Spatial and Numerical Predictors of Measurement Performance: The Moderating Effects of Community Income and Gender

Beth M. Casey, Eric Dearing, Marina Vasilyeva, Colleen M. Ganley, and Michele Tine
Boston College

Spatial reasoning and numerical predictors of measurement performance were investigated in 4th graders from low-income and affluent communities. Predictors of 2 subtypes of measurement performance (spatial-conceptual and formula based) were assessed while controlling for verbal and spatial working memory. Consistent with prior findings, students from the affluent community outperformed students from the low-income community on all measures examined. More importantly, the study revealed different patterns of relations between cognitive skills and academic performance in the 2 communities. Specifically, spatial skills were related to measurement performance in the affluent but not in the low-income community. These findings demonstrate that socioeconomic context impacts not only children’s levels of performance but also their capacity to apply basic cognitive skills, like spatial reasoning, to their academic performance.

Keywords: sex differences, gender, measurement, income level, spatial skills, mathematics

Measurement is one of the most widely used applications of mathematics in everyday life (Lehrer, 2003; Wilson & Rowland, 1993). Moreover, measurement plays a unique and critical role in mathematical and scientific understanding: It provides a link between continuous and discrete quantity. Researchers have long considered the measurement of space (including length, perimeter, area, and volume) to involve the integration of numerical and spatial thinking (Battista, 2003; Lehrer, 2003; Miller, 1989; Shaw & Puckett-Cliatt, 1989; Stephan & Clements, 2003; Wilson & Rowland, 1993). According to the National Council of Teachers of Mathematics (2000, p. 103), measurement “bridges two main areas of school mathematics—geometry and number.”

Students who are effective at measurement have good “measurement sense,” which means that they have a conceptual understanding of the processes underlying measurement procedures (Joram, 2003; Shaw & Puckett-Cliatt, 1989). For example, Battista (2003) described the underlying processes in gaining competence in measuring area and volume as understanding how to enumerate arrays of squares and cubes; he identified two mental processes essential to meaningful structuring of arrays: (a) forming and using mental models and (b) spatial structuring. Thus, effective measurement performance involves an understanding of the underlying spatial nature of measurement, as well as the numerical and procedural competence to use measuring tools and apply formulas.

Subtypes of Measurement

Although researchers in the field of measurement have made the assumption that effective measurement performance depends on spatial reasoning as well as numerical skills, the relation between these cognitive skills and measurement performance has not been directly investigated. In fact, there has been very little empirical research exploring the factors predicting success in measurement. Thus, a major goal of the present research was to determine whether numerical and spatial reasoning do, indeed, predict measurement performance in fourth-grade elementary school students. However, it is important to recognize that measurement is not a unitary construct, and thus numerical and spatial predictors may vary as a function of the type of skill assessed. A new fourth-grade assessment tool covering a broad range of measurement concepts (Vasilyeva, Ludlow, Casey, & St. Onge, 2009) revealed two major categories of measurement items: those tapping formula-based skills and those tapping spatial-conceptual skills. Recent findings (Vasilyeva, Casey, Dearing, & Ganley, 2009) confirm that students approach these formula-based and spatial-conceptual measurement problems using very different strategies.

Formula-based problems can be solved with known measurement formulas. For example, a formula problem, in which the child is given the length and width of a rectangle and is asked to determine the area of the rectangle, involves recalling the appropriate formula (Area = Length × Width) and carrying out the correct calculation. These types of items are considered to be analytical as opposed to spatial because solving them does not
need to involve spatial reasoning. That is, to solve these problems, children do not have to examine spatial relations between objects and units of measure; nor do they need to generate or manipulate spatial images. They can simply apply known formulas using numeric techniques. Many students may treat formula-based measurement problems as a purely numerical problem (e.g., Length × Width) without considering the spatial nature of the problem.

However, individuals with good measurement sense may draw on their spatial as well as numeric skills to solve formula-based problems. For example, they may visualize the object to be measured in order to select the most relevant formula to use. Even though the use of spatial reasoning to solve formula-based problems can be effective, it is not necessary for obtaining the correct answer, and not all students who are successful at solving these types of problems choose to use spatial reasoning. Thus, although numeric skills should relate to formula-based measurement, it is not clear whether spatial thinking will also be related for performance on this type of measurement problem.

In contrast to formula-based problems, effectively solving spatial–conceptual measurement items often depends on an understanding of the spatial relations underlying measurement procedures. The spatial component of these types of measurement items involves forming and manipulating mental images as well as reasoning about spatial relations between units and objects to be measured (Battista, 2003). In many cases, to solve a measurement problem without the use of formulas, it is necessary to subdivide space into equal parts and visualize unit structures (Battista, 2003; Campbell, Watson, & Collis, 1992; Outhred & McPhail, 2000). It should be noted that even though spatial–conceptual problems cannot be solved simply by relying on known formulas, they do depend, in part, on numeric skills. For example, they often involve (a) counting or estimating the number of units in a line, (b) calculating the number of square units in rows or columns within an area problem, or (c) calculating the number of cubes in a volume problem. Thus, we hypothesize that for spatial–conceptual problems, both spatial and numeric skills would be likely to predict measurement performance.

Rationale for the Numerical and Spatial Predictors

There has been little analysis of the specific types of spatial and numerical skills likely to be related to measurement performance. As discussed above, skills relating to counting and computations involving addition, subtraction, and multiplication would be the most likely numerical skills associated with measurement. Thus, we used items from the National Assessment of Educational Progress (NAEP) and Trends in International Mathematics and Science Study (TIMSS) fourth-grade assessments of numerical skills as our assessment tool (Mullis, Martin, Gonzalez, & Chrostowski, 2004; National Center for Education Statistics [NCES], 2004).

For spatial reasoning skills, the ability to generate and manipulate images would be useful for solving measurement problems (Battista, 2003; Shaw & Puckett-Cliatt, 1989). Both mental rotation and spatial visualization skills tap this type of reasoning, and they have been found to relate to mathematics performance (Casey, Nuttall, Pezaris, & Benbow, 1995; Sherman, 1979). Mental rotation involves the ability to generate images and to mentally rotate them in three-dimensional space. Spatial visualization involves the multistep process of generating images of different shapes and then mentally moving and combining those images to generate a new design (e.g., puzzles, tangrams, and block design tasks). There are several ways that these types of spatial reasoning are related to measurement. First, as indicated above, measuring often involves the ability to generate mental images, as, for example, in generating images of different size units (Shaw & Puckett-Cliatt, 1989), or generating models of arrays, such as visualizing the hidden cubes in a volume problem (Battista, 2003). Second, these images often need to be mentally manipulated, as in estimation of length, when the image of a unit is iterated by mentally moving it along the length of an object (Anderson, 2003; Bright, 1976; Shaw & Puckett-Cliatt, 1989). Thus, in the present study, we used a composite measure comprising mental rotation and spatial visualization to assess spatial reasoning skills.

Yet, there is one problem with pinpointing spatial and numerical skills as key predictors of measurement. More basic cognitive processes such as verbal and spatial working memory may underlie the relationship between these skills and measurement performance. Miyake, Friedman, Rettinger, Shah, and Hegarty (2001) found that spatial working memory was strongly implicated in the interrelations among different types of spatial measures, including spatial visualization and mental rotation tasks. Thus, individual differences in spatial working memory could be responsible for relations between spatial skills and measurement.

There are numerous studies showing a relation between verbal working memory and math performance (Hitch, 1978; Holmes & Adams, 2006; Logie, Baddeley, Mané, Donchin, & Sheptak, 1989), and several recent studies have shown the relation between spatial working memory and mental arithmetic (Barrouillet & Lépine, 2005; DeStefano & LeFevre, 2004; Kyttilä & Lehto, 2008). Furthermore, childhood poverty is inversely related to spatial working memory in young adults, and this prospective relationship has been shown to be mediated by elevated chronic stress during childhood (Evans & Schamberg, 2009). Given that we are studying children from both low-income and affluent communities in the present study, it was important to control for spatial and verbal working memory processes to determine whether spatial and numeric reasoning skills are critical in accounting for measurement performance, above and beyond the role of working memory.

Individual Differences in Measurement Skills

Income-Level Differences

The most recent findings on differences in overall math performance (NCES, 2009) show an effect size for the difference between children from low-income and more advantaged communities as great as 1.5 standard deviations. It is of particular relevance to the present study that large-scale national and international assessments show that measurement is the mathematical strand showing the widest achievement gap among children from different economic backgrounds (Barrett & Clements, 2003; Lehrer, Jenkins, & Osana, 1998; Lokan, Ford, & Greenwood, 1996; Lubinski, 2002; Strutchens & Silver, 2000).

A substantial body of research has examined factors contributing to observed differences in math performance as a function of income level. On the basis of this literature, poorer performance of
children in low-income communities can be explained, in large part, by the fact that, on average, they receive limited learning stimulation in their families, schools, and communities (Becker & Tomes, 1986; Yeung, Linver, & Brooks-Gunn, 2002). Children in low-income communities, for example, receive less exposure to cognitively stimulating materials (e.g., puzzles, building blocks, and construction sets) and activities (e.g., parent–child interaction around math) than children in more affluent communities (Dearing & Taylor, 2007; Votruba-Drzal, 2003). As a result of this limited learning stimulation, children are less likely to develop the cognitive skills necessary for high academic performance (Entwisle, Alexander, & Olson, 1994; Heckman, 2008). Thus, in comparison with the students from an affluent community, we would expect students from a low-income community to have lower numeric and spatial skills as well as lower measurement performance.

In addition, we suspect that the general environmental deprivation often occurring within low-income communities (Evans, 2004) has another effect: It constrains the effective application of these cognitive skills to academic performance. That is, children in low-income communities may not have the opportunities and affordances to make the connection between the cognitive skills they do possess and the mathematical tasks presented in the classroom. In fact, there is research on situated cognition (Nunes & Bryant, 1996; Rogoff, 1990; Saxe, 2002; Saxe, Dawson, Fall, & Howard, 1996) that shows that children may develop a relatively high level of a particular cognitive skill but apply it only within the specific context in which the skills were acquired.

In the present study, we were interested in examining a potential association between cognitive skills (spatial and numeric reasoning) and students’ performance on measurement problems, with community socioeconomics as a possible moderator of this association. More specifically, a goal of this study was to compare the pattern of predictive relations for students living in a low-income community with students living in an affluent community. Good spatial skills should predict good measurement performance for children growing up in more affluent communities. This association, however, may be significantly weaker for children growing up in low-income communities, if deprived home learning contexts limit the application of spatial skills to the types of measurement problems often encountered in school.

As an example, children from low-income communities are, on average, likely to have fewer opportunities to practice their spatial skills in ways that are intellectually similar to school problem solving (e.g., activity books that include puzzles, patterns, and mazes). As a result, students from low-income communities may have more difficulty than those from affluent communities applying their spatial skills to the types of measurement problems encountered in school. In other words, even when children growing up poor possess good spatial reasoning, we suspect that their deprived home contexts may limit their ability to transfer these spatial competencies into academic settings and, more specifically, measurement performance. The differences in opportunities within the home environment take on added importance when considering that spatial reasoning skills are usually not directly taught within school, whereas numeric skills are taught extensively in classrooms.

**Gender Differences**

Though less extreme than the effects of poverty, gender is another aspect of individual differences revealing variations in measurement performance. In the past, large-scale mathematics assessments showed a male advantage in measurement (e.g., Ansell & Doerr, 2000; Mullis et al., 2004; Mullis, Martin, Fierros, Goldberg, & Stemler, 2000), and an extensive body of research has demonstrated that boys excel compared with girls on some types of spatial skills (Halpern et al., 2007; Voyer, Voyer, & Bryden, 1995). However, the most recent findings of large-scale math assessments show that gender differences in mathematics have been decreasing during the last couple of years (NCES, 2009). On the latest TIMSS, the gender differences for overall math performance in the United States showed an effect size of only 0.1 (compared with an effect size of 1.5 for poverty-level differences). Further, the measurement and geometry items showed no gender differences for the United States (NCES, 2009). Thus, on the basis of the literature, there are no clear predictions related to gender.

**Overview of the Research Questions**

In summary, the present study was designed to examine the relations between key cognitive skills and measurement performance, as well as to determine whether community socioeconomics and gender moderated these relations. First, we asked whether numerical and spatial reasoning skills contribute to performance on the two subtests. One possibility, based on the conceptualization of measurement performance as involving the integration of numerical and spatial thinking, is that both numeric and spatial reasoning skills predict measurement performance across both subtypes of items (even when controlling for working memory skills). Alternatively, we might expect that numeric skills would predict both, whereas spatial reasoning skills would predict performance only on the spatial–conceptual measurement subtest. Because students can bypass spatial reasoning and still successfully solve formula-based problems, it is not clear whether spatial thinking will predict formula-based measurement. These findings may depend on individual differences, as students with good measurement sense may be able to integrate spatial and numerical thinking, whereas students who do not have a good conceptual grasp of measurement may simply apply numeric solutions to formula-based problems.

Second, the present study was designed to address the question of whether community socioeconomics and gender moderate the association between these predictor variables and types of measurement performance. We examined whether associations varied across boys and girls, but given the mixed results from previous work on gender differences in measurement, we did not formulate specific hypotheses relating to gender. With regard to community socioeconomics, however, we formulated a moderation hypothesis for one of the cognitive predictors: spatial reasoning. Namely, we hypothesized that the relation between spatial reasoning skills and measurement would be stronger for students from the affluent community than for students from the low-income community, who may have fewer naturally occurring opportunities in their environment to apply spatial skills toward academic performance. On the other hand, it is less clear whether type of community should moderate the relation of numeric skills to measurement.
performance. Given that children from both communities likely had sufficient opportunities to apply numeric skills toward academic performance (as a result of ubiquitous emphasis on numeric skills in school), more skilled students in each group may be able to effectively apply numeric skills when solving all types of measurement problems (relative to their own income group).

Note that in examining the effects of poverty, we consider the developmental impact of the broad social ecology that occurs within the context of a low-income community when compared with the broad social ecology that occurs within a more affluent community. Although we distinguish between these two communities as either low income or affluent, all aspects of these contexts such as their racial and ethnic compositions, the education levels of parents and community members, school quality, and families’ incomes are considered relevant. That is, rather than attempt to statistically disentangle the unique effects of each component of these ecologies, we consider these communities broadly, aiming to capture the correlated, additive, and interactive risks and protective factors occurring in these different types of communities.

Method

Participants

The sample included 124 fourth-grade students who participated in 3 days of testing; 54 (23 girls, 31 boys) were from schools in an affluent community, and 70 (37 girls, 33 boys) were from schools in a low-income community. Students were recruited in Massachusetts from three public schools in a low-income urban community and from two public schools in a neighboring suburban community. The student population in the affluent community was, on average, 8% African American, 61% White, 9% Hispanic, 18% Asian, and 4% other. In the low-income community, the student population was 39% African American, 13% White, 37% Hispanic, 8% Asian, and 3% other (Massachusetts Department of Education, 2008).

Economic indicators were assessed at community and school levels. The low-income schools served a community with a median family income of $44,151, which was approximately 12% lower than the national median income and 28% lower than that of the state of Massachusetts based on the 2000 census (U.S. Census Bureau, 2007). At the school level, the three schools sampled had 79%, 82%, and 87% of students, respectively, who qualified for free or reduced-price lunch (Massachusetts Department of Education, 2008). High-poverty level has been defined as communities with at least 75% of students eligible for free or reduced lunch (NCES, 2009). Thus, the schools in the low-income community in the present study fit this definition. The affluent schools served a community with a median family income of $92,993, which is 85% higher than the national median income and 51% higher than that of the state of Massachusetts based on the 2000 census (U.S. Census Bureau, 2007). At the school level, the two schools sampled had 8% and 14% of students, respectively, who qualified for free or reduced-price lunch. Given the wide differences between the two communities in the proportion of children qualifying for free or reduced-price lunch, we used the different income levels in the two communities as a proxy for socioeconomic status.

Students from both communities were assessed during the course of one academic year, with testing beginning in November and ending in February. All students participating in the study from both low-income and affluent communities received regular instruction in mathematics based on the same curriculum: Investigations in Number, Data, and Space (Scott Foresman, 2004). Most students had been taught math with this curriculum since kindergarten. Measurement content was integrated into a variety of math content throughout the academic year. The curriculum approaches measurement skills from a spatial perspective, which is a different approach from those used in most elementary math curricula. This approach involves a focus on the structure of units within arrays, rather than on the application of formulas. For example, when learning about area, students are shown pictures of different size rectangles divided into grids and asked to identify the amount of two-dimensional space a given shape covers as its area. The content covered during third and fourth grade included measuring with paces and steps (estimation of distance and the inverse rule), standard and metric measurement, distance in miles including estimation and map scaling, comparing heights using different measuring tools, flips and turns (mental rotation), measuring perimeter, measuring area (by combining shapes and using square units), and measurement formulas. The use of the same math curriculum enabled us to partly control for the type of math content that the students were receiving within the two economic communities. It should be noted, however, that even with the same curriculum, there could be variations in how the curriculum was implemented across the two school systems.

Procedure

The study included six test instruments: a measurement test, a numeric test, a spatial visualization task, a mental rotation task, a verbal working memory task, and a spatial working memory task. During the first testing session, the measurement test, which lasted 45–50 min, was administered to the whole class. Each student received a booklet with test items. The numeric test and the two spatial reasoning tasks were administered to the whole class about a week later; together these tests took about 40 min. The two working memory tasks were administered on a third day when students were assessed individually for about 15–20 min.

Materials

Measurement test. A recently developed measurement assessment tool was used in the present study (see Vasilevya, Ludlow, et al., 2009, for detailed information on the design and psychometric properties of this test). The 34-item test had a paper-and-pencil, multiple-choice format (similar to that of NAEP tests). This format made it possible to test children’s understanding under conditions in which they cannot rely on physical manipulation of measurement instruments. Thus, rather than ask children to use measurement tools and manipulatives, the test items assessed the ability to reason about measurement concepts. Children could write on scratch paper when solving the problems. However, to prevent them from bypassing mental imagery and manipulation, they were not allowed to draw either in the test booklet or on the scratch paper (e.g., make grids, mark off units).

The students were asked to read each problem silently as the test administrator read it aloud (in order to reduce the effects of reading
level on performance). Prior to solving test items, children did a sample problem. The test items varied along several dimensions. First, they covered different content areas involving measurement of space, including object length, distance, perimeter, area, and volume. Second, the items covered different difficulty levels that were determined on the basis of pilot results. Finally, the test was specifically designed to tap two subtypes of measurement performance: formula-based measurement and spatial–conceptual measurement.

The factor analysis (reported in Vasilyeva, Ludlow, et al., 2009) supported the distinction between formula-based and spatial–conceptual categories. When the data were submitted to a two-factor principal-component solution, the items loaded on the two factors in accordance with the expected distinction. One factor included the problems that could be solved analytically with known formulas, and the other factor included the problems that involved spatial–conceptual reasoning.

These two categories were then used to classify the items into two subtests. The 14 items in the formula-based subtest involved the application of measurement formulas taught at the fourth-grade level including perimeter and area, as well as conversion formulas for linear units. Whereas some of these problems involved using formulas in a straightforward way (e.g., calculating the perimeter of the rectangle given its length and width), others required manipulating formulas (e.g., calculating the length of a rectangle given its perimeter and width). Figures 1A and 1B provide examples of items from this subtest; the items varied in content area, level of difficulty, and presence or absence of a picture.

The 20 items in the spatial–conceptual subtest involved reasoning about spatial relations, in particular the relations between the unit and the object to be measured. Figures 1C and 1D provide examples of items from this subtest. In some problems, children had to estimate the linear size or distance either by mentally manipulating the unit of measure that was provided or by generating an image of a unit and making mental comparisons to the size of other objects (Figure 1C). In other problems, children had to reason about the relation between area and square units of measurement. Figure 1D shows an area problem testing children’s understanding of the inverse relation between the size of a unit and the number of units. There were also volume items, which did not require knowledge of formulas but rather required visualizing objects in three dimensions and reasoning about the relation between the size of the object and the cubic unit. On some problems, children had to mentally generate invisible parts of objects.

The score for each subtype of items was the proportion of items within that subtype that were answered correctly. To determine the internal consistency for the two categories of measurement items for the students in the present sample, we assessed the Cronbach’s alpha for the two-factor solution. The spatial–conceptual factor had a Cronbach’s alpha of .84, and the formula-based factor had a Cronbach’s alpha of .80.

Figure 1. Sample items from the measurement test: formula-based area item (A), formula-based perimeter item (B), spatial–conceptual linear item (C), and spatial–conceptual area item (D).
**Numeric test.** The procedure for the numeric test was identical to the procedure for the measurement test, with the exception that students were encouraged to write their calculations directly in the test booklet. The items were selected from a larger pool of computational problems based on the NAEP and TIMSS fourth-grade mathematics assessments (Mullis et al., 2004; NCES, 2004). On the basis of piloting conducted with an ethnically and economically diverse group of students, we selected 14 items that varied in difficulty and showed good discriminability among individual students as well as good internal consistency, with a Cronbach’s alpha of .77 with the present sample of students.

The items included a mix of number facts and word problems. Word problems were math problems embedded in real-world situations (e.g., “There are 53 hamburgers to serve 38 children. If each child is to have at least one hamburger, at most how many of the children can have more than one?”). Number fact problems were number sentences with “How much is” in front of them (e.g., “How much is 42 – 29?”). These problems involved arithmetic operations parallel to those involved in the word problems. We included both number fact and word problems because they are typically used at the elementary level to test students’ numeric skills. Our pilot work had shown that there were no differences in the level of difficulty between the two types of numeric problems, for participants overall and for each gender group.

**Spatial reasoning composite measure.** To tap a range of spatial skills as a predictor variable, we used a spatial reasoning composite measure that incorporated two key types of spatial tasks (Linn & Petersen, 1985; Voyer et al., 1995), which have been shown to be related to mathematics performance (Casey et al., 1995; Sherman, 1979): a spatial visualization task and a mental rotation task ($r = .37$). Scores on both these tasks were normalized and averaged to form the spatial reasoning composite measure. To address the reliability of the two spatial measures, we assessed their internal consistency. We found that the spatial imagery puzzle task had a Cronbach’s alpha of .81; the mental rotation test had a Cronbach’s alpha of .70.

The spatial visualization task was adapted for children from the Jigsaw-Puzzle Imagery Task developed by Richardson and Vecchi (2002) for adults. The pictures of objects used in the Jigsaw-Puzzle Imagery task were selected from objects rated for level of familiarity by Snodgrass and Vanderwart (1980). We chose this task because it was designed to incorporate mental visualization, spatial working memory, and a strong mental movement component—important elements of many measurement problems. Unlike with a standard puzzle, participants were not allowed to physically move the puzzle pieces; instead they had to mentally move the pieces to different locations on an empty grid to find where each piece should be placed to create the object. Thus, participants had to generate and hold images of the pieces of the object, imagine how the elements fit together, and then mentally move the pieces to the correct location.

The puzzles used in the present study were pilot-tested with an ethnically and economically diverse group of fourth graders to remove items that were either too easy or too difficult. To adapt the task for fourth graders, only puzzles with four or six pieces were included, some of the pictures used by Richardson and Vecchi (2002) were replaced by pictures more familiar to children, and the task was made up of fewer items.

The final task consisted of eight puzzle items taken from the pilot-tested items. For each puzzle, on the top page participants were given a drawing of a common object that had been cut into equal-sized rectangular puzzle pieces and placed on a grid in a mixed-up randomized order (see Figure 2A). Each piece was numbered based on this random order from left to right and top to bottom across the grid. On the bottom page there was a blank grid of the same size as the top grid (see Figure 2B). The goal was to write the number of the puzzle piece from the top page in the correct grid location on the bottom page so that if the pieces were placed in those locations, they would make the complete picture of the object. The number of correct responses consisted of the number of picture pieces placed in the correct location, and the score was the average proportion of puzzle pieces placed in the correct location across all eight items.

Students were given two sample problems. For each puzzle, students were first told the name of the object that the puzzle pieces would make when combined and then told they could write numbers in the grid locations but could not draw any pictures or make any other marks on their booklet. They were allowed a...
maximum of 2 min to solve each puzzle. The task consisted of two
sets of puzzles (nonrotated and rotated puzzles). Within each set,
the first puzzle was a four-piece puzzle placed on a $2 \times 2$ grid, and
the last three were six-piece puzzles on a $3 \times 2$ grid. For the four
puzzles in the rotated set, half the pieces in the top puzzle display
were randomly selected to be rotated $180^\circ$ from their original
orientation (see Figure 2A).

The mental rotation test was related to a previous measure used
with seventh-grade students (Anderson et al., 2008). This task was
based on figures designed by Shepard and Metzler (1971). All
drawings represented figures rotated in different orientations
around an axis in three-dimensional space (see Figure 3). Prior to
conducting the present study, the mental rotation task was pilot-
tested with an ethnically and economically diverse group of stu-
dents.

Each problem consisted of a row of three figures; the standard
figure (marked by a star) was presented on the left side, and a pair
of choice figures was presented on the right (see Figure 3). One of
the two choice figures was “the same as” (identical to) the standard
figure, whereas the other was “different from” (a mirror image of)
the standard figure. The students’ task was to choose which of the
two choice figures matched the figure with the star. They were
told:

If you turn this figure around in your head, it will match the figure
marked with the star. Now, the other figure is not the same as the one
with the star over it. No matter how you turn it around in your head,
it won’t go the same way. Let me show you what I mean.

As an example, the tester presented two identical three-
dimensional, multilink wooden cube figures, then placed the fig-
ures one on top of the other and rotated them so the students could
see that the figures could be made to be identical. Next, the tester
placed two mirror-image figures one on top of the other. The figures
were then rotated so the students could see that the figures
could not be made to be identical. Next, the students were referred
to the sample figures in their booklet and asked to try to decide
which of the two choice figures was the same as the one marked
with a star. After the sample item, students were presented with 12
items.

The working memory tasks. The verbal and spatial working
memory tasks were administered to participants individually in a
single session outside the classroom. Both tasks were presented
on a laptop computer and began with a practice trial, which was
immediately followed by test trials. Test trials were presented as a
series of blocks, each with six trials. Each consecutive block would
add one to the list length of the previous block. For all tasks, one
point was given for each correctly recalled trial. All tests followed
a move-on rule and a discontinue rule. Under the move-on rule, if
a child correctly responded to the first four trials within a block,
the program automatically proceeded to the next block, and a score
of 6 was given for the block just completed. If a child responded
correctly to four of five trials within a block, the program auto-
matically began the next block, and a score of 5 was given for the
block just completed. Under the discontinue rule, if a child made
three or more errors within a block, the program automatically
stopped. Therefore, the total score was the total number of trials
answered correctly until the program automatically stopped due to
the discontinue rule.

The verbal working memory task consisted of a subtest from the
Automated Working Memory Assessment (AWMA) created by
Alloway (2007). As presented in the AWMA, the backwards digit
recall task (a modified version of the verbal short-term memory
digit recall task) was used to assess verbal working memory. In
this task, children heard a sequence of digits before they were
asked to recall aloud the digit sequence in reverse order. The digit
sequence, which varied from one item to the next, increased in
length by one digit for each consecutive block of items, beginning
with two digits in the first block. A response was considered
correct when the child repeated back the entire digit sequence in
reverse order. The memory maintenance demand of this task was
the retention of the order of the digits, and the processing demand
was the reversal of the order of the digits retained in memory.

The spatial working memory task was a new task created by
systematically modifying the task used to assess spatial short-term
memory on the AWMA (the block recall task). In this spatial
working memory task, students were asked to recall a sequence of
blocks but in reverse order. First, children viewed a video clip
from the AWMA in which the tester on the screen tapped an
arrangement of nine blocks in a particular order. The blocks
remained on the computer screen, and the children were asked to
touch the blocks on the screen in the reverse order of the pattern
that had just been shown in the video. The block tapping sequence,
which varied from one item to the next, increased in length by one
block for each consecutive set of items, beginning with two blocks
in the first set. A response was considered correct if the child
touched every block from the tapping sequence in reverse order.
The memory maintenance demand of this task was the retention
of the order of the blocks that were presented, and the processing
demand was the reversal of the order of the blocks retained in
memory.

Analytic Approach

To examine the study hypotheses, we estimated three ordinary
least squares regression models—building hierarchically—for
each of our outcomes (spatial–conceptual and formula-based mea-
surement). Specifically, in the first model, we regressed the out-
comes on working memory covariates. In the second model, we
added four main-effect predictors (i.e., numerical skills, spatial
reasoning skills, gender, and income level). This approach allowed
us to examine variance in measurement explained by working
memory before examining the predictive value of the main effects
above and beyond the variance explained by these covariates.
In our third and final model, we added five interaction terms (i.e., gender by income level, numerical by gender, numerical by income, spatial reasoning by gender, and spatial reasoning by income), therefore, simultaneously estimating main effects and interactions. This model follows established methods for examining moderation hypotheses (Baron & Kenny, 1986; Cohen, Cohen, West, & Aiken, 2003; Dearing & Hamilton, 2006; Hamilton, 1992; McClelland & Judd, 1993; Warner, 2008). Specifically, we used the test statistic (t test in ordinary least squares) corresponding to the coefficients for the interaction terms to evaluate whether the moderator effects were statistically significant. In addition, we probed statistically significant interactions by estimating simple slopes and conditional effects (Cohen et al., 2003; Dearing & Hamilton, 2006; Preacher, Curran, & Bauer, 2006).

**Results**

**Preliminary Analyses**

Prior to our primary analyses, we examined the study variables for violations of normality and outliers. There were no univariate outliers (e.g., values more than 2.5 standard deviations above or below the mean), and tests for skewness among the predictors and outcomes were null, with one exception: Scores from the numeric test were moderately negatively skewed. However, the results of our primary analyses proved robust to using a transformed (i.e., reflected and squared) version of this predictor that corrected the skewness. As such, we present only the results with untransformed predictors and outcomes.

Descriptive statistics for the study variables are presented in Tables 1 and 2, broken down by community (i.e., affluent and low income) and gender. In addition, zero-order correlations are presented in Table 3 for the sample, as a whole, and in Table 4 by community and gender. In these tables, three findings are of particular interest. First, in Table 1, for the sample as a whole and for students from the affluent community, gender differences were small and generally not statistically significant. For students from the low-income community, however, two moderate-sized gender differences were evident: Girls showed higher scores on numerical skills and formula-based measurement than boys. Second, as shown in Table 2, it is evident that students from the affluent community scored higher on all indicators, including the measurement outcomes, the skill predictors, and the working memory covariates. Third, for the correlations, although both skill predictors were strongly associated with the two measurement outcomes in the sample as a whole, associations between spatial skills and measurement were more than twice as large for students in the affluent community than for those in the low-income community (i.e., .53 vs. .21 for formula based and .65 vs. .28 for spatial–conceptual; see Table 4).

**Main Analysis**

To test our study hypotheses, we estimated ordinary least squares regression models separately for each measurement outcome. Models were estimated in three analytic steps. In the first step, we regressed the outcomes on working memory covariates. We present the findings of both outcomes together in Table 5, including coefficients and standard errors. Note that verbal working memory was significantly and positively associated with spatial–conceptual measurement, and spatial working memory was significantly and positively associated with both outcomes. These significant associations underscore the value of controlling for children’s working memory while estimating the main effects and interactions of numerical skills, spatial reasoning skills, gender, and community.

In our second analytic step, we added four main effect predictors (i.e., numerical skills, spatial reasoning skills, gender, and community) to the regression models while controlling for working memory. In Table 5 we provide an overview of the results from this second analytic step for each of the two measurement outcomes. At least three results are noteworthy. First, associations between working memory and measurement were close to zero and insignificant; the predictors entered in the second step accounted for the variance previously explained by these covariates.

Second, controlling for working memory, both spatial reasoning and numerical skills were positively and significantly associated with the two measurement outcomes; better spatial reasoning and numerical skills predicted better measurement performance. It had been hypothesized that spatial skills would predict spatial–conceptual measurement. In addition, the results showed that this type of skill also predicted formula-based measurement. The significant relation between numeric skills and the two outcome measures was consistent with hypotheses.

Beyond statistical significance, the effect sizes for spatial reasoning and numerical skills were relatively large. For the spatial reasoning measure, effect sizes ranged from 0.60 (for formula based) to 0.74 (for spatial–conceptual). For numerical skills, effect sizes ranged from 0.89 (for formula based) to 0.95 (for spatial–conceptual). Together these two skills explained as much as 30% of the variance in the measurement outcomes (e.g., spatial reasoning uniquely explained 12.0%, and numerical skills uniquely explained 18.4% of the variance in spatial–conceptual measurement).

The third noteworthy result from this main-effects model was that gender and community were significantly associated with the two measurement subtests. Boys outperformed girls on spatial–conceptual problems ($d = 0.44$), and girls outperformed boys on formula-based problems ($d = -0.36$). Thus, boys and girls displayed contrasting strengths in the two measurement areas (with the magnitude of these differences being fairly similar in size). Also as expected, students from the affluent community scored higher than those from the low-income community on the two measurement outcomes, with the smaller differences evident for formula-based problems ($d = 0.42$) and the larger differences evident for spatial–conceptual problems ($d = 1.23$).

As our third and final analytic step, we added five interaction terms to our regression models: gender by community, numerical

---

1 In the zero-order correlations, the association between visual working memory and formula-based measurement was significantly stronger for the low-income group than for the affluent group. However, this violation was not evident in the multivariate regression models. Nonetheless, we reestimated our models with four additional interaction terms allowing both working memory covariates to vary by gender and income level. None of these additional interaction terms was statistically significant, and all our reported significant results remained significant in these reestimated models.
by gender, numerical by community, spatial reasoning by gender, and spatial reasoning by community. These interactions allowed the estimated effects of gender to vary by community and the estimated effects of spatial reasoning measure and numerical skills to vary by gender and community. An overview of the results from this step is presented in Table 5. The spatial reasoning measure and numerical skills predictors were mean centered, and simple slopes (i.e., conditional effects) were computed for all interactions, following best-practice recommendations (for a review, see Dearing & Hamilton, 2006).

In this last analytic step, numeric skills did not interact with either gender or type of community for either of the two outcomes; the main effect of numeric skills on measurement remained significant independent of gender and community. In contrast, the association between spatial reasoning skills and measurement significantly varied across type of community for both formula-based and spatial–conceptual measurement outcomes. These statistically significant interactions between spatial reasoning and the two outcome measures are graphed in Figures 4 and 5, with predicted outcome scores plotted at 1 standard deviation above and below the mean for the spatial reasoning composite.

As predicted, the effect of spatial reasoning on measurement performance was greater for students from the affluent than from the low-income community. For students from the affluent community, spatial reasoning skills were statistically significant and positive predictors of both measurement outcomes, with effect sizes ranging from 0.69 for formula-based measurement (simple slope of $b = 0.13, SE = 0.04, p < .001$) to 0.83 for spatial–conceptual measurement (simple slope of $b = 0.11, SE = 0.03, p < .001$). However, for students from the low-income community, this association was much smaller in size for both outcomes (effect sizes ranging from 0.16 to 0.33) and never reached statistical significance.2

The Gender $\times$ Community interaction for spatial–conceptual measurement was also significant. For students from the affluent community, there was no gender difference on this outcome ($b = -0.02, SE = 0.04, p = .64$). In contrast, for the students from the low-income community, there was a statistically significant difference such that boys outperformed girls ($b = 0.12, SE = 0.04$).

---

### Table 1

**Mean Proportion of Items Correct (Unadjusted) for the Predictor and Dependent Variables by Gender and Gender by Income**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Female</th>
<th>Male</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$N$</td>
<td>$M$</td>
<td>$SD$</td>
<td>$N$</td>
<td>$M$</td>
<td>$SD$</td>
<td>$F$</td>
<td>$d$</td>
<td></td>
</tr>
<tr>
<td><strong>Predictor variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verbal working memory</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low income</td>
<td>37</td>
<td>11.84</td>
<td>2.78</td>
<td>33</td>
<td>11.12</td>
<td>3.22</td>
<td>0.61</td>
<td>-0.24</td>
<td></td>
</tr>
<tr>
<td>Affluent</td>
<td>23</td>
<td>13.74</td>
<td>5.65</td>
<td>31</td>
<td>14.29</td>
<td>3.86</td>
<td>0.27</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>Total sample</td>
<td>60</td>
<td>12.57</td>
<td>4.18</td>
<td>64</td>
<td>12.66</td>
<td>3.86</td>
<td>0.01</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>Spatial working memory</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low income</td>
<td>37</td>
<td>15.19</td>
<td>3.85</td>
<td>33</td>
<td>15.12</td>
<td>4.23</td>
<td>0.00</td>
<td>-0.02</td>
<td></td>
</tr>
<tr>
<td>Affluent</td>
<td>23</td>
<td>19.39</td>
<td>5.48</td>
<td>31</td>
<td>20.77</td>
<td>4.56</td>
<td>1.27</td>
<td>0.27</td>
<td></td>
</tr>
<tr>
<td>Total sample</td>
<td>60</td>
<td>16.80</td>
<td>4.95</td>
<td>64</td>
<td>17.86</td>
<td>5.20</td>
<td>0.65</td>
<td>0.21</td>
<td></td>
</tr>
<tr>
<td>Numeric skills</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low income</td>
<td>37</td>
<td>0.60</td>
<td>0.17</td>
<td>33</td>
<td>0.44</td>
<td>0.22</td>
<td>12.94***</td>
<td>-0.81</td>
<td></td>
</tr>
<tr>
<td>Affluent</td>
<td>23</td>
<td>0.77</td>
<td>0.14</td>
<td>31</td>
<td>0.74</td>
<td>0.21</td>
<td>0.44</td>
<td>-0.17</td>
<td></td>
</tr>
<tr>
<td>Total sample</td>
<td>60</td>
<td>0.66</td>
<td>0.18</td>
<td>64</td>
<td>0.58</td>
<td>0.26</td>
<td>8.20**</td>
<td>-0.36</td>
<td></td>
</tr>
<tr>
<td>Spatial reasoning composite</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low income</td>
<td>37</td>
<td>-0.21</td>
<td>0.83</td>
<td>33</td>
<td>-0.21</td>
<td>0.85</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Affluent</td>
<td>23</td>
<td>0.12</td>
<td>0.70</td>
<td>31</td>
<td>0.43</td>
<td>0.87</td>
<td>1.83</td>
<td>0.39</td>
<td></td>
</tr>
<tr>
<td>Total sample</td>
<td>60</td>
<td>-0.08</td>
<td>0.79</td>
<td>64</td>
<td>0.10</td>
<td>0.91</td>
<td>1.06</td>
<td>0.21</td>
<td></td>
</tr>
<tr>
<td><strong>Dependent variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Formula-based measurement</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low income</td>
<td>37</td>
<td>0.46</td>
<td>0.19</td>
<td>33</td>
<td>0.33</td>
<td>0.18</td>
<td>6.34**</td>
<td>-0.70</td>
<td></td>
</tr>
<tr>
<td>Affluent</td>
<td>23</td>
<td>0.69</td>
<td>0.24</td>
<td>31</td>
<td>0.62</td>
<td>0.23</td>
<td>1.26</td>
<td>-0.30</td>
<td></td>
</tr>
<tr>
<td>Total sample</td>
<td>60</td>
<td>0.55</td>
<td>0.24</td>
<td>64</td>
<td>0.47</td>
<td>0.26</td>
<td>6.24**</td>
<td>-0.32</td>
<td></td>
</tr>
<tr>
<td>Spatial–conceptual measurement</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low income</td>
<td>37</td>
<td>0.39</td>
<td>0.16</td>
<td>33</td>
<td>0.45</td>
<td>0.17</td>
<td>2.38</td>
<td>0.36</td>
<td></td>
</tr>
<tr>
<td>Affluent</td>
<td>23</td>
<td>0.75</td>
<td>0.16</td>
<td>31</td>
<td>0.75</td>
<td>0.16</td>
<td>0.02</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Total sample</td>
<td>60</td>
<td>0.53</td>
<td>0.24</td>
<td>64</td>
<td>0.59</td>
<td>0.22</td>
<td>0.83</td>
<td>0.26</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* Tests for the homogeneity of variances were conducted for the three measurement tests by means of the traditional $F$ test and Levene’s test with estimators of central tendency robust to nonnormality (Brown & Forsythe, 1974); the results of these tests were all null ($p > .05$). Positive $d$ statistics indicate a male advantage.

$^*$ $p < .05$. $^*$*$p < .01$. $^*$*$*p < .001$. 

2 One possible explanation for the present results might be that the patterns of relations obtained in the regressions are due to individual differences in performance level rather than community affluence. To test for this, we divided the total sample into high and low groups for each dependent variable using the median split and included performance level in the regression analyses as well as type of community. When we examined the interaction effects of performance level with numeric and spatial skills in addition to the interaction of these variables with community income level, we found that only community poverty interacted significantly with the predictor variables.
p = .001); the effect size for this difference was 0.60, a larger difference than was indicated by the main effect of gender when averaged across type of community. It is worth noting that this gender difference for the students in the low-income community was not apparent in the zero-order correlations, underscoring the impact of controlling for the other predictors in the regression model (e.g., working memory and numerical skills).

Discussion

In the present study, we examined whether numerical and spatial reasoning skills contribute to performance on two types of measurement problems: those that tap spatial–conceptual reasoning and those that focus on formula-based solutions. A key purpose of our research was to investigate these relationships from the perspective of individual differences. The goal was to better understand how the pattern of relations between the two types of reasoning skills and the two measurement subtypes might differ for children from low-income and affluent communities and for boys and girls.

Before discussing our findings in detail, it is important to point out some of the limitations of the study. In particular, we note that our comparison of two communities based only on poverty rate did not allow us to disentangle the unique contributions of factors, such as family income, education, and ethnicity. In short, these results should be interpreted with consideration for potential omitted variable bias. Nonetheless, our macrofocus on these two social ecologies is an important first step in understanding the developmental context that may hamper children’s applications of cognitive skills.

More generally, we recognize that we cannot make causal conclusions based on our correlational design. Given this methodological limitation, we consider alternative explanations throughout the Discussion. Although the majority of work on income level and children’s growth is necessarily nonexperimental, future work that is able to include a wider range of covariates and a larger number of communities will be helpful in this regard. Furthermore, the findings need to be replicated in districts that use a more traditional (formula-based) method of measurement instruction. Also, experimental intervention studies could directly assess the contribution of numeric and spatial skills to measurement performance.

Differences in Measurement Performance Levels Between the Two Communities

As in previous studies (NCES, 2004, 2009), in terms of absolute levels of performance, there were substantial differences between the two communities in the average accuracy of measurement performance.
performance. It is discouraging that the students from the low-income community did substantially worse on the measurement test than the students from the affluent community even though they were taught with the same spatially based math curriculum.

It was important that we did control for the math curriculum used in the two communities. Nevertheless, there are many other factors that could have contributed to this difference above and beyond the particular math curriculum used. First, despite extensive training in both school systems, the teachers in the affluent community may have been more effective at using this curriculum than the teachers in the low-income community (e.g., Pianta, La Paro, Payne, Cox, & Bradley, 2002). Second, classroom context may have contributed to the differences; for example, the students may have differed in terms of the stimulation and input provided by their peers in the classroom, resulting from the differing ability levels and knowledge base of their classmates (e.g., Battistich, Solomon, Kim, Watson, & Schaps, 1995).

More generally, there are tremendous differences, on average, in the levels of cognitive stimulation that children in low-income versus affluent families receive in their home environments (e.g., Dearing & Taylor, 2007; Votruba-Drzal, 2003). In terms of level of spatial skills, children from low-income communities may not have been sufficiently exposed to spatial experiences and resources in the home to develop good spatial reasoning skills on their own (e.g., Bradley, Corwyn, McAaddock, & García Coll, 2001). This lack of exposure to spatial stimulation includes a deprived “spatial world” outside the home as well, with poor neighborhoods often lacking out-of-school activity resources, such as community centers or sports teams, and parks and playgrounds, as well as possibly exposing children to hazardous conditions in their environment (Evans, 2004; Larson, 2000; Leventhal & Brooks-Gunn, 2000).

Application of Cognitive Skills to Academic Performance

Our findings are consistent with previous research that has demonstrated the ways by which socioeconomic conditions can affect children’s opportunities to develop cognitive skills (Entwisle et al., 1994; Heckman, 2008). However, our study provides a critical extension to this literature by suggesting that socioeconomic context may also affect children’s capacities to use cognitive skills in the service of academic performance. We found different patterns of relations between predictor variables and measurement for the two communities studied; children growing up in low-income communities were less likely than children in affluent communities to benefit—in terms of their measurement performance—from good spatial skills. Whereas in the affluent community numerical and spatial skills were strongly related to both types of measurement, in the low-income community only numerical, and not spatial reasoning skills, appeared to benefit students’ measurement performance.

Theorists from developmental psychology (e.g., Ceci & Papierno, 2005) and economics (e.g., Heckman, 2008) have, in fact, pointed out that limited learning opportunities in the home and neighborhood may limit both the acquisition and the application of cognitive skills for children growing up poor. This phenomenon can be understood within the context of the developmental literature on situated cognition (Nunes & Bryant, 1996; Rogoff, 1990) in which children may develop a relatively high level of a particular cognitive skill but apply it only within the specific context in which the skills were acquired. For example, in some countries poor children use high-level computational skills when selling goods on the street but are unable to transfer those skills to solving school math problems (Saxe, 2002; Saxe et al., 1996).

In families of poverty there is, on average, a lack of enriching experiences with school-related spatial problem-solving activities (such as paper-and-pencil games, mazes and connect-the-dots, computer-related games, constructions sets and legos, and workbooks with geometry and measurement problems). The lack of opportunities for these types of spatial learning experiences, as well as for other activities in the home and neighborhood, may impede the transition to school mathematics for children in poverty, even for those who have innate above-average spatial skills. In other words, limited opportunities to apply spatial reasoning to school-related learning contexts may limit measurement performance, even for children from low-income communities with good basic spatial skills.

Type of Community as a Moderator of Measurement Performance

Pattern of results for students from the affluent community. Within the group of students from the affluent community, those who have an understanding of both numerical principles and spatial reasoning appear to be able to draw on these skills when solving measurement problems across the board. One way of explaining these findings is that these students may be combining
### Table 5
Regression Models Predicting Measurement Performance

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Step 1</th>
<th>Step 2</th>
<th>Step 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Formula based</td>
<td>Spatial–conceptual</td>
<td>Formula based</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>b</td>
<td>SE</td>
</tr>
<tr>
<td>Working memory</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verbal</td>
<td>0.16</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Spatial</td>
<td>0.40***</td>
<td>0.02</td>
<td>0.004</td>
</tr>
<tr>
<td>Cognitive skills</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Numerical</td>
<td>0.39***</td>
<td>0.35</td>
<td>0.07</td>
</tr>
<tr>
<td>Spatial reasoning</td>
<td>0.23**</td>
<td>0.23**</td>
<td>0.06</td>
</tr>
<tr>
<td>Demographics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>-0.13*</td>
<td>-0.07</td>
<td>0.03</td>
</tr>
<tr>
<td>Affluent community</td>
<td>0.19*</td>
<td>0.10</td>
<td>0.04</td>
</tr>
<tr>
<td>Gender × Community</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Numerical × Gender</td>
<td>-0.02</td>
<td>-0.03</td>
<td>0.19</td>
</tr>
<tr>
<td>Numerical × Community</td>
<td>-0.00</td>
<td>-0.01</td>
<td>0.19</td>
</tr>
<tr>
<td>Spatial Reasoning × Gender</td>
<td>0.01</td>
<td>0.01</td>
<td>0.04</td>
</tr>
<tr>
<td>Spatial Reasoning × Community</td>
<td>0.24*</td>
<td>0.11</td>
<td>0.04</td>
</tr>
</tbody>
</table>

$R^2$ | .22 | .23 | .53 | .69 | .56 | .72 |

* $p < .05$.  ** $p < .01$.  *** $p < .001$. 

---

---
an understanding of the underlying spatial nature of measurement with the numerical and procedural competence to use measuring tools and apply formulas. In other words, they may have developed good measurement sense, which means that they may have a better conceptual understanding of the processes underlying measurement procedures (Joram, 2003; Shaw & Pucket-Claiatt, 1989).

It is interesting to note that among the students from affluent communities, spatial reasoning was related to their performance even on formula-based problems. Although many formula-based problems can be solved with only numerical skills, there are a number of ways that applying spatial reasoning can be helpful in solving these problems (Battista, 2003; Campbell et al., 1992; Outhred & McPhail, 2000). For example, when students are asked to make conversions (e.g., finding out the height of a fence in centimeters when given the height in meters), an ability to visualize the size of a centimeter and a meter may be helpful in estimating the most likely answers, even before doing the calculations. Again, when students have to compute the width of a rectangle based on the length and the area, their ability to visualize the rectangle would likely be helpful in eliminating possible choices even before they do the calculations. Thus, those students who had an understanding of the spatial reasoning related to measurement may have had an edge over their classmates even on formula-based problems.

Given that our data are only correlational, it should be noted that there are other, alternative explanations of the findings for the students from the affluent community. For example, it may be that among the students from the affluent community, the brighter students simply did better across all tests as a function of high general intelligence, and these same students would be likely to outperform the other students on other tests that tap into general intelligence. Yet, this alternative explanation seems less likely; when we included measures of verbal and spatial working memory in the analysis, we found that the obtained relations of measurement to spatial and numeric skills occurred even when controlling for these basic processes. Moreover, the skill predictors were consistently more strongly associated with measurement performance than were the working memory predictors, an indication that these skills may be more proximally related to measurement performance than working memory mechanisms. This is important because working memory is a basic cognitive process that has been found to be closely associated with the g factor of general intelligence, with some researchers arguing that working memory may in fact be a proxy for IQ (Colom, Rebollo, Palacios, Juan-Espinosa, & Kyllonen, 2004; Kyllonen & Christal, 1990; Sūβ, Oberauer, Wittmann, Wilhelm, & Schultz, 2002). Nonetheless, additional research is needed to determine whether similar relations occur when controlling specifically for child general intelligence.

**Pattern of results for the students from the low-income community.** Similar results obtained in the affluent community, the performance of students from the low-income community showed a relation between numeric skills and measurement for both outcomes. Thus, their degree of skill at doing computations was significantly related to their ability to solve different types of measurement problems. This is perhaps not surprising due to the focus on arithmetic skill development throughout math instruction in elementary schools across communities of varying income levels (Clements & Battista, 1992). The relation between numeric skills and measurement was found even when working memory measures were included as covariates in our regression models.

Unlike students growing up in the more affluent community, however, students from the low-income community did not seem to be able to apply their spatial reasoning skills when solving either the formula-based or the spatial–conceptual measurement problems. It is important to note that this does not appear to be due to a lack of variability among students from the low-income community (i.e., truncated range of test scores). Indeed we found that there was substantial variability within this group as well as the affluent community for the working memory and spatial predictor variables and for the outcome measures (see Tables 1 and 2).

Nonetheless, it still could be that patterns of relations between cognitive skills and measurement could be an artifact of differences in performance levels between the two groups (but see Footnote 2). Because the students from the low-income community, on average, did worse on the spatial tasks tests than the students from the affluent community, high and low spatial per-

![Figure 4](image-url)

*Figure 4. Slope of association between spatial reasoning composite and formula-based measurement for students from the affluent and low-income groups. Spatial reasoning composite values are graphed from 1 standard deviation below to 1 standard deviation above the mean. The statistical significance of each simple slope is noted.*

![Figure 5](image-url)

*Figure 5. Slope of association between spatial reasoning composite and spatial–conceptual measurement for students from the affluent and low-income groups. Spatial reasoning composite values are graphed from 1 standard deviation below to 1 standard deviation above the mean. The statistical significance of each simple slope is noted.*
formance levels may be the critical moderator; for example, there may be a spatial skills threshold that must be met in order for measurement performance to benefit from these skills. However, we find little evidence that our pattern of results could be explained by mean differences in performance levels. Indeed, over 25% of the low-income group performed above the affluent group’s median, and over 21% of the low-income group were in the top third of all children on spatial reasoning skills, making this threshold alternative an unlikely explanation for our results.

Thus, statistical artifacts such as lack of variability in performance or floor effects (i.e., a failure to meet some high-performance threshold) among the students from low-income communities do not appear to be critical factors in explaining why their spatial skills were not related to their measurement performance. To better understand the specific nature of their difficulties, in a recent study we identified the types of mistakes and incorrect strategies used by the children from a low-income community when approaching measurement problems (Vasilyeva, Casey, et al., 2009). We found that even when students attempted to use the spatial procedures and strategies taught to them, they were easily confused about which spatial strategy to use for which type of item. In other words, they did try to use spatial procedures but applied them incorrectly. When trying to remember which procedure should be applied to which measurement context for solving area and volume problems, they were often tricked by superficial characteristics of the problem presentation. For the types of procedures needed to solve length, area, or volume problems, they did not seem to have a conceptual understanding of the underlying spatial differences among these diverse types of measurement problems; for example, some students calculated total surface area to solve volume estimation problems.

Gender as a Moderator of Measurement Performance

When gender was entered as a predictor in the second stage of the analysis, an interesting pattern emerged on the two subtypes of measurement items: The boys performed better than the girls on the spatial–conceptual items, whereas the girls outperformed the boys on the formula-based items. These findings reinforce the view that spatial–conceptual and formula-based subtests are best analyzed as unique domains of measurement performance.

In the final phase of the analysis (when the interaction terms were entered into the equation), the findings on gender revealed a more complex picture. There was a significant Gender × Community interaction for the spatial–conceptual items. Comparisons showed no significant gender differences for students from the affluent community; it was the students from the low-income community who showed a male advantage on spatial–conceptual measurement performance. Ceci, Williams, and Barnett (2009) found that the magnitude of gender differences in math performance was greater among low-income than among middle-income students, with low-income boys doing better than low-income girls. In contrast, when assessing basic spatial skills rather than math performance, a reverse interaction pattern has been found: The male advantage on spatial tasks was present in the middle- and higher-income students but not in the low-income group (Levine, Vasilyeva, Lourenco, Newcombe, & Huttenlocher, 2005). A key difference between the Levine et al. (2005) findings on basic spatial skills and the spatially related measurement skills assessed in the present study is that measurement skills are formally taught to children in school. Thus, the boys may be benefiting more from the spatial measurement instruction than the girls. Given the inconsistency in the literature, no specific predictions were made in the present study regarding the relation between gender and income level. Thus, the present finding of gender differences in the low-income group needs to be replicated further.

One direction in future research is to examine gender differences at a more molecular level to better understand whether there are differences in the patterns of errors and strategies used by boys and girls from low-income communities. A recent qualitative analysis of low-income students’ measurement performance revealed both similarities and differences in the ways that the boys and girls solved measurement problems (Vasilyeva, Casey, et al., 2009). One gender difference was revealed in the type of representations used for recording their measurement solutions: The girls generally wrote down calculations, whereas the boys made drawings. An analysis of problem types showed that girls had particular trouble with spatial–conceptual measurement items in which there was no pictorial representation of the problem, suggesting a difficulty with generating images.

Conclusions

Although numerous researchers in the field have assumed that both numerical and spatial skills are implicated when solving measurement problems (Battista, 2003; Lehrer, 2003; Miller, 1989; Shaw & Pucket-Cliatt, 1989; Stephan & Clements, 2003; Wilson & Rowland, 1993), the present study is the first to establish these relationships. We took this question one step further, providing a more nuanced understanding of measurement performance, by breaking it down into two key subtypes: one type requiring the application of measurement formulas and the other requiring the use of spatial reasoning as a basis for understanding measurement concepts.

Our findings present an interesting pattern of relationships from the perspective of individual differences: Spatial skills do not aid the measurement performance of children growing up in a low-income community, whereas they do contribute significantly to measurement performance in children coming from an affluent community. These findings are an important contribution to the literature on poverty and children’s development, because much of this literature has been focused on the ways in which poverty limits academic skill acquisition; our results indicate that growing up in a low-income community may also limit basic skill application to academic performance. One educational implication of the present findings is that educators should consider focusing on more on interventions that help children make the transition between their basic cognitive skills and school achievement.

References


problems using geometry clues matched to three different cognitive styles. Mind, Brain, and Education, 2, 188–197.


