Japan, although no published literature has reported an association between IQ and family SES, a positive correlation between family SES and children's school achievement or educational attainment has been shown in the fields of sociology and economics (Kariya, 2001; Tachibanaki, 2005). Given that IQ, school achievement, and educational attainment are closely intercorrelated (Shikishima, Ozaki, Hiraishi, Sasaki, & Ando, 2009), the family's SES and children's IQ as well as parents' IQ are assumed to be correlated in Japan. If so, for parents and children, performance on the *BAROCO Short* should be significantly associated with the SES of the family.

Hypothesis 3. The genetic effect increases although the shared environmental effect decreases during the progression from adolescence to adulthood.

A robust finding from behavioral genetic studies previously reported is that intrafamily resemblance in IQ scores is attributed to both genetics and shared environment in childhood or early adolescence; only genetics exert an effect after late adolescence (Plomin, DeFries, McClearn, & McGuffin, 2008). Meta-analyses of published twin study correlations for IQ and genetic and environmental contributions to IQ in accordance with age groups clearly demonstrated that from late adolescence to young adulthood the shared environmental effects on IQ drastically declined, and genetic effects increased instead (McCartney, Harris, & Bernieri, 1990; McGue, Bouchard, Iacono, & Lykken, 1993; Bergen, Gardner, & Kendler, 2007). Thus, late adolescence can be considered a transitional period for IQ formation, in terms of its underlying genetic and environmental structure. We predicted that *BAROCO Short* scores would exhibit relatively lower genetic effects and higher shared family environment effects in early adolescence, and higher genetic effects and lower shared family environment in late adolescence.

Hypothesis 4. The score is genetically correlated with other measures of cognitive ability.

Factor *g* refers to, by its definition, phenotypic covariation among diverse measures of cognitive ability as indexed by an unrotated first principal component score (Jensen, 1998). The phenotypic covariation observed should be yielded for underlying etiologies (i.e., either genetic or environmental, or both) common among measures. Behavioral genetics researchers have argued with substantial empirical evidence that the overlap is largely attributable to genetics. A plausible explanation is provided by the generalist genes hypothesis that genes contributing to diverse, distinct human abilities are pleiotropic (i.e., attributable to the same function in the same set of genes; Kovas & Plomin, 2006). Therefore, if the *BAROCO Short* assesses *g*, substantial genetic overlap with other measures of cognitive ability should be detected.

To conduct these hypothesis tests, we used genetically informative samples that covered wide life stages, including twins from early adolescence to late adolescence and their parents.

2. Method

2.1. Participants

All the data used in this study were collected through a cross-sectional postal survey conducted by the Keio Twin Research Center (KoTReC). The survey was designed to investigate sources of socialization in adolescence by surveying Japanese junior high and high school twins and their parents.

Prior to the survey, a twin registry was constructed by gathering the names of two family members who lived together and had the same dates of birth between April 1990 and March 1996. We used official residential records from city halls in Tokyo and three neighboring prefectures. Among those pairs extracted, the KoTReC randomly selected 2828 families (35% of the registry) and conducted an anonymous postal survey in 2009, which provided the dataset in the present study. A set of four booklets of questionnaires were sent to each family: two for twin pairs, one for fathers, and the other one for mothers. We asked participants to refrain from discussing the questions while answering and to close the booklet with attached seals immediately after they filled out the questionnaire.

The number of families from which at least one member provided an effective response was 569, constituting a participation rate of 23.4% when the number of undelivered mails was taken into account. Syllogism-solving data, included in each questionnaire, were collected from 487 pairs of twins (98 identical male, 151 identical female, 63 fraternal male, 67 fraternal female, and 108 fraternal male–female; ages 13–18, $M = 15.3$), and from 536 mothers (ages 34–62, $M = 46.2$) and 431 fathers (ages 36–68, $M = 48.7$), yielding 1988 individuals in total.

The zygosity of the same-sex twin pairs was diagnosed by a 3-item scale based on physical resemblance developed by Ooki, Yamada, Asaka, and Hayakawa (1990). This scale is a translated version of Torgersen's questionnaire for zygosity diagnosis (Torgersen, 1979). The items are as follows: “Are you and your co-twin as alike as two peas in a pod?”, “Were you and your co-twin confused for each other as children?” and “If so, who confused one for the other?” These items were included in the questionnaire, and each twin pair responding to the questionnaire was classified as either identical or fraternal, in accordance with the cut-off point established for this scale. We estimate the accuracy rate of the diagnoses to be 93% based on the agreement rate between questionnaire-based diagnoses and DNA-based diagnoses we previously obtained from another twin sample from the KoTReC (Shikishima, Ando, Ono, Toda, & Yoshimura, 2006). Blood type (A, B, AB, or O) was reported in the questionnaire, and those twin pairs in disagreement were diagnosed as fraternal twins.

2.2. Materials

2.2.1. Syllogism solving

The same five syllogism-solving problems were included in the questionnaire for twin children, mothers, and fathers. These five problems were selected as those best predicting full IQ scores of 448 individuals (a different sample from the present study) aged 17 to 36 who were tested both on a full IQ test (the *Kyodai Nx15-*) and on the 100-problem categorical syllogism-solving test (the *BAROCO* II.3) in a previous survey conducted by the KoTReC (Shikishima, Hiraishi, et al., 2009).

The five problems were selected from Belief-neutral, Belief-congruent, and Belief-incongruent *formats* of the *BAROCO* II.3 (Ando et al., 2006). In these formats, all the terms are stated in common words, but the conclusion describes a particular situation the truth of which is determined by the context. In the Belief-neutral format, the conclusions are independent of people's belief systems. In the Belief-congruent format, the conclusions are congruent with people's beliefs, whereas in the Belief-incongruent format, the conclusions are incongruent with people's beliefs. We did not include the problems from Graphical and Abstract formats where graphical representation and abstract letters (A, B, and C) are used, respectively, because inclusion of those formats might have offered respondents superfluous hints for solving problems.

Variations in levels of difficulty in solving syllogisms can be added by including different *types* of syllogisms. The types consist of combination of *moods* in both the premises and the conclusions in accordance with the Aristotelian categorical syllogism classification (Johnson-Laird, 1983). The four moods are as follows: universal affirmative (A), universal negative (E), particular affirmative (I), and particular negative (O). It has been reported that the EIO type of reasoning is more strongly correlated with spatial IQ than verbal IQ, while AAA/AII/AEE types of reasoning are more strongly correlated with verbal IQ than spatial IQ (Ando et al., 2006).

Moreover, in the premises of syllogisms there are three terms. The subject of the conclusion is called the minor term (described as S), the term of the predicate part of the conclusion is called the major term (P), and the third term is called the middle term (M), which appears in both premises. Combinations of the order of S, P, and M in the two premises generate four possible patterns called *figures*, and reversing the order of the two premises doubles the number of possible syllogisms, which provides further sources of difficulty in solving problems (Johnson-Laird, 1983; Evans, Newstead, & Byrne, 1993; Bara et al., 1995; Arend et al., 2003).

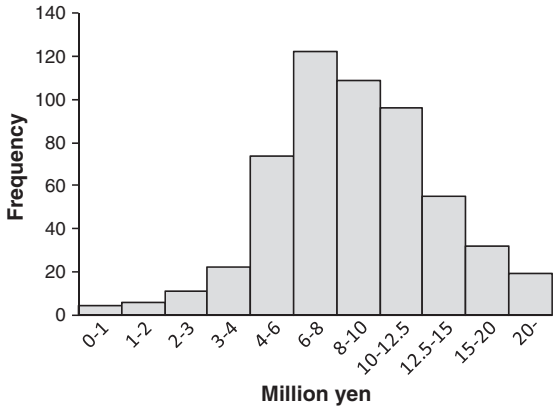


Fig. 1. Distribution of family income.

Of the five problems selected out of the 60 problems in the *BAROCO* II.3, four were from the Belief-incongruent format, and one was from the Belief-neutral format (Table 1). Two were the EIO type, two were the AII type, and one was the AEE type. There were three from the M–P S–M figure (*Figure* I; Johnson-Laird, 1983) and two from the P–M M–S figure (*Figure* IV). The correlation coefficient for the full IQ for the five problems ranged from .37 to .41. The problem that exhibited the highest correlation was the following (see Problem 1 in Table 1):

Some parrots are Babaros.
No Babaros are birds.
Therefore,

1. *Some parrots are birds.*
2. *All parrots are birds.*
3. *Some parrots are not birds.*
4. *No parrots are birds.*
5. *No valid conclusion.*

Following the two premises above, respondents were required to choose the most appropriate answer from the alternatives. As in the *BAROCO* II.3, the number of answer alternatives was five, which comprised four possible conclusions and a "no valid conclusion". Because the principal component analysis revealed a clear one-component solution with roughly equal loadings for the five problems (Table 2), the total score of the five problems (1 point for correct answers and 0 points for incorrect answers) ranging from 0 to 5 was used in the present study.

Table 1
Descriptions of the five problems.

Problem	r_{IQ}	Format	Type			Figure			Classification
			First premise	Second premise	Conclusion	First premise	Second premise	Conclusion	
1	.41	Belief-incongruent	I	E	O	S–M	M–P	S–P	EI1*O
2	.39	Belief-neutral	E	I	O	M–P	S–M	S–P	EI1O
3	.39	Belief-incongruent	A	I	I	M–P	S–M	S–P	AI1I
4	.39	Belief-incongruent	E	A	E	M–S	P–M	S–P	AE4*E
5	.37	Belief-incongruent	A	I	I	M–S	P–M	S–P	IA4*I

A: Universal affirmative E: Universal negative I: Particular affirmative O: Particular negative.
S: Minor term P: Major term M: Middle term.
* Denotes that the order of the major and minor premises is changed, although this makes no difference logically.

Table 2
Principal component analysis.

Problem	Component 1
1	.62
2	.60
3	.75
4	.64
5	.58
Eigenvalue	2.1
Variance explained	41%

Reliability of the scale in terms of internal consistency represented by Cronbach's α (based on tetrachoric correlations) of the five problems of the *BAROCO Short* was .77 for children (twins), .76 for mothers, and .72 for fathers.

When we looked at the *BAROCO* II.3 data used in Shikishima, Hiraishi, et al. (2009) the correlation between the factor score of the 100 problems in the *BAROCO* II.3 and the five problems in the *BAROCO Short* was .84. The correlation between the total score of the five problems in the *BAROCO Short* and the full IQ score was .54. A multiple regression analysis revealed that the five problems explained a third of the full IQ score variation (adjusted $R^2 = .33$).

In the postal survey, we wrote explicit instructions directing respondents to answer each problem within 1 min.

2.2.2. SES

As an index of SES, parental educational attainment was ascertained by asking both mothers and fathers their highest level of education achieved. A 6-point scale was used: 1 point for elementary school or junior high school, 2 points for high school, 3 points for technical college, 4 points for junior college, 5 points for university, and 6 points for graduate school. The distribution of educational attainment for mothers and fathers is shown in Table 3.

Our sample was biased towards higher education, with 22% of the mothers and 58% of the fathers having completed university or graduate school. According to the published report of the 2000 Population Census of Japan, 9.2–14.0% of females and 29.3–36.8% of males born in 1951–1970 (during which more than 95% of mothers and fathers in our sample were born) completed university or graduate school.

Family income in the previous year was reported on an 11-point scale (from 1 point for 0–1 million yen to 11 points for more than 20 million yen) from both mothers and fathers, and parental mean scores were used in the analysis. According to the Comprehensive Survey of Living Conditions of the People on Health and Welfare in 2006 conducted by the Ministry of Health, Labour and Welfare in Japan, average yearly family income was 6.998 million yen for individuals in their 40s and 7.346 million yen for those in their 50s. Compared with these national standards, our sample may have been biased towards the wealthier, but only slightly, with the distribution being close to average (Fig. 1).

Table 3
Distribution of educational attainment for mothers and fathers.

Highest level of education achieved	Mothers		Fathers	
	Frequency	Percent	Frequency	Percent
Junior high school	9	1.7	20	4.7
High school	168	31.5	107	25.1
Technical college	87	16.3	41	9.6
Junior college	154	28.8	11	2.6
University	109	20.4	216	50.6
Graduate school	7	1.3	32	7.5
Total	534	100.0	427	100.0
Missing	2	–	4	–

2.2.3. Cognitive abilities

As measures of cognitive ability, twin children's school achievements in English and Math were reported by mothers and fathers separately on a 5-point scale for each of the twins, and the parental mean score for each of the twins was used in the analysis. The correlation between maternal and paternal evaluation was .77 for English and .73 for Math.

The information regarding SES and cognitive abilities was reported in the same questionnaire containing the *BAROCO Short*.

2.3. Statistical analyses

For test of Hypothesis 1, the mean score for the *BAROCO Short* was compared in relation to age. Scores for twins were categorized into six one-year age groups ranging from an age 13 group to an age 18 group. Scores for mothers and fathers were jointly categorized into six age groups: 34–39, 40–44, 45–49, 50–54, 55–59, and 60–68.

To compare the performance among the six age groups comprising mothers and fathers, scores with an adjustment for level of educational attainment were used for the following reasons. First, the importance of controlling for a cohort effect of level of education has been pointed out in cross-sectional data where the mean level change in intelligence was examined (Horn & Hofer, 1992; Kaufman, 2001). In Japan, the standard level of educational attainment has been rising substantially during past decades because of educational expansion, which causes the level of formal education for younger individuals to be higher than that for those who are older (Kariya, 2001; Shikishima, Ozaki, et al., 2009). Second, all members of our adult sample were parents of children aged 13–18. It is assumed that the young parents in our sample, such as those in their 30s, were more likely to have married younger and thus had less formal education. Such a sample bias could contaminate the examination of change in intelligence level based on age groups.

In carrying out test of Hypothesis 2, to describe the impact of parental educational attainment on *BAROCO Short* scores, simple linear regression analyses were conducted using *BAROCO Short* scores for parents and children as dependent variables and educational attainment for parents as independent variables. Independent variables were entered one by one, given that levels of educational attainment for parents were substantially correlated ($r = .44$), a phenomenon known as assortative mating.

The degree of associations between *BAROCO Short* scores for parents and children and the categorized level of family income in the previous year was examined by Spearman's rank correlation coefficients. Furthermore, distributions of *BAROCO Short* scores for children as well as parents were

compared according to the categorized level of family income. Medians and upper and lower quartiles as well as minimum and maximum scores were graphed in a box plot for each family income group.

Tests of Hypotheses 3 and 4 were conducted using a behavioral genetic approach. Behavioral genetics is a statistical methodology designed to clarify genetic and environmental factors that influence individual differences in behavior through examination of the observed resemblance between twins, adoptees, or other family members (see Plomin et al., 2008). The twin method, a framework of behavioral genetic analysis most frequently used, was employed in this study. The logic of the twin method rests on the basis of comparisons of two different types of twins (i.e., identical or monozygotic: MZ twins, and fraternal or dizygotic: DZ twins). All MZ twin siblings share identical genes, whereas DZ twin siblings share approximately half of their genes with individual variations. Therefore, the relative ratio of genetic resemblance between MZ and DZ twin siblings is 2 to 1. However, the environmental conditions affecting the resemblance of twins can be considered almost equal between the two types of twins. Along with the comparison of resemblance between the two types of twins, resemblance between mothers–children and fathers–children can be combined. Parents and offspring share half of their genes as is the case with DZ twin siblings. Hence, any higher similarity for MZ twin siblings than DZ twin siblings or parents–offspring implies the presence of genetic effects in the trait. On the other hand, if there is no difference in the degree of similarity between both types of twins, shared – particularly those shared by siblings – environmental effects with no genetic effects are suggested. Moreover, if there is no difference in the degree of similarity between twin siblings and parents – offspring relations either, shared – particularly those shared between two generations in the family—environmental effects are suggested.

In test of Hypothesis 3, intrafamily correlations for MZ twin siblings, DZ twin siblings, mothers–children (twins) and fathers–children (twins) were compared by dividing the twins' ages into three categories: 13–14 years, 15–16 years, and 17–18 years. A parent's data were used twice to match the parent to each twin sibling's data.

In test of Hypothesis 4, multivariate genetic analysis with structural equation modeling was employed by fitting the model to the twins' covariance matrices of the three variables: scores for *BAROCO Short*, school achievement in English, and school achievement in Math. Multivariate genetic analysis can examine the degree to which genetic (A), shared environmental (C), and nonshared environmental (E) factors mediate the phenotypic covariance between traits. This technique also can describe the extent to which genetic effects on one trait overlap genetic effects on another trait, yielding a statistic referred to as genetic correlation (r_G; Fig. 2). The shared environmental correlation (r_C) and nonshared environmental correlation (r_E) can be likewise obtained (Neale & Maes, 2002; See Plomin & Spinath, 2002 for a description of multivariate genetic analysis among cognitive abilities).

The PASW Statistics 18 and Mx software package (Neale, 2004) was used for the analyses.

3. Results

3.1. Test of hypothesis 1

Regarding the *BAROCO Short* scores for twins, no significant mean differences were observed between female and male ($p = .28$); MZ and DZ ($p = .13$); and first-born and second-born ($p = .30$) twins. The score for mothers and fathers also did not show a significant mean difference ($p = .22$).

Fig. 3 depicts the mean of the total score of the *BAROCO Short* according to age. For children ages 13–18, the score increased gradually. For parents, after reaching the 30s, the score continued to increase gradually. It later decreased gradually after the late 50s. The education-adjusted scores also displayed a decline after the late 50s, but presented a plateau from the 30s to early 50s (Fig. 4). The mean difference between education-adjusted scores for those younger than age 55 and above age 55 was significant ($t = -2.21$, $df = 941$, $p < .05$). The level of the mean score showed systematical change according to age.

3.2. Test of hypothesis 2

Simple linear regression analyses revealed that the *BAROCO Short* scores for every family member were significantly explained by both mother's and father's educational attainment (Table 4). The effect of mother's education on

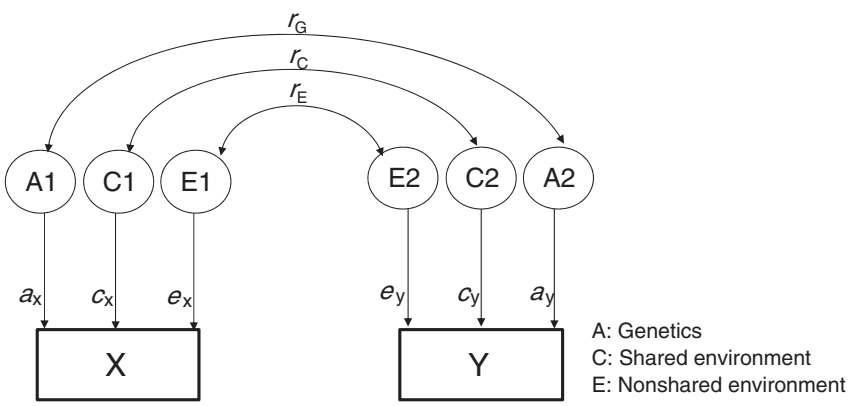
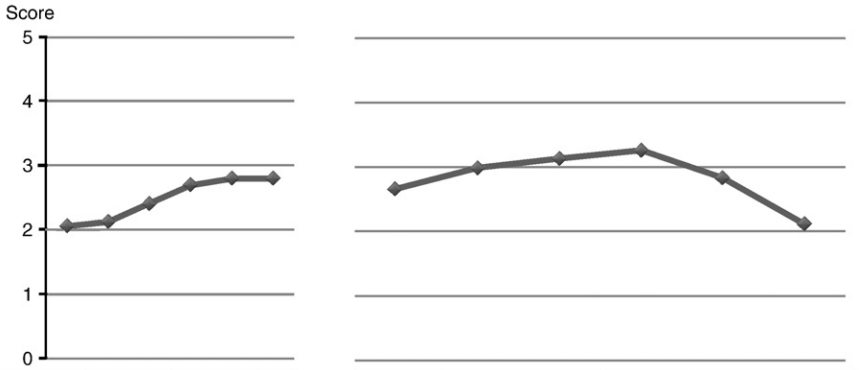


Fig. 2. Genetic (r_G), shared environmental (r_C), and nonshared environmental (r_E) correlations.



Age	13	14	15	16	17	18
N	180	176	203	180	149	131
Mean score	2.06	2.12	2.40	2.69	2.80	2.80
SD	1.64	1.61	1.61	1.55	1.38	1.47

34-39	40-44	45-49	50-54	55-59	60-68
45	249	374	198	64	17
2.67	2.98	3.13	3.26	2.83	2.12
1.52	1.48	1.46	1.48	1.52	1.54

Fig. 3. Mean score change according to age.

BAROCO Short scores was nearly equivalent among family members ($\beta = .20$ for children, $\beta = .19$ for mothers, and $\beta = .24$ for fathers). The effect of father's education on *BAROCO Short* scores was higher for fathers ($\beta = .29$) than for children ($\beta = .17$) and mothers ($\beta = .15$). Father's *BAROCO Short* scores were best explained by both the mother's and father's educational attainment.

Spearman's rank correlations exhibited significant associations between the *BAROCO Short* score and level of family income in the previous year: .33 for fathers, which was the highest, .25 for children, the second highest, .15 for mothers, the third highest.

In box plots, Fig. 5 graphically presents the distributions of *BAROCO Short* scores according to level of family income in the previous year. As the level of family income increased, the score became higher for both children and parents. Results consistent with the correlation analyses above showed that the tendency was particularly evident for fathers; it was not as evident for mothers. Scores for mothers with family income less than 2 million yen were relatively high. However, these mothers were suspected to be mostly single mothers with a score for father missing. Scores for mothers whose family income was over 2 million yen followed a similar increasing pattern as did scores for fathers and children.

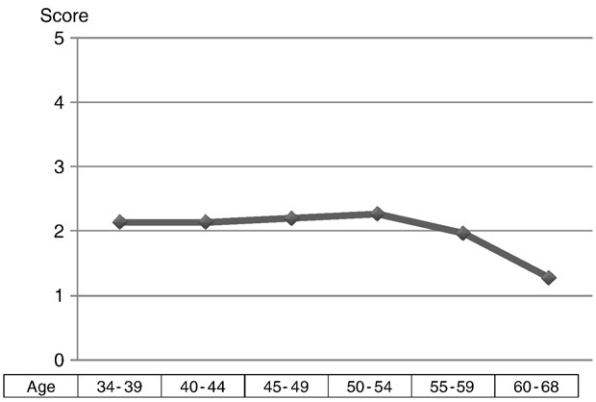


Fig. 4. Mean score change according to age with adjustment for educational attainment.

3.3. Test of hypothesis 3

Intrafamily correlations for *BAROCO Short* scores are presented in Fig. 6. Although DZ twins' correlations were stable across ages (.29 for ages 13–14, .25 for ages 15–16, and .25 for ages 17–18), the MZ twins' correlations increased as children's ages increased (.30 for ages 13–14, .37 for ages 15–16, and .44 for ages 17–18), which reflects the increasing influence of genetics from early adolescence to late adolescence. The respective correlations between children and fathers and between children and mothers also clearly indicated that at ages 13–14, the family resemblance was equivalent (.29 for father–child, and .27 for mother–child), irrespective of their genetic relatedness. However, older children showed less parent–offspring resemblance, which implies a substantial decrease in effects of the family environment, especially the environment shared by children and parents (.17 at ages 15–16 and .12 at ages 17–18 for father–child; .16 at ages 15–16 and .08 at ages 17–18 for mother–child).

3.4. Test of hypothesis 4

Intercorrelations, in terms of phenotypic, genetic, and environmental factor levels, among children's *BAROCO Short* scores, school achievement in English, and school achievement in Math are shown in Table 5. Phenotypic correlations (r_P) between the *BAROCO Short* and English, and between the *BAROCO Short* and Math were both .32. Genetic correlations (r_G) were very high: .77 between the *BAROCO Short* and English, and .70 between the *BAROCO Short* and Math. In contrast, environmental correlations were low: shared environmental (r_C) and nonshared environmental (r_E) correlations between the *BAROCO Short* and English, and between the *BAROCO Short* and Math were all in the range of .11–.18. Between English and Math, in comparison, the genetic correlation was high ($r_G = .76$), but environmental correlations were as high ($r_C = .98$: $r_E = .66$; Table 5).

Table 4
Regression analysis for *BAROCO Short* scores and parental educational attainment.

		BAROCO Short score								
		Children			Mothers			Fathers		
		β	t	p	β	t	p	β	t	p
Educational attainment	Mothers	.20	6.55	<.001	.19	6.47	<.001	.24	7.18	<.001
	Fathers	.17	5.00	<.001	.15	4.31	<.001	.29	8.69	<.001

4. Discussion

Tests of the hypotheses from the four perspectives described earlier all produced positive results. The four principles of IQ were true for the *BAROCO Short* scores.

4.1. Test of hypothesis 1: The mean score changes systematically according to age

Change in the mean level of the *BAROCO Short* scores computed by age group from adolescence to adulthood exhibited a plausible curve to that of *g* (Figs. 3 and 4). With a cross-sectional analysis using a small-size sample and with data of ages 19–33 missing, caution is necessary when interpreting the results. However, the curve gained in the present study suggests that the *BAROCO Short* measures a general mental ability across G_f and G_c. In the framework of Horn and Cattell's (1966, 1967) G_f–G_c theory, Kaufman (2001) reported that, based on analyzing education-adjusted WAIS III data on a standardization sample, the peak of Verbal IQ (G_c) was at ages 45–54 and notably declined in the 80s. The peak of Performance IQ (G_f) was at ages 20–24 and constantly and steadily declined afterwards throughout the adult life span. Hence, Kaufman's work supports the G_f–G_c theory and argues that Verbal IQ behaves as a maintained ability, and Performance IQ behaves as a vulnerable ability.

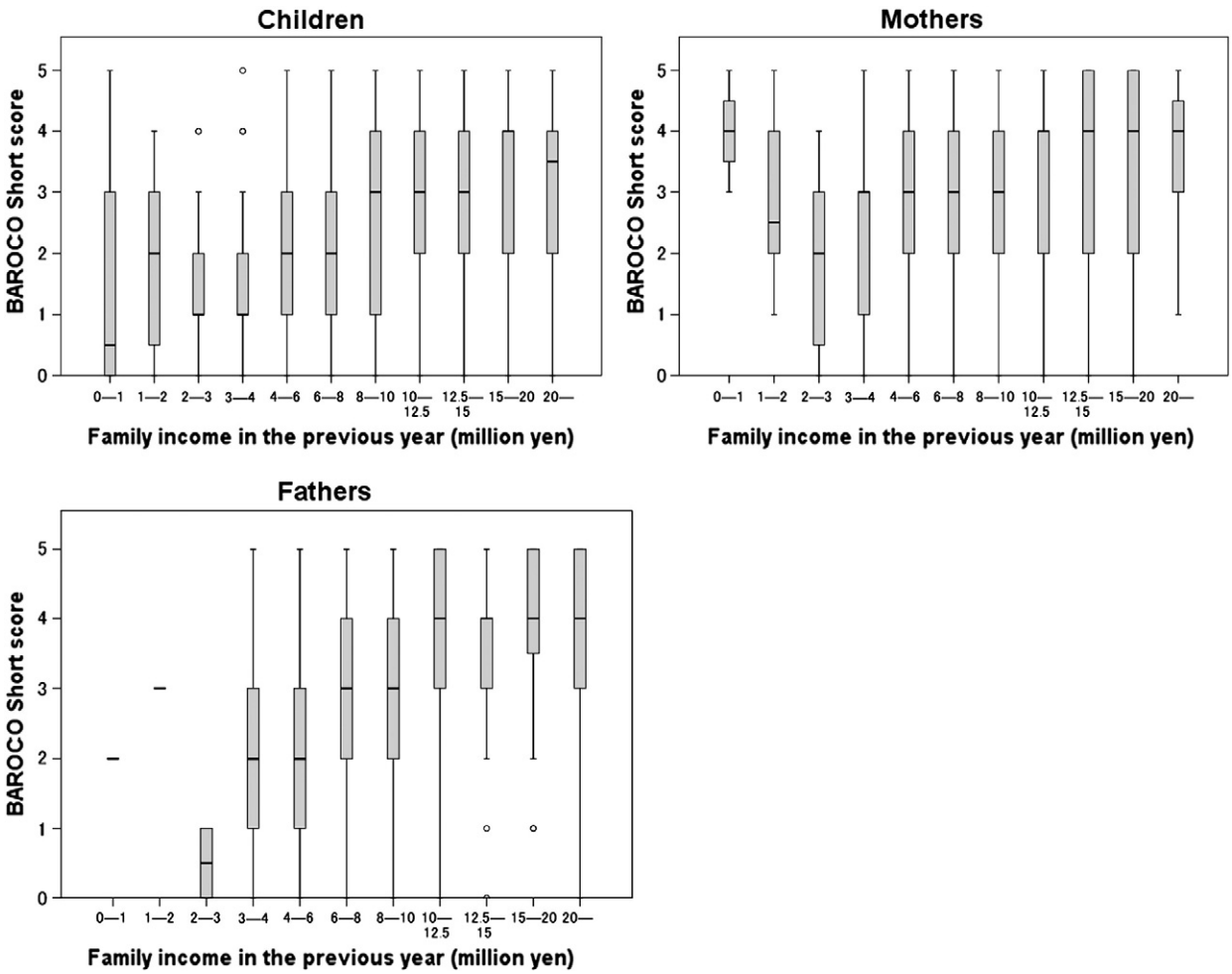


Fig. 5. Distributions of scores for children, mothers, and fathers according to family income in the previous year.

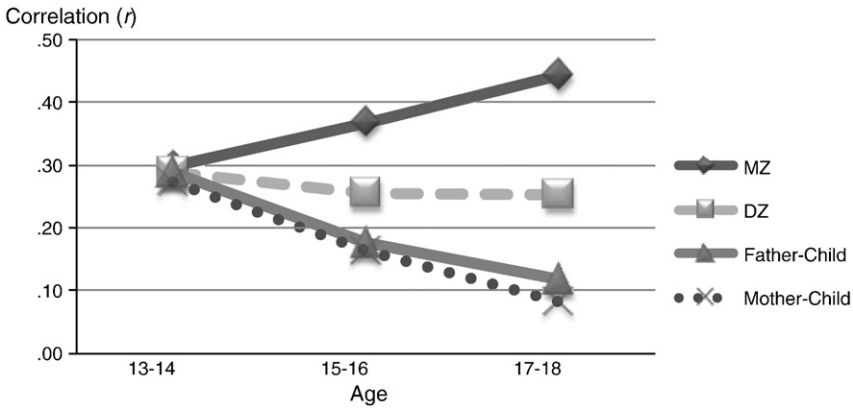


Fig. 6. Intrafamily correlations.

The curve obtained in the present study using education-adjusted (for adults) *BAROCO Short* scores did not represent either of Kaufman's (2001) two curves. From age 13 to 18, the level of performance elevated gradually. From the 30s to the early 50s, the level stayed unchanged and subsequently declined after the late 50s. The peak age is not known, because we lack data for ages 19–33, but within the corresponding age range, the shape of the curve followed the median points between Kaufman's two curves—education-adjusted Verbal IQ and Performance IQ of the WAIS III (Kaufman, 2001; Fig. 3). The same is true for the two curves by Horn and Hofer (1992); Figs. 3 and 4 where education-adjusted quadratic factor score curves for G_f and G_c from teens to 50s were predicted.

The findings offer evidence that what is measured by the *BAROCO Short* represents a higher order construct that integrates G_f and G_c, indicating general intelligence, *g*. However, a thorough comparison spanning the adult age range with ample sample size is needed to confirm this result.

4.2. Test of hypothesis 2: The score is predicted by socioeconomic status (SES)

Performance on the *BAROCO Short* for children, mothers, and fathers was significantly predicted by educational attainment of both mothers and fathers (Table 4) and was associated with family income (Fig. 5).

The SES variables, educational attainment for mothers and fathers as well as family income, showed the highest impact on father's *BAROCO Short* scores. The difference between fathers and mothers might reflect a gender difference based on intelligence and social status. The difference between parents and children might reflect a difference in causal direction between intelligence and SES. For parents, both intelligence and social outcomes can be a cause and an effect. For children, however, family SES cannot be an effect of intelligence.

The twin sample used in this study was too young to examine these hypothetical causalities. Investigation of the etiological overlap between intelligence and social outcomes would be possible by analyzing the data for the *BAROCO Short* and social achievement variables collected from adult twin pairs.

4.3. Test of hypothesis 3: The genetic effect increases, although the shared environmental effect decreases during progression from adolescence to adulthood

Late adolescence was the stage for a transitional period when factors affecting *BAROCO Short* scores dramatically changed. From ages 13 to 18, shared environmental effects drastically decreased and nearly disappeared, while genetic effects appeared and increased remarkably (Fig. 6). The evident change in underlying genetic and environmental structure during late adolescence corresponded exactly to that of IQ. However, the absence of genetic effects at ages 13–14, represented by equivalent intrafamily resemblance irrespective of genetic relatedness, was not in agreement with previous IQ research (e.g., McGue et al., 1993). This might have been caused by components specific to syllogism solving. A thorough examination is needed to clarify what produced the discrepant results.

Contributions of genetic and environmental effects were not computed in the present study. However, respective intraclass correlations for MZ and DZ of .44 and .25 at age 17–18 can allow us to simply estimate a genetic effect (heritability) of 38%, shared environmental effect of 6%, and nonshared environmental effect

Table 5
Phenotypic, genetic, shared environmental, and nonshared environmental correlations among *BAROCO Short*, English and Math.

	Phenotypic (r_P)		Genetic (r_G)		Shared environmental (r_C)		Nonshared environmental (r_E)	
	English	Math	English	Math	English	Math	English	Math
Math	.73		.76		.98		.66	
BAROCO Short	.32	.32	.77	.70	.11	.12	.11	.18

of 56%, using Falconer's formula (Falconer & MacKay, 1996). A heritability of 38% for a measure of *g* appears rather low as "genetic *g*" (Plomin & Spinath, 2002) which Plomin estimated to be around 50%. However, with Cronbach's α of around .75 for the *BAROCO Short*, approximately a fourth of the variance amounts to error components, which yields underestimation of heritability. Corrected for unreliability in measurement, the heritability estimates would be higher. On the other hand, error components should be minimal for *g*, since *g* is, by definition, factor components extracted from multiple measures. Given that, we can say that the heritability of 38% is plausible.

How heritability changes during the adult life span is not yet known. Analysis for *BAROCO Short* scores of adult twins covering a wide age range is needed in future research.

4.4. Test of hypothesis 4: The score is genetically correlated with other measures of cognitive ability

Children's school achievements in English and Math were equivalently phenotypically correlated with *BAROCO Short* scores, and both correlations were largely mediated by the same underlying genetic factor. At the phenotypic level, correlations between the *BAROCO Short* and school achievement were modest: .32 for both English and Math. However, at the genetic level, correlations were much higher: .77 for English and .70 for Math, and environmental correlations were negligible (Table 5). These values would be higher if low reliabilities of parental reports are taken into account. Phenotypically modest, genetically high, and environmentally low correlations were reported between primary school achievement and intelligence subtests as well (Thompson, Detterman, & Plomin, 1991). The commonality between school achievement and syllogistic reasoning ability is almost entirely genetic in origin, whereas specificities between them are almost entirely environmental. Such salient presence of genetic, not environmental, overlap between measures is a typical feature of *g* as a manifestation of the pleiotropy of genes in affecting diverse cognitive abilities (Plomin & Spinath, 2002).

Moreover, the fact that English and Math performance showed an equivalent structure of overlap with the *BAROCO Short* provides evidence that syllogistic reasoning ability is general, not specific, which further captures the characteristic of *g*.

4.5. Implications

With the four hypotheses supported, it was confirmed that the performance on the five syllogism-solving problems, included in a self-administered postal survey, yielded results very similar to those known as the nature of *g*. Consequently, we may reasonably suppose that the score obtained using a simple syllogism-solving test, the *BAROCO Short*, could have construct validity as an indicator of *g* and accordingly as a proxy for intelligence tests.

Most intelligence research to date has largely depended on standardized IQ tests. These tests have contributed to the enormous progress in this field of study. However, at the same time, because of the time-consuming, labor-intensive, and capital-intensive properties for both examiners and examinees that have inevitably accompanied IQ tests, access to information regarding administering an IQ test has been far from open to social scientists in general.

This simple syllogism-solving test is promising in this respect. The easy administration of the test, simply involving a questionnaire in a postal survey, will enable us to collect data from various samples, including those we have not accessed so far. Moreover, syllogisms are very easy to translate among languages, making intercultural direct comparisons possible. Future studies using this simple syllogism-solving test in various cultures and with diverse research backgrounds could be a big step in intelligence research.

4.6. Limitations

It should be noted that the association between the five syllogism-solving problems and IQ remains, in a sense, hypothetical, because our dataset did not include any direct measures of IQ. Consequently, the possibility cannot be completely ruled out that the results obtained in the present study might have been observed because components specific to syllogism solving happened to show correlation patterns similar to IQ. However, given the fact that four out of the four typical features of *g* were replicated with the five syllogism-solving problems, we believe that the possibility is unlikely. Still, checking criterion-related validity for the five problems of the *BAROCO Short* with IQ measures using the postal survey design is definitely needed in the future. In addition, more extensive work, including identifying characteristics for syllogism solving independent of *g*, is necessary to further clarify the association between syllogism solving and *g*.

It should also be pointed out that generalization of the results obtained in the present study requires caution because the sample was restricted to adolescent twins and their parents living in the Tokyo area with a higher than average SES. Replications using independent samples including non-twin families are necessary to justify the results.

4.7. Summary

The empirical findings obtained in the present study further strengthen the association between syllogism solving and *g* that were hitherto described merely as the degree of correlations or loadings from *g*. The substantive results yielded in performance on the five syllogism-solving problems – developmental changes in mean scores and the genetic and environmental structures, correlations with SES, and genetic correlations with school achievements – corresponded to typical features of *g*. This suggests that the overlapping components between the two constructs are substantial.

It is intriguing that this simple test requiring only a basic task measures complex human mental abilities. This might well be because the mechanism to solve syllogisms successfully reflects the essence of *g*.

Aristotle's insights over 2000 years ago were right in that logical reasoning ability expressed in syllogism solving is a symbol of human intelligence. An antiquarian tool required for human logical thinking, syllogism solving, is still alive as a vehicle of *g*, and can elucidate the mysterious nature of *g*.

Acknowledgements

This study was given the ISIR 2009 Travel Award at the 10th Annual Conference of the ISIR, and was revised to reflect the many useful comments from the participants. We greatly appreciate this chance to deepen our understanding of issues surrounding general intelligence. We wish to thank Yusuke Takahashi, Koken Ozaki, and Kunitake Suzuki for their help in carrying out the present study. We also would like to express our gratitude to Mitsuhiro Okada and Ryo Takemura for their collaborative work in developing the *BAROCO* tests. This research was supported in part by a Grant-in-Aid for Young Scientists from the Japan Society for the Promotion of Science (No. 19–5497: Shinji Yamagata).

References

Ando, J., Shikishima, C., Hiraishi, K., Sugimoto, Y., Takemura, R., & Okada, M. (2006). At the crossroads of logic, psychology, and behavioral genetics: Development of the deductive reasoning test in the Keio Twin-Baroco Project. In D. Andler, Y. Ogawa, M. Okada, & S. Watanabe (Eds.), *Reasoning and cognition: Interdisciplinary conference series on reasoning studies*, *vol. 1*. (pp. 9—36)Tokyo: Keio University Press.

Arend, I., Colom, R., Botella, J., Contreras, M. J., Rubio, V., & Santacreu, J. (2003). Quantifying cognitive complexity: Evidence from a reasoning task. *Personality and Individual Differences*, *35*, 659—669.

Bara, B. G., Bucciarelli, M., & Johnson-Laird, P. N. (1995). The development of syllogistic reasoning. *The American Journal of Psychology*, *108*, 157—193.

Beaumariage, G. N., Jr. (1960). Syllogistic reasoning. *Journal of Experimental Education*, *29*, 61—71.

Bergen, S. E., Gardner, C. O., & Kendler, K. S. (2007). Age-related changes in heritability of behavioral phenotypes over adolescence and young adulthood: A meta-analysis. *Twin Research and Human Genetics*, *10*, 423—443.

Bickersteth, P., & Das, J. P. (1981). Syllogistic reasoning among school children from Canada and Sierra Leone. *International Journal of Psychology*, *16*, 1—11.

Bochenski, I. (1970). *A history of formal logic.* New York: Chelsea.

Evans,, J. St. B. T., Newstead, S. E., & Byrne, R. M. J. (1993). Human reasoning: The psychology of deduction. Hillsdale, NJ: Lawrence Erlbaum Associates.

Falconer, D. S., & MacKay, T. F. C. (1996). Introduction to quantitative genetics (4th ed.). Harlow: Longmans Green.

Frandsen, A. N., & Holder, J. R. (1969). Spatial visualization in solving complex verbal problems. *The Journal of Psychology*, *73*, 229—233.

Guyote, M. J., & Sternberg, R. J. (1981). A transitive-chain theory of syllogistic reasoning. *Cognitive Psychology*, *13*, 461—525.

Heckman, J., & Krueger, A. (2003). *Inequality in America: What role for human capital policy?* Cambridge, MA: MIT Press.

Herrnstein, R. J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life.* New York: The Free Press.

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized intelligence. *Journal of Educational Psychology*, *57*, 253—270.

Horn, J. L., & Cattell, R. B. (1967). Age difference in fluid and crystallized intelligence. *Acta Psychologica*, *26*, 107—129.

Horn, J. L., & Hofer, S. M. (1992). Major abilities and development in the adult period. In R. J. Sternberg, & C. A. Berg (Eds.), *Intellectual development* (pp. 44—99). Cambridge: Cambridge University Press.

Jensen, A. R. (1998). *The g factor: The science of mental ability.* New York: Praeger.

Johnson-Laird, P. N. (1983). *Mental models.* Cambridge: CUP.

Kariya, T. (2001). Class in Japan and crisis in education. Tokyo: Yushindo-Kobunsha (in Japanese).

Kaufman, A. S. (2001). WAIS-III IQs, Horn's theory, and generational changes from young adulthood to old age. *Intelligence*, *29*, 131—167.

Kneale, W., & Kneale, M. (1962). *The development of logic*: Oxford University Press.

Kovas, Y., & Plomin, R. (2006). Generalist genes: Implications for the cognitive sciences. *Trends in Cognitive Sciences*, *10*(5), 198—203.

McCartney, K., Harris, M. J., & Bernieri, F. (1990). Growing up and growing apart: A developmental meta-analysis of twin studies. *Psychological Bulletin*, *107*, 226—237.

McGue, M., Bouchard, T. J., Iacono, W. G., & Lykken, D. T. (1993). Behavioral genetics of cognitive ability: A life-span perspective. In R. Plomin, & G. E. McClearn (Eds.), *Nature, nurture, and psychology* (pp. 59—76). Washington DC: American Psychological Association.

Neale, M. C. (2004). Mx software and documentation. http://www.vcu.edu/mx/.

Neale, M. C., & Maes, H. H. M. (2002). *Methodology for genetic studies of twins and families.* Dordrecht: Kluwer Academic Publishers.

Ooki, S., Yamada, K., Asaka, A., & Hayakawa, K. (1990). Zygosity diagnosis of twins by questionnaire. *Acta Geneticae Medicae et Gemellogiae*, *39*, 109—115.

Plomin, R., DeFries, J. C., McClearn, G. E., & McGuffin, P. (2008). Behavioral genetics (5th ed.). New York: Worth Publishers.

Plomin, R., & Spinath, F. M. (2002). Genetics and general cognitive ability (*g*). *Trends in Cognitive Sciences*, *6*(4), 169—176.

Shikishima, C., Ando, J., Ono, Y., Toda, T., & Yoshimura, K. (2006). Registry of adolescent and young adult twins in the Tokyo area. *Twin Research and Human Genetics*, *9*, 811—816.

Shikishima, C., Hiraishi, K., Yamagata, S., Sugimoto, Y., Takemura, R., Ozaki, K., et al. (2009). Is *g* an entity? A Japanese twin study using syllogisms and intelligence tests. *Intelligence*, *37*, 256—267.

Shikishima, C., Ozaki, K., Hiraishi, K., Sasaki, S., & Ando, J. (2009). Genetic and environmental factors contributing to educational attainment in Japanese society. *Behavior Genetics*, *39*, 680.

Spearman, C. (1904). "General intelligence", objectively determined and measured. *The American Journal of Psychology*, *15*, 201—293.

Stanovich, K. E., & West, R. F. (1998). Individual differences in rational thought. *Journal of Experimental Psychology*, *127*, 161—188.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *The Behavioral and Brain Sciences*, *23*, 645—726.

Tachibanaki, T. (2005). *Confronting income inequality in Japan: A comparative analysis of causes, consequences, and reform.* Cambridge: MIT Press.

Thompson, L. A., Detterman, D. K., & Plomin, R. (1991). Associations between cognitive abilities and scholastic achievement: Genetic overlap but environmental differences. *Psychological Science*, *2*, 158—165.

Torgersen, S. (1979). The determination of twin zygosity by means of a mailed questionnaire. *Acta Geneticae Medicae et Gemellogiae*, *28*, 225—236.

Wetherick, N. E. (1989). Psychology and syllogistic reasoning. *Philosophical Psychology*, *2*(1), 111—124.