Journal of Educational Psychology
2015, Vol. 107, No. 2, 486–501

0022-0663/15/$12.00 http://dx.doi.org/10.1037/a0037519

Gender Differences and School Influences With Respect to Three Indicators of General Intelligence: Evidence From Saudi Arabia

Sascha Hein and Mei Tan
Yale University

Abdullah Aljughaiman
King Faisal University

Elena L. Grigorenko
Yale University and Moscow State University of Psychology and Education

This study utilized multilevel modeling to examine the student characteristics (gender and age) and school factors (private vs. public school, urban vs. rural school, socioeconomic status, curricular organization, resources, activities, and equipment) associated with individual and between-school differences in the verbal, numerical, and figural intelligence test scores of 7,189 students in Grades 4–9 in the Kingdom of Saudi Arabia. Regarding student characteristics, results showed no substantial gender differences in the 3 intelligence subtests, whereas age accounted for 0.01%–0.06% of the individual differences in subtest performance. School factors accounted for 57.2%–71.6% of the differences in mean intelligence scores across schools. Moreover, 16.3% (figural) to 19.5% (verbal) of the variance was conditioned by different school experiences. Children in public schools achieved higher scores in all subtests, and a higher number of core, extra, and special courses consistently predicted higher scores. These findings extend the view that factors from the educational system are important correlates of intelligence, though they may vary across countries.

Keywords: intelligence, multilevel analysis, individual differences, school context, Kingdom of Saudi Arabia

Supplemental materials: http://dx.doi.org/10.1037/a0037519.supp

Understanding individual differences in intelligence[1] has been of great interest to researchers from the educational and psychological disciplines for more than 100 years (Neisser et al., 1996; Nisbett et al., 2012). Such differences have mainly been attributed to the complex interplay of child immanent predispositions and environmental influences of formal education and school contextual factors (Ceci & Williams, 1997; Christian, Bachnan, & Morrison, 2001; Eccles & Roeser, 2012; Mayer, 2000) as well as features of the home and family context (Bronfenbrenner & Morris, 2006; Fiese, 2001; Maccoby, 1992; Ramey & Ramey, 2012). Both the school and the family context are embedded in a larger cultural milieu that provides another major developmental context for a child's cognitive abilities to unfold (Harkness & Super, 2012; Nisbett, 2009; Rogoff, 2003).

Evidence of the impact of schooling on individual differences in intelligence has predominantly emerged from Western countries. However, these days the cultural milieu of children's development is strongly shaped by globalization, leading to increased fusion of and interdependence among cultures (Arnett, 2002). While the importance of formal education is being recognized among cultures worldwide, particularly Western-style modes of universal education leave their impact in countries around the globe (Grigorenko, 2008). Whereas many studies from Western countries have identified a number of school contextual factors associated with cognitive abilities (Ceci, 1991; Ceci & Williams, 1997; Christian et al., 2001; Eccles & Roeser, 2012), it is unlikely that the same factors will operate universally across different countries, cultures, and societies (Hanushek & Luque, 2003; Heyneman & Loxley, 1983). For instance, it has been noted that smaller classes (Hanushek, 1999) and smaller schools (Leithwood & Jantzi, 2009) do not necessarily yield better student outcomes, as class-size and school-size effects on student achievement may rather be nonlinear (Borland & Howsen, 2003; Borland, Howsen, & Trawick, 2005), indicating that there may be an "optimal" size.

This article was published Online First August 11, 2014.

Sascha Hein and Mei Tan, Child Study Center, Yale University; Abdullah Aljughaiman, Special Education Department, College of Education, King Faisal University; Elena L. Grigorenko, Child Study Center, Yale University, and Moscow State University of Psychology and Education.

This work was supported by a grant from the King Abdulaziz City for Science and Technology, Kingdom of Saudi Arabia. We would also like to acknowledge the assistance of Fatin Al-Alawi and Bushra Awidi, whose knowledge of the Arabic language and culture, and whose readiness to translate on the spur of the moment, were very much appreciated. We would also like to thank Mahmoud Al Bawaleez for his generous contributions to this study.

Correspondence concerning this article should be addressed to Elena L. Grigorenko, Child Study Center, Yale University, 230 South Frontage Road, New Haven, CT 06519-1124. E-mail: elena.grigorenko@yale.edu

[1] In this study, we use the term *intelligence* interchangeably with similar terms such as *IQ* and *general cognitive ability* (or *cognitive ability*).

Yet, important questions remain about whether the impacts of school factors on cognitive development generalize further, particularly among lower-middle-income countries, henceforth abbreviated as LMIC[2] (Fuller, 1987; Hanushek, 2006). There has been little systematic empirical research on the correlates of intelligence in LMIC (Engle et al., 2007; Grantham-McGregor et al., 2007; Walker et al., 2007); the available literature is even more sparse for Arab countries in the Middle East in general (Heyneman, 1997; Wagner & Spratt, 1987) and the Kingdom of Saudi Arabia in particular (e.g., Wiseman, Sadaawi, & Alromi, 2008).

There are three major reasons that school influences on cognitive skills in this understudied population need to be further delineated. First, the different organization of formal schooling around the globe and their influences on cognitive abilities might vary in other cultures and across countries (Demetriou & Papadopoulos, 2004). Second, these Western-style modes of formal schooling might vary in specific cultures (Elliott & Grigorenko, 2008; Sternberg, 2008). This implies a potentially different effect of the school context on intellectual functioning in different cultures. For example, as we discuss in more detail below, Saudi Arabia's educational system is based on gender-segregated formal schooling, a factor that might render the effects of the school context on cognitive development differently. For instance, women's choice of professions is restricted to certain occupations (mostly in the fields of medicine, science, and education), which results in women staying in school longer than men (Aljughaiman & Grigorenko, 2013). Third, the implications of the tragic terrorist attacks on the United States on September 11, 2001, have triggered changes and strains in the relationship between the two countries that have leveraged the debate and educational discourse about the Arab system of education in general and the role of Saudi textbooks in inciting violence against the West through its curriculum in particular (Center for Religious Freedom, 2008, 2011). This debate emphasizes that studying certain aspects of the educational environment is important to better understand the impact of school on children's and adolescents' development. Interestingly, despite this debate and media attention, the aftermath of September 11, 2001, did not produce as many educational studies as one would expect.

The relevant literature considers two groups of factors that account for individual differences in intelligence: student (child) based and school based.

Student-Based Characteristics and Cognitive Abilities

Gender and age are two of the most considered factors that may have a bearing on intelligence (e.g., Chen & Siegler, 2000; Halpern & LaMay, 2000). While there are no substantial gender differences in scores of general intelligence tests (Halpern & LaMay, 2000; Hyde, 2005; Lynn, Fergusson, & Horwood, 2005; Mackintosh, 1996; Neisser et al., 1996), some investigations have ascertained reliable differences in tests of certain cognitive abilities. Males, on average, score higher on subtests that require general knowledge, transformation in visual working memory such as mental rotation and spatial cognition (with a large range of effect sizes, ranging up to 0.9 standard deviations for spatial tasks), and fluid reasoning, especially in abstract mathematical and scientific domains such as arithmetic computations and arithmetical reasoning (small to moderate differences, ranging between 0.2 and 0.6 standard deviations); females, on average, score higher on tests that require use of language, verbal fluency (with differences up to 1 standard deviation), associative and phonological memory (with differences of about 0.2 to 0.5 standard deviation), and perceptual speed (Geary, Saults, Liu, & Hoard, 2000; Halpern, 1997; Halpern & LaMay, 2000; Johnson & Bouchard, 2007; Mackintosh, 1996). With regard to age, individual differences in intelligence test scores show high stability (Baltes, Staudinger, & Lindenberger, 1999; Deary et al., 2012; Ramsden et al., 2011), as expressed in substantial correlations (with smaller coefficients of about 0.3 from infancy to adult IQ, 0.5 from middle childhood to adolescence and larger correlation of up to 0.7 in adulthood) between intelligence test scores at different time points across an individual's life span (e.g., Deary, Whalley, Lemmon, Crawford, & Starr, 2000; Fagan, Holland, & Wheeler, 2007; Gow et al., 2011; Larsen, Hartmann, & Nyborg, 2008).

School-Based Factors and Cognitive Abilities

Formal schooling has important effects on cognitive development via a number of factors such as school structure and size, school resources, instructional practice, teacher–student relationships, academic organization, social composition, and curricular differentiation (N. Barber, 2005; Eccles & Roeser, 2012; Lee, 2000).

Early research in the United States has focused on the effects of material inputs such as textbook availability or overall school expenditure on academic achievement (Coleman et al., 1966). Later studies have classified the broader term of school resources as materials, human resources (student and teacher body), and instructional time (e.g., Barr & Dreeben, 1983). In research emerging from Western countries (e.g., the United States), school size has been found to be negatively linked to learning in secondary schools (Lee, 2001). Regarding class size, there is less consensus about the negative impact of large class sizes (or the beneficial impact of small class sizes) on achievement (with an average effect of reducing class sizes from 25 to 15 of about 0.10–0.20; Hattie, 2005). Also, the socioeconomic status of the student body at a school has been related to student achievement (Dronkers & Robert, 2008; McLoyd, 1998), with, for example, effect sizes of about 0.97 for mathematics achievement (Battistich, Solomon, Kim, Watson, & Shaps, 1995).

Differential performances have often been observed between public and private schools (Coleman, Kilgore, & Hoffer, 1982; Dronkers & Robert, 2008). The corresponding literature in the United States (e.g., Benveniste, Carnoy, & Rothstein, 2003; Lubienski, Crane, & Lubienski, 2008; Sander, 1999) is voluminous. Across the majority of these studies, findings indicate that private schools support higher student achievement (e.g., higher mathematics and literacy skills), even while accounting for a multitude

[2] We use the term *lower-middle-income countries* instead of the more commonly employed term *developing world* or *third world* in order to adequately capture the difficulties that emerge in educational research as a result of varying economies, and to avoid an ideological bias that puts higher income countries into a more privileged category. Although Saudi Arabia is considered a high-income non-Organisation for Economic Co-operation and Development country, the literature on LMIC provides the best possible background for the present study given the scarcity of research from Arab countries in the Middle East.

of student and family background factors (e.g., socioeconomic status, education levels of parents, child motivation, previous performance, malnutrition). The apparent effectiveness of private schools has been partially attributed to two factors: consistent and higher rates of engagement in academic activities (i.e., better attendance, engagement in homework, and more rigorous course offerings) and better disciplinary climate (i.e., more effective discipline and fairness leading to better peer behavior).

The Educational System in Saudi Arabia

In examining Saudi Arabian education, two characteristics stand out. First, the educational system of Saudi Arabia is shaped by strong centralization, with the Ministry of Education overseeing the construction of schools, the training of teachers, and the development of the national curriculum and its accompanying tests (Alarfaj, 2011). Second, the educational policies, determined by the national government and carried out by the ministry, derive largely from the religious and moral laws of Islam (Al-Sadan, 2000). The main focus of education in Saudi Arabia continues to be the Islamic Sharia (the totality of the religious and moral laws of Islam), which demands gender-segregated education and a curriculum that emphasizes the importance of Islamic education and Arabic language studies. While Islamic education and Arabic language studies play a major role in Saudi education still (UNESCO, 2011), the curriculum, set for all boys' and girls' schools, Grades 1–12, by the Ministry of Education, has diversified to include mathematics, science, and English, among other subjects (Rugh, 2002a). The curriculum in girls' schools has been described as a less comprehensive version of that taught in boys' schools (e.g., no sports education), since nonreligious education was initially regarded as unsuitable for girls after opening schools for girls (Mobaraki & Söderfeldt, 2010).

The Saudi way of teaching is still characterized primarily by dependency on rote learning (e.g., memorization of the Quran and other religious texts), low teaching standards, and minimal orientation toward the skill requirements of the global market (M. Barber, Mourshed, & Whelan, 2007). Under this system of education, children predominantly study the same academically oriented, rigid curriculum everywhere with textbooks printed centrally and distributed to schools, leaving only a limited range of flexibility for teachers to develop the curriculum (Al-Sadan, 2000). Thus, one might not expect to find much variation in children's cognitive development between Saudi schools (notwithstanding islands of special programs, such as those for gifted and talented children; Aljughaiman & Grigorenko, 2013).

Moreover, increasing privatization is a noteworthy trend in Saudi education in recent years (Rugh, 2002a, 2002b), and there are private schools in Saudi Arabia that may conduct themselves somewhat differently from the government schools. As of 2002, there were only very few privately owned educational K–12 institutions in the kingdom. Currently, however, private schools constitute about 12.9% of the total number of schools in Saudi Arabia. These were established by Saudi business families in Jeddah and the Eastern Province with the permission of the government, which requires them to teach the basic national curriculum but allows them to add to this curriculum if they choose to (Ministry of Education, Kingdom of Saudi Arabia, 2008; Rugh, 2002a). Many of them do this by adding English, math, and science courses taught in English (Mandelman, Tan, Aljughaiman, & Grigorenko, 2010). Other aspects that differentiate private from public schools in Saudi is the teacher salary, which is kept low by hiring primarily non-Saudi teachers, whose salaries are not subject to government regulation. In addition, private schools are funded largely through tuition fees paid by families, but do receive some funding (not more than 10%) from the government (Mandelman et al., 2010).

At first sight, the strongly centralized educational system and the rigid curriculum suggest only small differences might be found in students' intelligence test scores between schools. However, there is often a considerably greater between-school variation in LMIC, indicating larger differences and greater heterogeneity in average students' achievement between the schools for LMIC than in Western countries (Riddell, 1997; Scheerens, 2001). To delineate the factors that may explain these differences, the last 3 decades have been marked by research on school effects conducted in LMIC (e.g., Glewwe, Hanushek, Humpage, & Ravina, 2011; Scheerens, 2001). Most of the studies about LMIC have examined the impact of factors such as resource inputs (e.g., class size, teacher training and teacher salaries, availability of textbooks, general facilities and equipment; Fuller, 1987; Fuller & Clarke, 1994) and instructional processes (e.g., teacher's use of instructional time, the amount and type of curriculum covered; Fuller & Heyneman, 1989) on student achievement. Moreover, studies about the relative effectiveness of public versus private schools in LMIC often conclude that there is a private school advantage in academic achievement (Ammermueller, 2012; Dronkers & Avram, 2010; Jimenez, Lockheed, & Paqueo, 1991). However, the international literature has not yet reached consensus regarding the effect of school quality and school inputs on cognitive performance and academic achievement in LMIC (Hanushek, 1995, 2006; Kremer, 1995).

For example, some researchers highlight that school resources are important determinants of student achievement, but characteristics of the teachers and student–teacher ratios play relatively little role in explaining variation in school performance (Khoo & Khoo, 2005; Scheerens, 2001). Others still find certain school and teacher characteristics to have a positive effect on student achievement, for example, higher levels of school quality, class size, instructional time, teachers' qualifications, and availability of instructional materials such as textbooks (Fuller, 1987; Fuller & Clarke, 1994).

In contrast, some results from studies lead researchers to doubt the importance of these effects in LMIC (Baker, Goesling, & Letendre, 2002; Hanushek & Luque, 2003). These studies conclude that increasing the quality or quantity of most of the traditional inputs is not likely to improve student achievement, referring to inputs such as teacher training and expenditures per student (Simmons & Alexander, 1978), higher teacher salaries and reduced class size (e.g., Buchmann & Hannum, 2001; Cohn & Rossmiller, 1987), or physical and pedagogical characteristics such as physical plant, curriculum, instructional time, and teacher quality (Glewwe, Grosh, Jacoby, & Lockheed, 1995). Taken together, the clustering according to school structure and organization, physical resources, human resources, and instructional processes is still the predominant approach applied in recent research on school effects in LMIC (Glewwe et al., 1995; Lee & Zuze, 2011).

The Present Study

Characteristics of the school context play a vital role in the children's cognitive development. It has been argued (e.g., Cliffordson & Gustafsson, 2008) that in order to obtain a better understanding of the factors associated with intelligence, research should (a) adopt models of intelligence that allow for different dimensions of intellectual performance and (b) examine effects of different school characteristics rather than study effects of an undifferentiated amount of schooling. However, only little is known about the effects of specific school-contextual factors on different aspects of intelligence, especially in LMIC and Middle East in general and Saudi Arabia in particular. Therefore, in the present study, we illustrate gender differences and the associations between school factors and verbal, numerical, and figural intelligence test scores of a sample of fourth to ninth graders in Saudi Arabia.

It should be noted that this study is exploratory in nature. We applied translations or adaptations of Western instruments in this study, aware that results may be culturally distinct due to culture-specific notions of education and intelligence. We used these instruments partially because there are no Saudi-developed instruments currently available, but also because the aim of the larger study was to explore the nature of and contributing factors to Saudi Arabian students' cognitive performance and to differentiate students within the culture (not compare them to children of other cultures).

Thus, we assessed three indicators of intellectual functioning of students in Saudi Arabia with three aims: (a) to investigate gender differences in three indicators of general intelligence as well as the relation of these indicators with age, (b) to examine school differences in intelligence test scores, and (c) to elucidate which factors pertaining to school organization (i.e., school type, location, student–teacher ratio, socioeconomic status of the student body, curriculum, and academic activities) and school resources (i.e., classroom equipment and school facilities) are associated with mean intelligence test scores. Given the nationally unified and centralized educational system in Saudi Arabia, we did not expect to find substantial difference in children's intelligence test scores between the schools. However, based on the findings of research on school effects, we expected higher scores on the three intelligence subtests in private schools compared to public schools.

Method

Participants

A sample of 7,189 students (2,506 female, 4,683 male; age range: 7.3–18.7 years; $M_{\text{age}} = 12.28$ years, $SD_{\text{age}} = 1.81$) was recruited from 34 schools (18 primary and 16 secondary schools) in six regions (Riyadh, Jeddah, Tabuk, Alhassa, Abha, and Aljubail) across the Kingdom of Saudi Arabia. Out of these schools, 11 within four regions were located in villages or towns with a rather small population. Three out of these 11 schools can be characterized as located in a rural area. The remaining schools were located in cities with between 900,000 (where six of the private schools are located) and 4,000,000 inhabitants. Table 1 shows the distribution of the participating students across grades, private versus public schools, and rural versus urban schools. A set of chi-square tests revealed a relationship between gender and grade, $\chi^2(1) = 18.61$, $p < .01$, Cramér's $V = .05$, that is attributable to the higher percentage of boys in fifth and seventh grade; no difference in the number of boys and girls in private versus government schools, $\chi^2(1) = 2.06$, $p = .15$, Cramér's $V = .02$; and since only males were enrolled in schools in rural areas, there is a large relationship between gender and school location, $\chi^2(1) = 97.59$, $p < .01$, Cramér's $V = .12$.

Measures

General intelligence. The Aurora-*g* (*g* for *g* factor) test battery (Chart, Grigorenko, & Sternberg, 2008; Tan et al., 2009) was utilized as a paper-and-pencil measure to assess three indicators of general intelligence (i.e., figural, verbal, and numerical); it included three types of tasks (i.e., analogies, classification, series). The answers for all items had to be selected from four possible responses and each item was scored with 1 = correct or 0 = incorrect.

All of the items of Aurora-*g* (see supplemental materials) were reviewed in depth by the research team in Saudi Arabia for cultural and linguistic anomalies that Saudi Arabian students might not understand. As expected, no changes were required in the figural sections; few in the numerical sections (only those where the shape or configuration of the number may be important); but several adjustments were required in the verbal section. For example, items that were based on word sounds (such as rhyming, initial- or final-word sounds) had to be changed to Saudi words that preserved the intentions of the items. In addition, items that used words or referred to concepts unfamiliar to Saudi culture were

Table 1
Sample by Grade, Gender, School Type, and School Location

	Grade						School type		School location	
Gender	4	5	6	7	8	9	Private	Public	Urban	Rural
Male	790	824	765	930	720	654	539	4,144	4,505	178
Female	449	380	456	465	440	316	317	2,187	2,504	0
Total	1,239	1,204	1,221	1,395	1,160	970	856	6,331	7,009	178

Note. $N = 7{,}189$ (2,506 female, 4,683 male) clustered in 34 schools. School type and location information was missing for two female students.

replaced or deleted—such as sports that are not commonly played in the Kingdom of Saudi Arabia (e.g., hockey). Therefore, while we assume that all cultures have some set of cognitive mechanisms in common, we understand that interference from cultural differences should be avoided.

In fact, as expected for a newly translated and adapted version of an assessment being administered within a very different culture from the culture of origin, several items behaved contrary to expectation in Aurora-*g*. In general, low performing and mistranslated items were eliminated to bring the reliability of all to an acceptable level; the details of this process are shared below. This is in keeping with the notion that while types of assessment (e.g., subtests that present multiple series, analogy and classification subtests) may be culture fair, particular items may not be.

Aurora-*g* was originally developed for students in Grades 4–7, oversampling for difficult items with an intention to cover the "giftedness" spectrum of intellectual functioning. However, after initial screening of the data and preliminary data analyses of test performance, we found that Saudi students in Grade 9 did not score at the ceiling of any of the subtests. For the sake of comparison to other Aurora-*g* data sets that have pilot tested the battery across a broader grade range in different countries, we chose to report results of students up to Grade 9 in this study.

The *series tasks* of Aurora-*g* required students to figure out what comes next given a series of four words (verbal, 20 items), four numbers (numerical, 20 items), and four geometric figures (figural, 10 items). Due to item-total correlations below .20, three items were removed from the verbal subtest, four items from the numerical subtest, and two items from the figural subtest. Due to item mistranslation, an additional item was removed from the numerical subtest.

The *classification tasks* of Aurora-*g* required students to select an item that does not belong in a set of four words (verbal, 20 items), numbers (numerical, 20 items), or figures (figural, 20 items). Due to item-total correlations below .20, ten items were removed from the numerical subtest and two items from the figural subtest.

The *analogies tasks* of Aurora-g required students to select a missing piece in analogous pairs of words (verbal, 20 items), numbers (numerical, 20 items), and shapes (figural, 10 items). Due to item-total correlations below .20, one item was removed from the verbal subtest and five items were excluded from the numerical subtest. Due to translation problems, an additional item was removed from the verbal subtest.

The three types of tasks (series, analogies, and classification) were combined for each subtest in order to create scales as outcomes for the present study. The resulting verbal subtest comprised 55 items with an acceptable reliability of Cronbach's α = .75; the numerical subtest comprised 40 items with an acceptable reliability of Cronbach's α = .76; the figural subtest comprised 36 items with an acceptable reliability of Cronbach's α = .76. For comparison, a total score of all three subtests was computed that comprised 131 items and yielded a good reliability of Cronbach's α = .88.

School environment. Information on school-level variables was collected using a survey that was filled out by school officials of the participating schools. It was structured around two major sections: (a) school organizational characteristics (i.e., student and teacher body, curricular context, activities, and evaluation) and (b) school resources (i.e., school facilities and classroom equipment).

Organizational characteristics. Regarding general school characteristics, we asked whether the school was a private or public school and whether the school was located in a rural or urban area. School type (0 = private, 1 = public) and location (0 = rural, 1 = urban) were used as school-level predictors. The total number of students was divided by the total number of teachers at the school to compute the student–teacher ratio at each school. We also asked for the total number of students who qualified for reduced or no tuition in order to compute the percentage of these students for each school as an indicator of the average socioeconomic status of the students on the school level.

With regard to the curricular context, the school officials provided the number of core courses, number of special courses, number of extra courses, number of special programs for individualized learning as well as the number of courses for gifted students. Here, and in the following instances, a principal component analysis (oblique rotation) was conducted to examine the dimensionality and reduce the complexity of school-level information as well as the number of variables in the multilevel model. Results yielded a two-factor solution explaining about 75.7% of variance, with both factors being weakly related (r = .03). The number of core courses, special courses, and extra courses clustered together as one factor (henceforth called *core curriculum*; factor loadings were .87, .74, and .76, respectively; Cronbach's α = .67), as did the number of special programs for individualized learning and the number of courses for gifted students (henceforth called *specialized curriculum*; factor loadings were .94 and .93, respectively; Cronbach's α = .83). We used sum scores for the composite measure for both curriculum factors for further analyses.

Furthermore, we asked three questions regarding whether the school provided activities for students such as training (e.g., first aid, communication skills, and self-defense), visiting educators, and lecturers (henceforth called *activities*; Cronbach's α = .68). These questions had to be answered with no = 0 or yes = 1. A sum score of these three questions was created to assess the activities on the school level.

School resources. This section contained questions regarding the classroom equipment and the overall facilities at a school. All of the questions in this section had to be answered with no = 0 or yes = 1. With regard to equipment, we asked five questions about whether classrooms in general were equipped with their own computer, television, white board, and projector, as well as whether chairs and tables could be rearranged into different teaching configurations (henceforth called *classroom equipment*; Cronbach's α = .61).

Regarding the school's facilities, officials indicated whether the school had facilities, buildings, or classrooms specially designed for sports, workshops, folklore, gardening activities, concerts, specialized learning resource room, student clubs, library, and a science laboratory. A principal component analysis (oblique rotation) yielded a two-factor solution explaining about 56.8% of variance, with both factors being moderately negative related (r = −.30). The items loading highest of the first factor were facilities for sports, workshops, specialized learning room, library, and science laboratory (pattern matrix loadings were .77, .56, .58, .70, and .68, respectively). The items loading highest on the second factor were facilities for folklore, gardening activities, concerts, and student clubs (pattern matrix loadings were −.76, −.85, −.56, and −.58, respectively). Due to cross-loadings higher than .32 (i.e., about 10% of explained variance in an item explained by another factor) for facilities for workshops, concerts, and student clubs, we used factor scores for further analyses to account for the contribution of these items to both factors. The first factor was labeled as *core-curricular resources* (five items; Cronbach's α = .75); the second factor was labeled as *extracurricular resources* (four items; Cronbach's α = .74).

Procedure

The data were gathered in the year 2010 as part of a larger research project. The goal of this project was the characterization of the home environment, genetic predispositions, and adaptive behavior of a smaller sample of students with high levels of general intelligence. The Aurora-*g* battery was administered as a paper-and-pencil measure to groups of children to be completed in a single session. For the purpose of this study, Aurora-*g* was translated into Arabic and cross-validated (i.e., evaluated for familiarity, difficulty and suitability) through back-translation into English by a bilingual member of the project team. Trained members of the research team from the collaborating university in the kingdom collected the data on students' intelligence.

As indicated above, 34 schools participated in the collection of school-level information. Twenty-seven school officials (i.e., supervisors of the schools who work at the school district and are responsible for 10–15 schools) provided information about the school environment of the 34 schools. This means that two school officials evaluated two schools each, one person evaluated three schools, and one person evaluated four schools.

Data Analysis

At the student-level, independent sample *t* tests were computed along with Cohen's *d* effect size measure (Cohen, 1988) to compare average verbal, numerical, and figural test scores between male and female students. In addition, a 2 (between-subjects: male vs. female) × 9 (within-subjects: subtests) repeated-measures analysis of variance was conducted to analyze the profile of the nine subtest scores (i.e., analogies, classification, series for verbal, numerical, and figural, respectively) for boys and girls. Finally, variance ratios (i.e., dividing the male variance by the female variance) were computed to examine whether males compared to females showed greater variability in their intelligence subtest scores (Feingold, 1992). A variance ratio smaller than 1.0 indicated more variability among females than among males, whereas a variance ratio equal to 1.0 yielded equal variance in both genders.

Hierarchical linear modeling (Peugh, 2010; Raudenbush & Bryk, 2002) procedures were used to estimate the effects of school-level contextual variables on students' scores in the verbal, numerical, and figural subtests. In a first step (Model 1), we estimated an unconditional model with no predictor variables in order to provide estimates of the variance within and between schools before the inclusion of additional predictors (Model 1). The proportion of variance conditioned by the schools was estimated as the intraclass correlation coefficient (ICC). To consider multilevel methods, the ICC should indicate a substantial amount of the total variance (10%–20%) between the schools (Muthén, 1991, 1994; Peugh, 2010). In a second step (Model 2), we built a model with random intercepts (i.e., means as outcomes) that contained grand-mean-centered (i.e., subtracting the overall mean from a variable) age as a fixed effect on the three intelligence indicators across all schools. In a third step (Model 3), the proportion of variance in the three indicators (adjusted for age at the student level) that occurs between schools was modeled as a function of grand-mean-centered school-level variables. School gender was treated as a between-school variable because of the gender-segregated school system in Saudi Arabia. Moreover, as outlined above, schools for boys might differ from schools for girls in their curriculum (i.e., number of courses) and curriculum resources. Thus, we included four interaction terms in the model to examine differential effects of these factors in schools for boys versus schools for girls. Figure 1 shows the full contextual model (Model 3). The software Mplus (Version 6.11; Muthén & Muthén, 2010) was used to estimate parameters and treat missing data, utilizing a full-information robust maximum likelihood estimator (Satorra & Bentler, 1994).

Results

Preliminary Analyses

The total amount of missing values ranged between 1.32% (verbal–series subtest) and 4.01% (figural–analogies subtest). Due to this small fraction of missing data, we derived subtest scores by computing the mean when at least 70% of the items of a subtest had a valid value, multiplying it by the overall number of items in the scale, and rounding this value. This procedure resulted in an adjusted sum score. Under this procedure, only the numerical (n = 4) and the figural (n = 10) subtests showed missing values. There were no missing values in the school environment survey.

Descriptive Statistics

Table 2 shows the outcome means and standard deviations, the scale reliabilities, and skewness and kurtosis statistics for the three indicators of general intelligence and the total score. The overall low skewness and kurtosis values indicate that the subtest scores were approximately normally distributed, and the internal consistencies were acceptable for all subtests. The three intelligence subtests showed moderate intercorrelations (r = .41–.48, ps < .01) and were positively related to age (r = .23, for verbal; r = .24, for figural; r = .28, for numerical; r = .31, for the total score; all ps < .01).

Gender differences. We first examined gender differences in the three subtests and the total scores using independent-samples *t* tests with Bonferroni correction for four comparisons (i.e., 0.05/4 = 0.013). Although the effect sizes were rather small in magnitude, females, on average, scored higher in the verbal subtest, $t(6097.19) = -3.42$, p = .001, Cohen's d = 0.08; the figural subtest, $t(5970.59) = -6.87$, p < .001, Cohen's d = 0.17; and in the total score, $t(6700.97) = -3.15$, p = .001, Cohen's d = 0.07. In contrast, males, on average, scored higher in the numerical subtest, $t(6219.61) = 2.82$, p = .005, Cohen's d = 0.07. In order to determine the nature of these differences, we conducted a profile

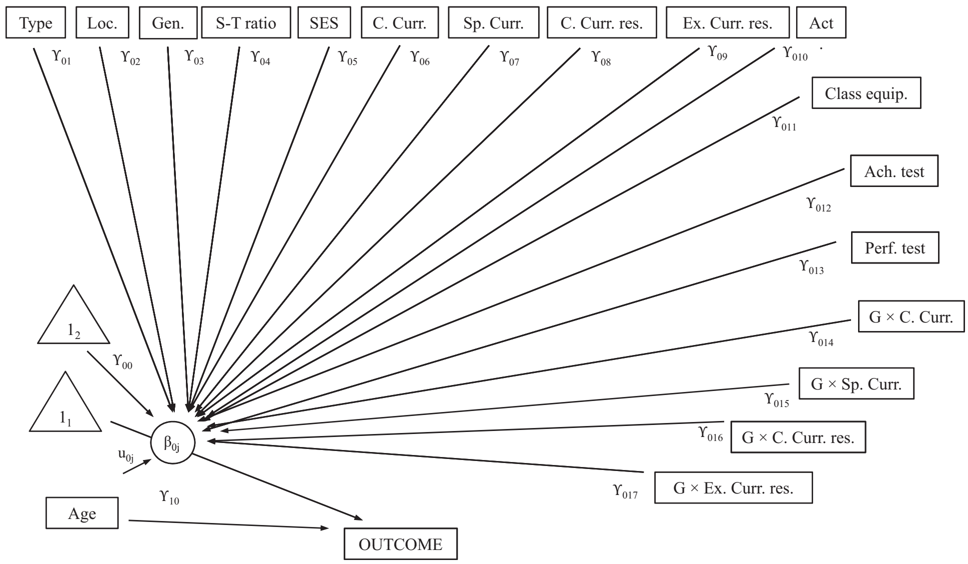

Figure 1. Path diagram of the full contextual model (Model 3). Models 1 and 2 were estimated prior to estimating this model but are not displayed here. The notation for Model 1 (i.e., the unconditional model) can be expressed as $Y_{ij} = \gamma_{00} + u_{0j} + r_{ij}$, with Y_{ij} representing intelligence test scores of student i in school j, γ_{00} reflecting the sum of a grand-mean intelligence test score, u_{0j} representing a school-specific deviation from the grand mean, and r_{ij} reflecting individual student differences around the mean of school j. Model 2 adds grand-mean-centered age to Model 1. The notation of Model 2 can be expressed as $Y_{ij} = \beta_{0j} + \beta_{1j}(\text{Age}_i) + r_{ij}$, with $\beta_{0j} = \gamma_{00} + u_{0j}$; $\beta_{1j} = \gamma_{10}$. The estimation of this model provides regression coefficients as fixed effects for age (γ_{10}) as a predictor of the three intelligence indicators. Model 3 include grand-mean-centered school-level variables (all except the interaction terms) as predictors to explain variation in the three intelligence subtest scores between schools. The notation for Model 3 can be expressed as $Y_{ij} = \beta_{0j} + \beta_{1j}(\text{Age}_i) + r_{ij}$, with $\beta_{1j} = \gamma_{10}$. In this model, β_{0j} contains school-level fixed effects that can be expressed as $\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{School Type})_j + \gamma_{02}(\text{School Location})_j + \gamma_{03}(\text{School Gender})_j + \gamma_{04}(\text{Student–Teacher Ratio})_j + \gamma_{05}(\text{Socioeconomic Status})_j + \gamma_{06}(\text{Core Curriculum})_j + \gamma_{07}(\text{Specialized Curriculum})_j + \gamma_{08}(\text{Core-Curricular Resources})_j + \gamma_{09}(\text{Extracurricular resources})_j + \gamma_{010}(\text{Activities})_j + \gamma_{011}(\text{Classroom Equipment})_j + \gamma_{012}(\text{Achievement Test})_j + \gamma_{013}(\text{Performance Test})_j + \gamma_{014}(\text{Gender} \times \text{Core Curriculum})_j + \gamma_{015}(\text{Gender} \times \text{Specialized Curriculum})_j + \gamma_{016}(\text{Gender} \times \text{Core-Curricular Resources})_j + \gamma_{017}(\text{Gender} \times \text{Extracurricular Resources})_j + u_{0j}$. The indices $\gamma_{01}, \gamma_{02}, \ldots, \gamma_{017}$ denote regression coefficients of predictors at the school level. Loc. = location (urban or rural); Gen. = school gender (boys' school or girls' school); S-T ratio = student–teacher ratio; SES = socioeconomic status; C. Curr. = core curriculum; Sp. Curr. = specialized curriculum; C. Curr. res. = core-curricular resources; Ex. Curr. res. = extracurricular resources; Act. = activities; Class equip. = classroom equipment; Ach. test = achievement test; Perf. test = performance test; G × C. Curr. = school gender by core curriculum interaction; G × Sp. Curr. = school gender by specialized curriculum interaction; G × C. Curr. res. = school gender by core curricular resources interaction; G × Ex. Curr. res. = school gender by extracurricular resources interaction.

analyses for the three types of tasks (analogies, classification, series) for the verbal, numerical, and figural subtests, respectively. A repeated measures analysis of variance with one within-subjects factor (i.e., nine z-standardized subtest scores) and one between-subjects factor (i.e., gender) was conducted to examine the scores on the nine subtest scores of boys and girls. Mauchly's test of sphericity showed a violation of the sphericity assumption (Mauchly's $W = .85$, Approx. $\chi^2(35) = 1092.11$, $p < .001$, Greenhouse–Geisser $\varepsilon = .961$). Thus, Greenhouse–Geisser corrected test statistics with adjusted degrees of freedom were used for the tests of significance. A significant between-group main effect for subtests suggested unequal levels of the profiles, that is, a difference in the grand mean of all subtests between boys and girls, $F(1, 6733) = 4.48$, $p = .034$, $\eta = .001$. A nonsignificant within-subjects main effect for subtests, $F(7.69, 51757.56) = 1.72$, $p = .092$, $\eta = .000$, indicated that the slope of each line segment was 0 and the profile flat. Most importantly, a significant interaction between gender and subtests, $F(7.69, 51757.56) = 25.08$, $p < .001$, $\eta = .004$, indicated that the profiles were not parallel and therefore different for boys and girls. Post hoc analyses utilizing Bonferroni correction indicated that boys and girls differed on six subtest scores but not on the remaining three (i.e., verbal–analogies, verbal–series, figural–series).

In addition to this mixed picture of mean differences between males and females, variance ratios greater than 1 indicate that male students varied more in their performance on all subtests than female students (see Table 2). Thus, males were overrepresented at both extreme ends of the distributions of the verbal, numerical, and figural intelligence test scores. In general, males were overrepresented in our sample at a rate of 1.87:1. After correcting for this overrepresentation, males were mostly overrepresented in the top 10th percentile of numerical scores (1.67:1), followed by verbal

Table 2
General Statistics for the Subtests of Aurora-g

Variable	Total		Male		Female									
	M	*SD*	*M*	*SD*	*M*	*SD*	*t*	*df*	*p*	*d*	VR	Skew	Kurt	α
Verbal	25.79	6.86	25.60	7.31	26.15	5.92	−3.42	6097.19	.001	0.08	1.53	−0.06	0.06	.75 (55)
Numerical	11.97	5.52	12.10	5.92	11.74	4.66	2.82	6219.61	.005	0.07	1.62	0.77	1.07	.76 (40)
Figural	14.11	5.56	13.79	5.87	14.69	4.88	−6.87	5970.59	.000	0.17	1.45	0.23	−0.05	.76 (36)
Total	51.97	14.14	51.63	15.59	52.62	10.89	−3.20	6691.72	.001	0.07	2.04	0.24	0.69	.88 (131)

Note. $N = 7{,}189$. t = t value for independent samples t test comparing scores of males and females; *df* = degrees of freedom of the t statistic; d = Cohen's d effect size; VR = variance ratio (i.e., variance of males divided by variance of females); skew = skewness; kurt = kurtosis; α = Cronbach's alpha with numbers of items for the pertinent scale given in parentheses.

scores (1.32:1), and only slightly in figural scores (1.12:1). In contrast, for the bottom 10th percentile, males were mostly overrepresented in the figural scores (2.19:1), followed by verbal scores (1.95:1) and numerical scores (1.44:1).

School-level variables. Table 3 contains the descriptive statistics for school-level variables. We conducted independent sample t tests (using Bonferroni-corrected alpha of $0.05/8 = 0.063$) to examine differences between private and public schools and between schools for boys and schools for girls with regard to the continuous variables on the school level. The results show significant differences between private and public schools in favor of private schools with regard to the core curriculum composite, $t(32) = 14.02$, $p < .001$; the amount of activities, $t(28.50) = 4.46$, $p < .001$; and the amount of classroom equipment, $t(31.66) = 4.75$, $p < .001$. In contrast, public schools showed a marginally significant higher level of extracurricular resources, $t(32) = -2.91$, $p = .007$. Chi-square tests showed no difference between private and public schools regarding the use of achievement tests, $\chi^2(1) = 0.19$, $p = .66$, Cramér's $V = .08$, and the use of performance tests, $\chi^2(1) = 1.13$, $p = .29$, Cramér's $V = .18$. For school gender (boys' schools vs. girls' schools), the only difference found was for core-curricular resources, with schools for males having higher levels on this variable ($M = 0.53$, $SD = 0.51$) compared to schools for females ($M = -0.76$, $SD = 1.05$), $t(17.37) = 4.24$, $p = .001$.

Regarding correlations among school-level variables (see Table 3), we found negative correlations between student–teacher ratio and the core curriculum composite ($r = -.73$, $p < .01$), the core-curricular resources ($r = -.54$, $p < .01$), and the activities ($r = -.35$, $p < .05$) and classroom equipment ($r = -.44$, $p < .01$). School-level socioeconomic status and the specialized curriculum composite were not related to any of the other school variables. Core curriculum composite, core-curricular resources, and activities and classroom equipment yielded moderate positive intercorrelations ($r = .41$–$.46$). Interestingly, the extracurricular resource factor score was negatively related to the other school-level variables (and significantly negatively related with activities and classroom equipment), but positively (though not significantly) with the student–teacher ratio.

Effects of School-Level Variables on Intelligence Test Scores

Before examining the effects of school-level variables on intelligence test scores, an important question has to be answered: Is multilevel modeling needed? To answer this question, unconditional models were estimated for each subtest score of Aurora-*g* as well as the total score (Model 1 in Table 4). Although most of the variance can be attributed to the student level, the ICCs indicate that a substantial proportion (ranging between 16.3% for the figural subtest and 24.3% for the total score) of test score variation occurs across schools. Accordingly, the relatively high design effects (between 5.94 for the figural subtest and 7.22 for the total score) indicate a nested structure of the data and support the need

Table 3
Descriptive Statistics of Continuous Variables on the School Level

Variable	*M*	*SD*	α	Private		Public										
				M	*SD*	*M*	*SD*	*t*(32)	*p*	2	3	4	5	6	7	8
1. Student–teacher ratio	13.60	5.10		7.70	2.39	15.42	4.26	−4.86	.000	−.06	−.73	−.12	−.54	.29	−.35	−.44
2. Socioeconomic status	2.67	7.70		2.47	3.09	2.74	8.70	−0.09	.932	—	.01	.07	.11	−.01	.17	−.08
3. Core curriculum	9.38	1.74	.67 (3)	12.25	1.04	8.50	0.51	14.02	.000		—	.05	.41	−.30	.46	.45
4. Specialized curriculum	1.00	1.78	.83 (2)	1.25	0.71	0.92	2.00	0.45	.656			—	.24	−.12	−.11	−.05
5. Core curriculum resources	0.00	1.00	.75 (5)	0.65	0.51	−0.20	1.04	2.21	.035				—	−.30	.47	.46
6. Extracurricular resources	0.00	1.00	.74 (4)	−0.81	0.96	0.25	0.89	−2.91	.007					—	−.40	−.42
7. Activities	1.82	1.11	.68 (3)	2.75	0.46	1.54	1.10	4.46	.000						—	.36
8. Class equipment	3.09	1.40	.61 (5)	4.25	0.46	2.73	1.41	4.75	.000							—

Note. $N = 34$ schools (eight private and 26 public schools; 20 schools for males, 14 schools for females). Three of the public schools were located in rural areas; the remaining 31 schools were located in urban areas. Twenty-eight schools (seven private and 21 public schools) utilized achievement tests; 20 schools utilized performance tests (six private and 14 public schools). α = Cronbach's alpha with numbers of items for the pertinent scale given in parentheses; t = t value for independent samples t test comparing scores of private versus public schools.

Table 4
Variance Estimates for Unconditional Models and Models Containing Within-Level Predictors

Parameter estimate	Verbal	Numerical	Figural	Total
		Model 1		
Student-level σ^2	38.99	25.63	26.90	160.28
School-level τ_{00}	9.45	5.82	5.24	52.69
Grand mean ($\hat{\gamma}_{00}$)	25.94	12.06	14.42	52.56
ICC	.195	.185	.163	.243
M_c	211.38	211.27	211.09	211.24
Design effect	6.48	6.32	5.94	7.22
Deviance	46847.43	43805.74	44111.36	56977.56
		Model 2		
Age (γ_{10})	0.087	0.212*	0.222*	0.478
R^2	.001	.006	.006	.005
Deviance	46845.70	43791.01	44095.67	56965.46

Note. Model 1 represents an unconditional model with variance estimates for the student level (within level) as well as the school level (between level). Model 2 contains age as a within-level predictor. Estimates for within-level predictors reflect unstandardized estimates. σ^2 = variance on the student level; τ_{00} = variance across schools; ICC = intraclass correlation coefficient; M_c = average cluster size; design effect = $\sqrt{1 + ICC \cdot (M_c - 1)}$; R^2 = variance explained at the within level; deviance = −2 log-likelihood.
* $p < .05$.

for multilevel modeling to separately estimate the test score variance that occurs both across students and across schools. The grand mean (γ_{00}) of intelligence test scores across the 34 schools was 25.94 for verbal, 12.06 for numerical, 14.42 for figural, and 52.56 for total score (note the similarity to the means for the students in the sample; see Table 2). On the basis of the mean and the school-level variance estimates from the unconditional model, we calculated a 95% confidence interval for these means. Results indicate a substantial range in average intelligence test scores among the 34 schools, with a range of values of 19.91–31.97 for verbal scores, 7.33–16.79 for numerical scores, 9.93–18.91 for figural scores, and 38.03–66.49 for total scores. These intervals indicate that average intelligence test scores are much higher in the highest performing compared to the lower performing schools (i.e., 62.28% higher for verbal, 43.67% higher for numerical, 52.51% higher for figural, and 57.20% higher for the total score).

Student-level predictions. In a second step (Model 2 in Table 4), age was added as a predictor to explain the proportion of variations in test scores on the student level. Results showed that age was positively related to the numerical ($B = 0.21$, $SE = 0.09$, $p = .012$, $\beta = 0.08$), the figural ($B = 0.23$, $SE = 0.11$, $p = .037$, $\beta = 0.08$), and the total test score ($B = 0.48$, $SE = 0.25$, $p = .05$, $\beta = 0.07$). Thus, for each increase of 1 standard deviation in age, numerical test scores increase by 0.08 standard deviation, figural test scores increase by 0.08 standard deviation, and total test scores increase by 0.07 standard deviation. The variance explained by age on the student level was rather small, with 0.1% for the verbal test scores, 0.6% for the numerical and figural test scores, and 0.5% for the total score.

School-level predictions. In a third step (see Table 5), school-level variables were added to the model to explain between-school variability in test scores. Overall, the complete set of contextual indicators accounted for 57.2% (figural subtest) to 71.6% (verbal subtest) of variance between schools. Contrary to expectations, results indicated a performance difference between private and public schools with higher intelligence test scores in the verbal and numerical subtests as well as the total score for students in public schools.[3] This advantage of public over private schools is quantified below along with other school-level variables that were significantly related to the three intelligence subtest scores and the total score.

Verbal subtest scores were significantly negatively related to student–teacher ratio ($B = -0.32$, $SE = 0.13$, $p = .015$, $\beta = -0.54$) and the amount of activities ($B = -1.25$, $SE = 0.35$, $p < .001$, $\beta = -0.46$) and positively related to the core curriculum composite ($B = 2.96$, $SE = 0.83$, $p < .001$, $\beta = 1.70$) and the use of performance tests ($B = 1.62$, $SE = 0.77$, $p = .036$, $\beta = 0.27$). For numerical subtest scores, the only significant relationship with school-level variables was found for the core curriculum composite ($B = 2.49$, $SE = 0.75$, $p = .001$, $\beta = 1.95$). However, the interaction between school gender and core curriculum was significant ($B = -1.47$, $SE = 0.59$, $p = .012$), as was the interaction between school gender and extracurricular resources ($B = -2.34$, $SE = 1.03$, $p = .023$). These interactions indicate that the main effect of core curriculum cannot be interpreted directly, since numerical test scores increase more steeply for boys' schools compared to girls' schools with increasing levels of core curriculum. Figural subtest scores were positively related to school gender ($B = 2.26$, $SE = 1.11$, $p = .042$, $\beta = 0.53$), indicating higher test scores in schools for females. Finally, the total test scores were significantly positively related to the core curriculum composite ($B = 6.69$, $SE = 2.05$, $p = .001$, $\beta = 1.72$). School location, socioeconomic status of the student body, core-curricular resources, extracurricular resources, specialized curriculum, classroom equipment, and the use of achievement tests did not explain a significant proportion of between-school variation in intelligence test scores.

Taken together, the effect of school factors on students' intelligence test scores mattered more than expected. Figure 2 shows the standardized regression weights (β) and associated standard errors for all subtests and the total score. In particular, students' mean verbal test scores were 9.57 points ($\beta = 1.36$) higher in public compared to private schools and 1.62 points ($\beta = 0.27$) higher in schools that use a performance test. On the one hand, a 1 standard deviation increase in student–teacher ratio and the amount of activities implied a decrease of verbal test scores of about 0.54 and 0.46 standard deviation, respectively. On the other

[3] Note that in this set of multilevel analyses, only the between-school variation is predicted by the set of school-level variables. Therefore, the results presented here refer to students' mean intelligence test scores when school type (private vs. public) is part of a set of variables that predict the intelligence test scores. In contrast, independent-samples t tests suggest that students in private schools outperform their public school counterparts in verbal subtest scores, $t(1188.07) = 9.31$, $p < .01$, $M_{\text{private}} = 27.62$, $M_{\text{public}} = 25.55$; numerical subtest scores, $t(1213.14) = 7.82$, $p < .01$, $M_{\text{private}} = 13.17$, $M_{\text{public}} = 11.81$; and figural subtest scores, $t(1165.33) = 5.25$, $p < .01$, $M_{\text{private}} = 14.96$, $M_{\text{public}} = 13.99$; as well as the total score, $t(1433.39) = 11.24$, $p < .01$, $M_{\text{private}} = 55.77$, $M_{\text{public}} = 51.48$. However, when the assumption of independent observations is violated, one cannot draw reliable conclusions from results derived from univariate tests such as the independent-samples t test.

Table 5
Results From Final Contextual Model Analysis for Each Domain of General Intelligence

Parameter estimate	Verbal		Numerical		Figural		Total	
	Coeff.	*SE*	Coeff.	*SE*	Coeff.	*SE*	Coeff.	*SE*
Fixed effects								
Intercept (γ_{00})	26.110***	0.51	12.275***	0.54	14.299***	0.50	52.685***	1.34
Student level								
Age (γ_{10})	0.093	0.12	0.205*	0.09	0.220*	0.11	0.475	0.25
School level								
Type (γ_{01})	9.568*	3.88	12.138**	3.50	5.856	3.03	27.900**	9.63
Location (γ_{02})	1.614	2.11	0.409	2.20	−0.102	1.91	1.641	5.91
Gender (γ_{03})	0.550	1.47	2.008	1.13	2.255*	1.11	4.527	3.33
Student–teacher ratio (γ_{04})	−0.321*	0.13	−0.149	0.16	−0.214	0.14	−0.676	0.38
Socioeconomic status[a] (γ_{05})	−0.03	0.03	−0.008	0.03	−0.013	0.03	−0.056	0.08
Core curriculum (γ_{06})	2.961***	0.83	2.487**	0.75	1.122	0.69	6.694**	2.049
Specialized curriculum (γ_{07})	−0.456	0.269	−0.035	0.20	−0.043	0.22	−0.591	0.614
Core-curricular resources (γ_{08})	−0.526	1.00	0.057	0.73	0.265	0.83	−0.311	2.60
Extracurricular resources (γ_{09})	−0.862	0.52	−0.861	0.49	−0.475	0.41	−2.256	1.34
Activities (γ_{010})	−1.253***	0.35	0.132	0.32	−0.250	0.34	−1.373	0.88
Classroom equipment (γ_{011})	−0.753*	0.35	−0.038	0.35	−0.157	0.31	−0.981	0.99
Achievement test (γ_{012})	−1.017	0.95	1.002	0.89	−0.761	0.87	−0.593	2.37
Performance test (γ_{013})	1.619*	0.77	0.626	0.61	0.325	0.67	2.872	1.82
Gender × Core Curriculum (γ_{014})	−1.135	0.79	−1.467*	0.59	−0.360	0.61	−2.862	1.79
Gender × Specialized Curriculum (γ_{015})	−0.508	0.72	−0.428	0.46	0.552	0.61	−0.459	1.56
Gender × Core-Curricular Resources (γ_{016})	2.475	1.77	0.406	1.31	−0.175	1.47	2.532	4.51
Gender × Extracurricular Resources (γ_{017})	0.340	1.09	−2.344*	1.03	−1.094	0.79	−2.964	2.65
Variance components								
R^2								
Within level	.001	.002	.005	.005	.006	.006	.005	.005
Between level	.716***	.08	.580***	.10	.572***	.10	.616***	.10
σ^2								
Within level	38.99***	3.16	25.60***	2.66	26.87***	2.22	160.13***	23.27
Between level	2.53**	0.76	2.01***	0.55	1.86***	0.48	17.17**	4.94
Model summary								
ICC	.186		.156		.138		.217	
Deviance	46805.43		43763.50		44069.54		56934.42	

Note. Positive values for gender at the school level indicate higher scores for girls' schools. Positive values for school type indicate higher scores for public schools. Positive values for school location indicate higher values for urban schools. All coefficients were centered at the grand mean to measure the impact of a 1 standard deviation change in the variable on the respective score. The estimates for predictor variables reflect unstandardized regression coefficient (fixed effects). Because of grand-mean centering of all predictors, the intercept is interpreted as the school mean of the expected intelligence test score for students with grand mean levels of all school variables. Coeff. = coefficient; R^2 = explained variance; σ^2 = residual variance; ICC = intraclass correlation coefficient; deviance = −2 log-likelihood.
[a] Percentage of students who qualify for reduced or no tuition.
* $p < .05$. ** $p < .01$. *** $p < .001$.

hand, a 1 standard deviation increase in the core curriculum composite is associated with a 1.70 standard deviation increase in verbal test scores. Furthermore, mean numerical subtest scores were 12.14 points higher ($\beta = 2.36$) in public compared to private schools, and a 1 standard deviation increase in the core curriculum composite was associated with a 1.95 standard deviation increase in numerical subtest scores. However, as the interaction effect indicates, schools for males and females had the same levels of numerical test scores for a value of 1.47 on the core curriculum composite. For the figural subtest, mean scores for schools for females were 2.26 points ($\beta = 0.53$) higher compared to schools for males. Finally, mean total test scores were 27.9 points ($\beta = 1.77$) higher in public compared to private schools. Moreover, a 1 standard deviation increase in the core curriculum composite yielded a 1.72 standard deviation increase in total test scores.

Discussion

We presented a study that sought to explore gender differences in verbal, numerical, and figural intelligence test scores and school-level correlates of intelligence. Examining intelligence test scores of a sample of 7,189 fourth to ninth graders, we found (a) negligible gender differences in general intelligence (with the largest difference in the figural subtest favoring females); (b) between 57.2% and 71.6% of variance in intelligence test scores conditioned by the schools; (c) higher mean intelligence test scores

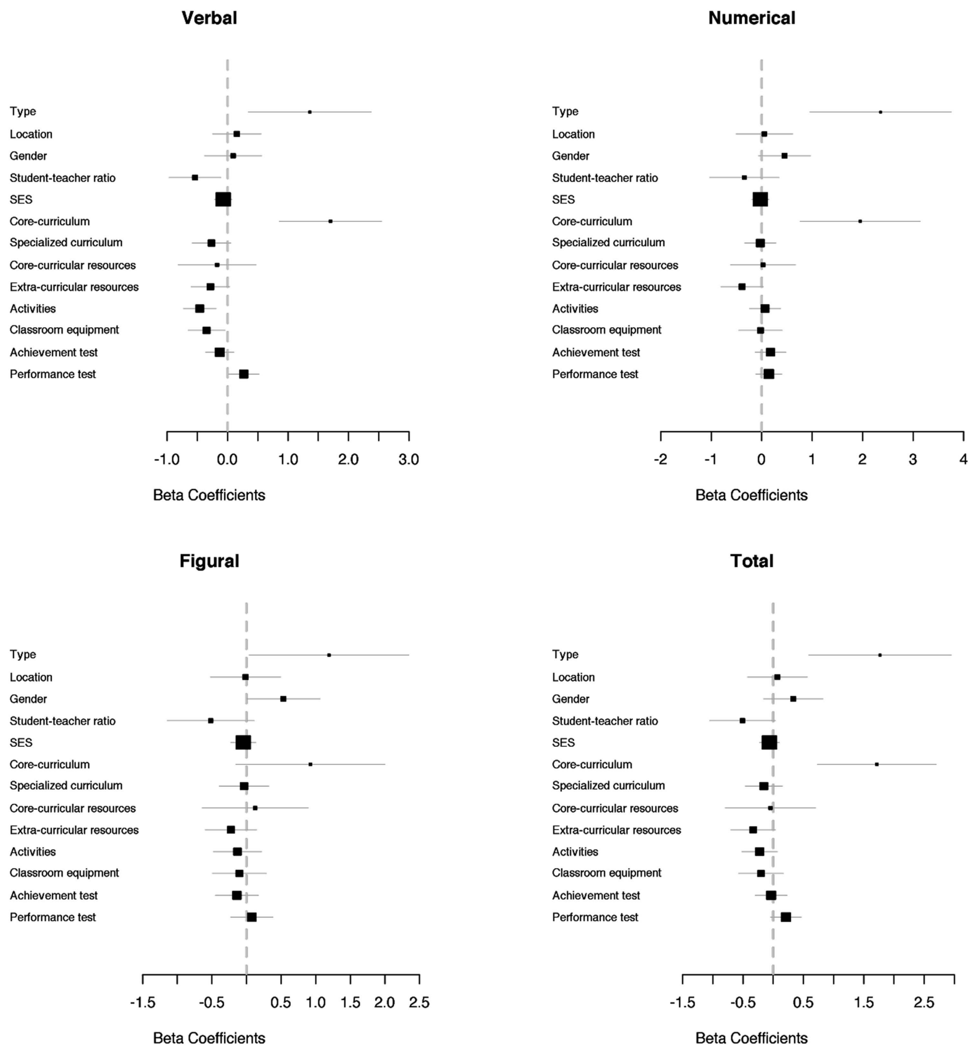

Figure 2. Standardized regression weights (β) and associated standard errors of school factors for three indicators of intelligence as well as the total score. SES = socioeconomic status.

in public compared to private schools; and (d) school contextual factors—especially the amount of core-curricular courses offered at a school—to matter more than expected given the strongly centralized Saudi school system. Compared to findings from large-scale studies that have investigated a unitary construct of IQ internationally for narrow age groups (e.g., Lynn & Meisenberg, 2010; Rindermann, 2007), to our knowledge, this study is the first systematic investigation of different indicators of intelligence in children of a broad age range in Saudi Arabia.

Negligible Gender Differences in Intelligence

With respect to individual factors, our findings were generally consistent with our expectations: Age was positively related to the three intelligence subtest scores, and gender differences were rather small in magnitude. The strongest effect for gender was found for the figural subtest in favor of females. The direction and magnitude of these findings are in line with other research; that is, the magnitude of gender differences varies with the specific cognitive ability being assessed, but the overall low gender differences found in the present study fall in the range typically found in a majority of studies (Feingold, 1992; Halpern, 1997; Hyde, 2005; Lynn et al., 2005). Complex psychobiosocial models (e.g., Halpern, 1997; Halpern & LaMay, 2000) have highlighted their major premise that neither biological nor environmental factors alone can explain differences in performance on cognitive ability tests and that gender differences are mutually attributable to both identifiable biological factors and socialization factors (Nisbett et al., 2012). We are therefore cautious in drawing conclusions about the nature of the findings. It has been argued that females score higher on tests in which questions are similar to material that was learned in school, whereas males score higher on tests of mathematics and science that are not directly tied to their school curriculum (Halpern, 2004; Halpern et al., 2007). Moreover, recent reports in the United States highlighted that males, on average, tend not to outperform girls on standardized tests that reflect more closely the mathematics content that is taught in school (Hyde, Lindberg, Linn, Ellis, & Williams, 2008). Thus, the figural subtest of Aurora-*g* supposedly assesses content that is more closely related to the curriculum taught at girls' schools. Similarly, the rather marginally higher scores of males on the numerical subtest in this study seems to reflect that the content of this subtest is linked to the mathematics curriculum taught at boys' schools as well.

For the context of Saudi Arabia, these findings are particularly interesting. The results of this study seem to imply a small gender gap in mean performance on tests that assess intellectual skills. Interestingly, this pattern is also highlighted in the changes of math test performance in the Trends in International Mathematics and Science Study results, in which Grade 8 boys outperformed girls by 10 points in 2003 (Mullis, Martin, Gonzalez, & Chrostowski, 2004) but girls outperformed boys in 2007 by 22 points (Mullis, Martin, & Foy, 2008), then again in 2011 by 16 points (Mullis et al., 2008). However, the question remains whether overrepresentation of males at the extreme ends of the distribution implies that a gender gap remains in access and admission to special programs, for example, for gifted students. That is, the results of this study imply that a greater number of males than females will qualify for advanced training that places emphasis on numerical skills (Halpern et al., 2007), and that more boys will be selected for participation in gifted programs in Saudi Arabia if the identification of gifted students stresses numerical skills over verbal and figural skills. Although gender-segregated education has a long history in Saudi Arabia, the overall finding of small gender differences in general intelligence seems to reflect the diminished effect of other gender differentiated environmental norms and inputs in this country (Halpern, 1997). To a certain degree, the small magnitude of differences ties in to the conclusions in Western literature that there is no evidence for the advantage of single-sex schools in improving students' average intellectual skills (Halpern et al., 2011; Mael, Alonso, Gibson, Rogers, & Smith, 2005).

School Matters

Against what one would expect based on the strongly centralized approach of the Saudi government toward education, we found a substantial variability in children's intelligence test scores to be conditioned by placement in different schools. Moreover, a striking finding was that up to about 72% of the observed variance in mean intelligence test scores among schools could be explained by differences in school contextual variables, mainly attributable to organizational characteristics such as school type and curricular context.

Regarding school type, we found that the variability in children's intelligence test scores appeared to be split along public versus private school lines but not in accordance with previous findings from Western contexts. Specifically, results showed that students' mean scores in public schools were higher in all intelligence subtests. This finding is even more surprising given that private schools in our sample offer a higher number of core, special, and extra courses; a factor that was positively related to mean intelligence scores. This higher number of courses in private schools is consistent with previous observations, as many private schools in Saudi Arabia seek to differentiate themselves from public schools by adding English, math, and sciences courses taught in English to their curriculum (Mandelman et al., 2010). With this difference in mind, an explanation for the advantage of public over private schools in line with other research (e.g., Psacharopoulos, 1987) could be that parental pressure (social demand) exerted on private schools makes them emphasize subjects that parents view as particularly helpful for their child (and that may be the reason they send their child there). Given the lack of data to compare the private schools in this study to other private schools in Saudi Arabia, we are cautious in drawing strong conclusions at this point. However, over the course of this study, anecdotal evidence has indicated that teachers in private schools often come from other Middle Eastern or Northern African countries, and that private schools often provide rather low salaries compared to public schools. This might also set a different general level of training and experience for teachers at private compared to public schools, which could have an impact on student achievement, especially in Gulf countries where the emphasis has traditionally been placed more on the number of teachers and low student–teacher ratios than on quality of teaching (M. Barber et al., 2007). In addition, average private schools have been selected in this study that ought to be more comparable to average public schools than to other private schools in the country. Although there is a lack of data supporting this claim, it has been reported that often those children who do not perform well in public schools are sent to private schools.

To draw stronger conclusions, future studies should examine both the cognitive performance and academic achievement of children at private and public schools in addition to a more detailed characterization of these schools, along with other student and home background characteristics. For example, it has been noted that variations in school attributes, proximity, and fees across neighborhoods affect the decision of parents to send their children to public or private schools, but also that private schools are utilized by poor households (e.g., in Pakistan; Alderman, Orazem, & Paterno, 2001; Aslam, 2009). This observation is reflected in the similar levels of school-level socioeconomic status in the present study and the lack of any relationship between the socioeconomic status of the student body and the three intelligence subtests.

Regarding the curricular context of the schools, we found consistent evidence that children in schools with a higher number of core, special, and extra courses scored higher in the verbal and numerical intelligence subtests. Other researchers have highlighted the associations between the amount of instruction and cognitive competencies as well (Baenninger & Newcombe, 1995; Ceci, 1991; Fuller & Heyneman, 1989; Rutter, 1983). As Christian et al. (2001) pointed out, cognitive skills change as a function of relatively specific experiences in the instructional environment. Thus, a higher number of courses in the curriculum seem to specifically enrich or stimulate verbal and numerical, but not figural skills. However, it is not possible to draw further conclusions based on the existing data. For example, we analyzed a rather undifferentiated amount of instruction, quantified in general approximation of the number of courses, but not differentiated by topic. We did not have detailed information on the differentiation of the curriculum available that could, for example, inform the differences between private and public schools, and how these differences might be reflected in other academic achievement tests rather than mean intelligence test scores. Future studies should examine in more detail the number of on-topic instructional minutes as a predictor of domain-specific achievement (Rindermann & Ceci, 2009) or the official duration of lesson time that is actually spent on task-related work (Scheerens, 1990).

Limitations

Two aspects have to be acknowledged here that limit the generalizability of the findings in this study. First, not all the variance in student intelligence test scores at the school level can be attributed to the effects of schools. Some of that variance is due to the individual background characteristics of the students, which affect student outcomes no matter where they attend school. Second, in a similar vein, the figures and patterns of findings derived from multilevel analyses are averages and therefore uncontrolled for other characteristics in which students attending the schools may differ (e.g., controlling for family background and aptitude). Gender and age only explained a small amount of variation on the student level, which highlights the importance of examining other correlates of students' cognitive performance in future studies (e.g., socioeconomic status and proximal indicators of the home and family environment; Seifer, 2001). Answering the question of whether schools influence students' cognitive test scores after accounting for their family background could inform future research (and policy decisions) on the relative importance of schools and family life for the cognitive development of children in Saudi Arabia.

Conclusion

Formal education is a nearly universal experience among children and is perhaps the most significant environmental influence on cognitive development beyond the family. We embarked on this study with the motivation to augment findings from Western literature about school effects and gender differences with respect to cognitive abilities by examining data from a largely understudied country: Saudi Arabia. As was outlined in this study, different aspects of the educational system such as the school type and the diversity of the curriculum imply different formal schooling experiences for children and are related to differences in intellectual functioning. We conclude that the factors from the school environment that were found to be associated with different indicators of intelligence in this study could provide an intellectually stimulating education, keeping the children cognitive challenged, thereby fostering the development of general intelligence. However, it remains to be investigated to what extent these empirical findings are generalizable across other cultural contexts and to what extent there is cross-cultural variation in the patterns of findings. This study will hopefully spark the motivation of other researchers to embark on a similar endeavor to add to the particular scarcity of research from countries in the Middle East. Studies with this focus may add to a more comprehensive theory on sources of individual differences in intelligence, but may also help us to understand more of the culturally related phenomena that work indirectly though education to influence the acquisition of intellectual abilities of children in diverse parts of the world, who are surely connected by an ever-strengthening globalization.

References

Alarfaj, A. (2011). *Developing a classroom science enrichment programme for gifted primary school boys in Saudi Arabia* (Unpublished doctoral thesis). University of Southampton, Southampton, England.

Alderman, H., Orazem, P. F., & Paterno, E. M. (2001). School quality, school cost, and the public/private school choices of low-income households in Pakistan. *Journal of Human Resources, 36,* 304–326. doi:10.2307/3069661

Aljughaiman, A. M., & Grigorenko, E. L. (2013). Growing up under pressure: The cultural and religious context of the Saudi system of gifted education. *Journal for the Education of the Gifted, 36,* 307–322. doi:10.1177/0162353213493153

Al-Sadan, I. A. (2000). Educational assessment in Saudi Arabian schools. *Assessment in Education: Principles, Policy & Practice, 7,* 143–155. doi:10.1080/713613320

Ammermueller, A. (2012). Institutional features of schooling systems and educational inequality: Cross-country evidence from PIRLS and PISA. *German Economic Review, 14,* 190–213. doi:10.1111/j.1468-0475.2012.00565.x

Arnett, J. J. (2002). The psychology of globalization. *American Psychologist, 57,* 774–783. doi:10.1037/0003-066X.57.10.774

Aslam, M. (2009). The relative effectiveness of government and private schools in Pakistan: Are girls worse off? *Education Economics, 17,* 329–354. doi:10.1080/09645290903142635

Baenninger, M., & Newcombe, N. (1995). Environmental input to the development of sex-related differences in spatial and mathematical ability. *Learning and Individual Differences, 7,* 363–379. doi:10.1016/1041-6080(95)90007-1

Baker, D. P., Goesling, B., & Letendre, G. K. (2002). Socioeconomic status, school quality, and national economic development: A cross-national analysis of the "Heyneman–Loxley Effect" on mathematics and science achievement. *Comparative Education Review, 46,* 291–312. doi:10.1086/341159

Baltes, P. B., Staudinger, U. M., & Lindenberger, U. (1999). Lifespan psychology: Theory and application to intellectual functioning. *Annual Review of Psychology, 50,* 471–507. doi:10.1146/annurev.psych.50.1.471

Barber, M., Mourshed, M., & Whelan, F. (2007). Improving education in the Gulf. *The McKinsey Quarterly,* 39–47.

Barber, N. (2005). Educational and ecological correlates of IQ: A cross-national investigation. *Intelligence, 33,* 273–284. doi:10.1016/j.intell.2005.01.001

Barr, R., & Dreeben, R. (1983). *How schools work*. Chicago, IL: University of Chicago Press.

Battistich, V., Solomon, D., Kim, D., Watson, M., & Shaps, E. (1995). Schools as communities, poverty levels of student populations, and students' attitudes, motives, and performance: A multilevel analysis. *American Educational Research Journal, 32,* 627–658. doi:10.3102/00028312032003627

Benveniste, L., Carnoy, M., & Rothstein, R. (2003). *All else equal: Are public and private schools different?* New York, NY: RoutledgeFalmer.

Borland, M. V., & Howsen, R. M. (2003). An examination of the effect of elementary school size on student academic achievement. *International Review of Education, 49,* 463–474. doi:10.1023/A:1026348922511

Borland, M. V., Howsen, R. M., & Trawick, M. W. (2005). An investigation of the effect of class size on student academic achievement. *Education Economics, 13,* 73–83. doi:10.1080/0964529042000325216

Bronfenbrenner, U., & Morris, P. (2006). The bioecological model of human development. In R. M. Lerner & W. Damon (Eds.), *Handbook of child psychology* (pp. 793–828). Hoboken, NJ: Wiley.

Buchmann, C., & Hannum, E. (2001). Education and stratification in developing countries: A review of theories and research. *Annual Review of Sociology, 27,* 77–102. doi:10.1146/annurev.soc.27.1.77

Ceci, S. J. (1991). How much does schooling influence general intelligence and its cognitive components? A reassessment of the evidence. *Developmental Psychology, 27,* 703–722. doi:10.1037/0012-1649.27.5.703

Ceci, S. J., & Williams, W. M. (1997). Schooling, intelligence, and income. *American Psychologist, 52,* 1051–1058. doi:10.1037/0003-066X.52.10.1051

Center for Religious Freedom. (2008). *Saudi Arabia's curriculum of intolerance*. Washington, DC: Center for Religious Freedom.

Center for Religious Freedom. (2011). *Ten years on Saudi Arabia's textbooks still promote religious violence*. Washington, DC: Center for Religious Freedom.

Chart, H. E., Grigorenko, E. L., & Sternberg, R. J. (2008). Identification: The Aurora Battery. In J. A. Plucker & C. M. Callahan (Eds.), *Critical issues and practices in gifted education* (pp. 345–365). Waco, TX: Prufrock Press.

Chen, Z., & Siegler, R. S. (2000). Intellectual development in childhood. In R. J. Sternberg (Ed.), *Handbook of intelligence* (pp. 92–116). Cambridge, England: Cambridge University Press. doi:10.1017/CBO9780511807947.006

Christian, K., Bachnan, H. J., & Morrison, F. J. (2001). Schooling and cognitive development. In R. J. Sternberg & E. L. Grigorenko (Eds.), *Environmental effects on cognitive abilities* (pp. 287–335). Mahwah, NJ: Erlbaum.

Cliffordson, C., & Gustafsson, J.-E. (2008). Effects of age and schooling on intellectual performance: Estimates obtained from analysis of continuous variation in age and length of schooling. *Intelligence, 36,* 143–152. doi:10.1016/j.intell.2007.03.006

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Cohn, E., & Rossmiller, R. A. (1987). Research on effective schools: Implications for less developed countries. *Comparative Education Review, 31,* 377–399. doi:10.1086/446697

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). *Equality of educational opportunity*. Washington, DC: U.S. Department of Health, Education, and Welfare.

Coleman, J., Kilgore, S., & Hoffer, T. (1982). Public and private schools. *Society, 19,* 4–9. doi:10.1007/BF02712901

Deary, I. J., Whalley, L. J., Lemmon, H., Crawford, J. R., & Starr, J. M. (2000). The stability of individual differences in mental ability from childhood to old age: Follow-up of the 1932 Scottish mental survey. *Intelligence, 28,* 49–55. doi:10.1016/S0160-2896(99)00031-8

Deary, I. J., Yang, J., Davies, G., Harris, S. E., Tenesa, A., Liewald, D., . . . Visscher, P. M. (2012). Genetic contributions to stability and change in intelligence from childhood to old age. *Nature, 482,* 212–215. doi:10.1038/nature10781

Demetriou, A., & Papadopoulos, T. C. (2004). Human Intelligence: From local models to universal theory. In R. J. Sternberg (Ed.), *International Handbook of Intelligence* (pp. 445–474). Cambridge, England: Cambridge University Press. doi:10.1017/CBO9780511616648.017

Dronkers, J., & Avram, S. (2010). A cross-national analysis of the relations of school choice and effectiveness differences between private-dependent and public schools: A cross-national analysis. *Educational Research and Evaluation, 16,* 151–175. doi:10.1080/13803611.2010.484977

Dronkers, J., & Robert, P. (2008). Differences in scholastic achievement of public, private government-dependent, and private independent schools. *Educational Policy, 22,* 541–577. doi:10.1177/0895904807307065

Eccles, J. S., & Roeser, R. W. (2012). School influences on human development. In L. C. Mayes & M. Lewis (Eds.), *The Cambridge handbook of environments in human development* (pp. 259–283). New York, NY: Cambridge University Press. doi:10.1017/CBO9781139016827.017

Elliott, J., & Grigorenko, E. L. (2008). Are Western educational theories and practices truly universal? In J. Elliott & E. L. Grigorenko (Eds.), *Western psychological and educational theory in diverse contexts* (pp. 1–4). Abingdon, England: Routledge.

Engle, P. L., Black, M. M., Behrman, J. R., Cabral de Mello, M. C., Gertler, P. J., Kapiriri, L., . . . International Child Development Steering Group. (2007). Child development in developing countries 3. Strategies to avoid the loss of developmental potential in more than 200 million children in the developing world. *Lancet, 369,* 229–242. doi:10.1016/S0140-6736(07)60112-3

Fagan, J. F., Holland, C. R., & Wheeler, K. (2007). The prediction, from infancy, of adult IQ and achievement. *Intelligence, 35,* 225–231. doi:10.1016/j.intell.2006.07.007

Feingold, A. (1992). Sex differences in variability in intellectual abilities: A new look at an old controversy. *Review of Educational Research, 62,* 61–84. doi:10.3102/00346543062001061

Fiese, B. H. (2001). Family matters: A systems view of family effects on children's cognitive health. In R. J. Sternberg & E. L. Grigorenko (Eds.), *Environmental effects on cognitive abilities* (pp. 39–59). Mahwah, NJ: Erlbaum.

Fuller, B. (1987). What school factors raise achievement in the third world? *Review of Educational Research, 57,* 255–292. doi:10.3102/00346543057003255

Fuller, B., & Clarke, P. (1994). Raising school effects while ignoring culture? Local conditions and the influence of classroom tools, rules, and pedagogy. *Review of Educational Research, 64,* 119–157. doi:10.3102/00346543064001119

Fuller, B., & Heyneman, S. P. (1989). Third world school quality current collapse, future potential. *Educational Researcher, 18,* 12–19.

Geary, D. C., Saults, S. J., Liu, F., & Hoard, M. K. (2000). Sex differences in spatial cognition, computational fluency, and arithmetical reasoning. *Journal of Experimental Child Psychology, 77,* 337–353. doi:10.1006/jecp.2000.2594

Glewwe, P., Grosh, M., Jacoby, H., & Lockheed, M. (1995). An eclectic approach to estimating the determinants of achievement in Jamaican primary education. *World Bank Economic Review, 9,* 231–258. doi:10.1093/wber/9.2.231

Glewwe, P. W., Hanushek, E. A., Humpage, S. D., & Ravina, R. (2011). *School resources and educational outcomes in developing countries: A review of the literature from 1990 to 2010 (NBER Working Paper No. 17554)*. Cambridge, MA: National Bureau of Economic Research.

Gow, A. J., Johnson, W., Pattie, A., Brett, C. E., Roberts, B., Starr, J. M., & Deary, I. J. (2011). Stability and change in intelligence from age 11 to ages 70, 79, and 87: The Lothian Birth Cohorts of 1921 and 1936. *Psychology and Aging, 26,* 232–240. doi:10.1037/a0021072

Grantham-McGregor, S., Cheung, Y. B., Cueto, S., Glewwe, P., Richter, L., & Strupp, B. (2007). Developmental potential in the first 5 years for children in developing countries. *Lancet, 369,* 60–70. doi:10.1016/S0140-6736(07)60032-4

Grigorenko, E. L. (2008). Hitting, missing, and in between: A typology of the impact of western education on the non-Western world. In J. Elliott & E. L. Grigorenko (Eds.), *Western psychological and educational theory in diverse contexts* (pp. 165–186). Abingdon, England: Routledge.

Halpern, D. F. (1997). Sex differences in intelligence: Implications for education. *American Psychologist, 52,* 1091–1102. doi:10.1037/0003-066X.52.10.1091

Halpern, D. F. (2004). A cognitive-process taxonomy for sex differences in cognitive abilities. *Current Directions in Psychological Science, 13,* 135–139. doi:10.1111/j.0963-7214.2004.00292.x

Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S., & Gernsbacher, M. A. (2007). The science of sex differences in science and mathematics. *Psychological Science in the Public Interest, 8,* 1–51.

Halpern, D. F., Eliot, L., Bigler, R. S., Fabes, R. A., Hanish, L. D., Hyde, J., . . . Martin, C. L. (2011). The pseudoscience of single-sex schooling. *Science, 333,* 1706–1707. doi:10.1126/science.1205031

Halpern, D. F., & LaMay, M. L. (2000). The smarter sex: A critical review of sex differences in intelligence. *Educational Psychology Review, 12,* 229–246. doi:10.1023/A:1009027516424

Hanushek, E. A. (1995). Interpreting recent research on schooling in developing countries. *World Bank Research Observer, 10,* 227–246. doi:10.1093/wbro/10.2.227

Hanushek, E. A. (1999). Some findings from an independent investigation of the Tennessee STAR experiment and from other investigation of class size effects. *Educational Evaluation and Policy Analysis, 21,* 143–163. doi:10.3102/01623737021002143

Hanushek, E. A. (2006). School resources. In E. A. Hanushek & F. Welch (Eds.), *Handbook of the economics of education* (Vol. 2, pp. 865–908). Amsterdam, the Netherlands: North-Holland.

Hanushek, E. A., & Luque, J. A. (2003). Efficiency and equity in schools around the world. *Economics of Education Review, 22,* 481–502. doi:10.1016/S0272-7757(03)00038-4

Harkness, S., & Super, C. M. (2012). The cultural organization of children's environments. In L. C. Mayes & M. Lewis (Eds.), *The Cambridge handbook of environments in human development* (pp. 498–516). New York, NY: Cambridge University Press. doi:10.1017/CBO9781139016827.028

Hattie, J. (2005). The paradox of reducing class size and improving learning outcomes. *International Journal of Educational Research, 43,* 387–425. doi:10.1016/j.ijer.2006.07.002

Heyneman, S. P. (1997). The quality of education in the Middle East and North Africa (MENA). *International Journal of Educational Development, 17,* 449–466. doi:10.1016/S0738-0593(97)00022-9

Heyneman, S. P., & Loxley, W. A. (1983). The effect of primary-school quality on academic achievement across twenty-nine high- and low-income countries. *American Journal of Sociology, 88,* 1162–1194. doi:10.1086/227799

Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist, 60,* 581–592. doi:10.1037/0003-066X.60.6.581

Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., & Williams, C. C. (2008). Gender similarities characterize math performance. *Science, 321,* 494–495. doi:10.1126/science.1160364

Jimenez, E., Lockheed, M. E., & Paqueo, V. (1991). The relative efficiency of private and public schools in developing countries. *World Bank Research Observer, 6,* 205–218. doi:10.1093/wbro/6.2.205

Johnson, W., & Bouchard, T. J., Jr. (2007). Sex differences in mental abilities: *g* masks the dimensions on which they lie. *Intelligence, 35,* 23–39. doi:10.1016/j.intell.2006.03.012

Khoo, C., & Khoo, L. (2005). Effects of resources, inequality, and privilege bias on achievement: Country, school, and student level analyses. *American Educational Research Journal, 42,* 575–603. doi:10.3102/00028312042004575

Kremer, M. R. (1995). Research on schooling: What we know and what we don't know. A comment on Hanushek. *World Bank Research Observer, 10,* 247–254. doi:10.1093/wbro/10.2.247

Larsen, L., Hartmann, P., & Nyborg, H. (2008). The stability of general intelligence from early adulthood to middle-age. *Intelligence, 36,* 29–34. doi:10.1016/j.intell.2007.01.001

Lee, V. E. (2000). Using hierarchical linear modeling to study social contexts: The case of school effects. *Educational Psychologist, 35,* 125–141. doi:10.1207/S15326985EP3502_6

Lee, V. E. (2001). *Restructuring high schools for equity and excellence: What works*. New York, NY: Teachers College Press.

Lee, V. E., & Zuze, T. L. (2011). School resources and academic performance in Sub-Saharan Africa. *Comparative Education Review, 55,* 369–397. doi:10.1086/660157

Leithwood, K., & Jantzi, D. (2009). A review of empirical evidence about school size effects: A policy perspective. *Review of Educational Research, 79,* 464–490. doi:10.3102/0034654308326158

Lubienski, C., Crane, C., & Lubienski, S. T. (2008). What do we know about school effectiveness? Academic gains in public and private schools. *Phi Delta Kappan, 89,* 689–695.

Lynn, R., Fergusson, D. M., & Horwood, L. J. (2005). Sex differences on the WISC–R in New Zealand. *Personality and Individual Differences, 39,* 103–114. doi:10.1016/j.paid.2004.12.009

Lynn, R., & Meisenberg, G. (2010). National IQs calculated and validated for 108 nations. *Intelligence, 38,* 353–360. doi:10.1016/j.intell.2010.04.007

Maccoby, E. E. (1992). The role of parents in the socialization of children: An historical overview. *Developmental Psychology, 28,* 1006–1017. doi:10.1037/0012-1649.28.6.1006

Mackintosh, N. J. (1996). Sex differences and IQ. *Journal of Biosocial Science, 28*(04). doi:10.1017/S0021932000022586

Mael, F., Alonso, A., Gibson, D., Rogers, K., & Smith, M. (2005). *Single-sex versus coeducational schooling: A systematic review*. Washington, DC: U.S. Department of Education, Office of Planning, Evaluation and Policy Development.

Mandelman, S. D., Tan, M., Aljughaiman, A. M., & Grigorenko, E. L. (2010). Intellectual giftedness: Economic, political, cultural, and psychological considerations. *Learning and Individual Differences, 20,* 287–297. doi:10.1016/j.lindif.2010.04.014

Mayer, R. E. (2000). Intelligence and education. In R. J. Sternberg (Ed.), *Handbook of intelligence* (pp. 519–533). Cambridge, England: Cambridge University Press. doi:10.1017/CBO9780511807947.024

McLoyd, V. C. (1998). Socioeconomic disadvantage and child development. *American Psychologist, 53,* 185–204. doi:10.1037/0003-066X.53.2.185

Ministry of Education, Kingdom of Saudi Arabia. (2008, November). *National report on education development in the Kingdom of Saudi Arabia*. Paper presented at the 48th Session of the International Conference on Education, Geneva, Switzerland.

Mobaraki, A. E. H., & Söderfeldt, B. (2010). Gender inequity in Saudi Arabia and its role in public health. *Eastern Mediterranean Health Journal, 16,* 113–118.

Mullis, I. V. S., Martin, M. O., & Foy, P. (2008). *TIMSS 2007 international mathematics report: Findings from IEA's Trends in International Mathematics and Science Study at the fourth and eighth grades*. Chestnut Hill, MA: Boston College.

Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., & Chrostowski, S. J. (2004). *TIMSS 2003 international mathematics report: Findings from IEA's Trends in International Mathematics and Science Study at the fourth and eighth grades*. Chestnut Hill, MA: Boston College.

Muthén, B. O. (1991). Multilevel factor analysis of class and student achievement components. *Journal of Educational Measurement, 28,* 338–354. doi:10.1111/j.1745-3984.1991.tb00363.x

Muthén, B. O. (1994). Multilevel covariance structure analysis. *Sociological Methods and Research, 22,* 376–398. doi:10.1177/0049124194022003006

Muthén, L. K., & Muthén, B. O. (2010). *Mplus (Version 6.11)* [Computer software]. Los Angeles, CA: Muthén & Muthén.

Neisser, U., Boodoo, G., Bouchard, T. J., Jr., Boykin, A. W., Ceci, S. J., Halpern, D. F., . . . Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51,* 77–101. doi:10.1037/0003-066X.51.2.77

Nisbett, R. E. (2009). *Intelligence and how to get it: Why schools and cultures count*. New York, NY: Norton.

Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., & Turkheimer, E. (2012). Intelligence: New findings and theoretical developments. *American Psychologist, 67,* 130–159. doi:10.1037/a0026699

Peugh, J. L. (2010). A practical guide to multilevel modeling. *Journal of School Psychology, 48,* 85–112. doi:10.1016/j.jsp.2009.09.002

Psacharopoulos, G. (1987). Public versus private schools in developing countries: Evidence from Colombia and Tanzania. *International Journal of Educational Development, 7,* 59–67. doi:10.1016/0738-0593(87)90007-1

Ramey, S. L., & Ramey, C. T. (2012). Understanding the developmental influence of the family environment. In L. C. Mayes & M. Lewis (Eds.), *The Cambridge handbook of environment in human development* (pp. 222–242). New York, NY: Cambridge University Press. doi:10.1017/CBO9781139016827.015

Ramsden, S., Richardson, F. M., Josse, G., Thomas, M. S., Ellis, C., Shakeshaft, C., . . . Price, C. J. (2011). Verbal and non-verbal intelligence changes in the teenage brain. *Nature, 479,* 113–116. doi:10.1038/nature10514

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage.

Riddell, A. R. (1997). Assessing designs for school effectiveness research and school improvement in developing countries. *Comparative Education Review, 41,* 178–204. doi:10.1086/447429

Rindermann, H. (2007). The *g*-factor of international cognitive ability comparisons: The homogeneity of results in PISA, TIMSS, PIRLS and IQ-tests across nations. *European Journal of Personality, 21,* 667–706. doi:10.1002/per.634

Rindermann, H., & Ceci, S. J. (2009). Educational policy and country outcomes in international cognitive competence studies. *Perspectives on Psychological Science, 4,* 551–577. doi:10.1111/j.1745-6924.2009.01165.x

Rogoff, B. (2003). *The cultural nature of human development*. Oxford, England: Oxford University Press.

Rugh, W. A. (2002a). Arab education: Tradition, growth and reform. *Middle East Journal, 56,* 396–414.

Rugh, W. A. (2002b). Education in Saudi Arabia: Choices and constraints. *Middle East Policy, 9,* 40–55. doi:10.1111/1475-4967.00056

Rutter, M. (1983). School effects on pupil progress: Research findings and policy implications. *Child Development, 54,* 1–29. doi:10.2307/1129857

Sander, W. (1999). Private schools and public school achievement. *Journal of Human Resources, 34,* 697–709. doi:10.2307/146413

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications to developmental research* (pp. 399–419). Thousand Oaks, CA: Sage.

Scheerens, J. (1990). School effectiveness research and the development of process indicators of school functioning. *School Effectiveness and School Improvement, 1,* 61–80. doi:10.1080/0924345900010106

Scheerens, J. (2001). Monitoring school effectiveness in developing countries. *School Effectiveness and School Improvement, 12,* 359–384. doi:10.1076/sesi.12.4.359.3447

Seifer, R. (2001). Socioeconomic status, multiple risks, and development of intelligence. In R. J. Sternberg & E. L. Grigorenko (Eds.), *Environmental effects on cognitive abilities* (pp. 59–82). Mahwah, NJ: Erlbaum.

Simmons, J., & Alexander, L. (1978). The determinants of school achievement in developing countries: A review of the research. *Economic Development and Cultural Change, 26,* 341–357. doi:10.1086/451019

Sternberg, R. J. (2008). Culture, instruction, and assessment. In J. Elliott & E. L. Grigorenko (Eds.), *Western psychological and educational theory in diverse contexts* (pp. 5–22). Abingdon, England: Routledge.

Tan, M., Aljughaiman, A. M., Elliott, J. G., Kornilov, S. A., Ferrando-Prieto, M., Bolden, D. S., . . . Grigorenko, E. L. (2009). Considering language, culture, and cognitive abilities: The international translation and adaptation of the Aurora Assessment Battery. In E. L. Grigorenko (Ed.), *Multicultural psychoeducational assessment* (pp. 443–468). New York, NY: Springer.

UNESCO. (2011). *World data on education: Saudi Arabia*. Washington, DC: International Bureau of Education.

Wagner, D. A., & Spratt, J. E. (1987). Cognitive consequences of contrasting pedagogies: The effects of Quranic preschooling in Morocco. *Child Development, 58,* 1207–1219. doi:10.2307/1130615

Walker, S. P., Wachs, T. D., Meeks Gardner, J., Lozoff, B., Wasserman, G. A., Pollitt, E., & Carter, J. A. (2007). Child development: Risk factors for adverse outcomes in developing countries. *Lancet, 369,* 145–157. doi:10.1016/S0140-6736(07)60076-2

Wiseman, A. W., Sadaawi, A., & Alromi, N. H. (2008, September). *Educational indicators and national development in Saudi Arabia*. Paper presented at the Third IEA International Research Conference, Taipei, Taiwan.

Received March 12, 2013
Revision received June 24, 2014
Accepted June 27, 2014 ■