Socioeconomic Status and Computer Science Achievement

Spatial Ability as a Mediating Variable in a Novel Model of Understanding

Miranda C. Parker
Georgia Institute of Technology
Atlanta, Georgia
miranda.parker@gatech.edu

Amber Solomon
Georgia Institute of Technology
Atlanta, Georgia
asolomon30@gatech.edu

Brianna Pritchett
Georgia Institute of Technology
Atlanta, Georgia
bpritchett8@gatech.edu

David A. Illingworth
Georgia Institute of Technology
Atlanta, Georgia
david.illingworth@gatech.edu

Lauren E. Margulieux
Georgia State University
Atlanta, Georgia
lmargulieux@gsu.edu

Mark Guzdial
Georgia Institute of Technology
Atlanta, Georgia
guzdial@cc.gatech.edu

ABSTRACT

Socioeconomic status (SES) has a measurable impact on many educational outcomes and likely also influences computer science (CS) achievement. We present a novel model to account for the observed connections between SES and CS achievement. We examined possible mediating variables between SES and CS achievement, including spatial ability and access to computing. We define access as comprised of measurements of prior learning opportunities for computing, perceptions of computer science, and encouragement to pursue computing. The factors (SES, spatial ability, access to computing, and CS achievement) were measured through surveys completed by 163 students in introductory computing courses at a college level. Through the use of exploratory structural equation modeling, we found that these variables do impact each other, though not as we originally hypothesized. For our sample of students, we found spatial ability was a mediating variable for SES and CS achievement, but access to computing was not. Neither model explained all the variance, and our subject pool of US college students had higher than average SES. Our findings suggest that SES does influence success in computer science, but that relationship may not be due to access to computing education opportunities. Rather, SES might be influencing variables such as spatial ability which in turn influence CS performance.

CCS CONCEPTS

• **Social and professional topics** → *Computer science education*;

KEYWORDS

socioeconomic status; spatial ability; access; CS achievement

1 INTRODUCTION

There is a strong, positive relationship between socioeconomic status (SES) and academic achievement [17, 34, 48]. Students from low-SES households are less likely to attain high scores on achievement tests and grade-point average (GPA) measures, while being from a high-SES household tends to predict academic success. This finding has been replicated in STEM fields [18], and we have evidence that this holds true for computer science achievement as well [24]. Obviously, it is not the mere presence of money that produces the ability to achieve in computer science. SES leads to other benefits, such as living in a neighborhood with less crime and better schools. Those other factors are more likely having an impact on academic achievement rather than just SES.

If we can define how SES impacts CS achievement, we might be able to mitigate the effect by designing interventions that would affect the intermediate variables. Socioeconomic status could affect access to computing hardware, broadband networks, community and family members with positive perceptions of computer science, encouragement to pursue computer science, availability of toys or trips to the museum that develop spatial reasoning skills, or other variables that might give a student a better chance at achieving in computer science [9, 13]. Giving every student enough wealth to boost their SES would likely be impossible. But some of those other intervening variables might be significant and be manipulable with reasonable resources. For example, we might be able to distribute low-cost hardware, if access to computing hardware turned out to be a significant intervening variable.

We wanted to begin to explore the *intervening variables* (also referred to as *mediating variables*) between SES and CS achievement. A better understanding of this could help inform interventions to help level the playing field for all students in CS. Our research question is: *What are the mediating variables X between socioeconomic status and computer science achievement such that socioeconomic status affects X and X affects CS achievement?*

We focus on two possible intervening variables: spatial ability and access to computing. Spatial ability, spatial reasoning, or spatial cognition deals with the locations of objects, their shapes, their relationship to each other, and the manipulation of them [26]. We refer to spatial reasoning as the assessment of spatial ability. Spatial ability is connected to SES [7, 23] and to CS achievement [9]. In this study, access to computing is defined by access to learning opportunities, as well as encouragement to pursue computing, and

perceptions of computing. Access to computing is also connected to SES [24] and CS achievement [2, 10]. We chose these variables because of their known connections to SES and CS achievement, but their unknown roles as intervening variables to describe the effect of SES on CS achievement.

We sought to build a novel model for computer science education to account for the observed connections between SES and CS achievement. To do this, we surveyed undergraduate students in their first college computer science course. We administered four surveys to assess SES, spatial ability, prior access to computing, and CS achievement. We created methods to score the surveys and then analyze the relationships between them. We began analyses with Pearson's correlations, which showed significant correlations between each of our four variables. We continued with exploratory structural equation modeling which resulted in a model of spatial ability as an intervening variable between SES and CS achievement, but access to computing was not found to be an intervening variable. We discuss the implications of our findings for the CS Education community.

2 EXISTING LITERATURE

Our work is grounded in the literature of the connections between SES, access to computing, spatial reasoning, and CS achievement. We separate our literature into the two intervening variables we explore: spatial reasoning and access to computing.

2.1 Spatial Reasoning

Our discussion of spatial reasoning in Computer Science stems from a previous report on the connection between these two variables, and the incidental role SES played in the study. Cooper et al. found that students did better on a CS assessment when they offered spatial skills training for 45 minutes for eight of the ten days of a workshop [9]. The other students received additional review time during those 45 minutes. In other words, even with less CS content, students with spatial skills training performed better than students without that spatial skills training and with more time on CS content. Their argument was built on the literature that shows correlations between spatial and programming abilities by attempting to improve spatial abilities in hope of improving programming ability [11, 12, 19, 20, 25, 46].

While SES was not part of the Cooper et al. initial hypotheses, their *post hoc* analysis found an interesting result. They found that spatial skills training helped low-SES students perform at the same level as their high-SES counterparts. The control group had a significant difference in performance between high- and low-SES students.

We build on the Cooper et al. study to understand why these connections occur. Below, we detail prior literature on the connections between SES and spatial reasoning, and spatial reasoning and CS achievement.

2.1.1 SES and Spatial Reasoning. Work in the areas of Developmental and Cognitive Sciences has found a connection between SES differences and disparities in spatial cognition, along with other neurocognitive functions [15, 27]. Furthermore, research has found that spatial reasoning is sensitive to SES differences, moderating the differences in spatial reasoning among gender groups; boys from

middle- to high- SES backgrounds outperform girls from similar backgrounds on spatial tasks, but there is little difference in performance among low-SES boys and girls [23]. Other studies have found that SES can affect a student's ability to apply spatial reasoning skills to their academic performance [7]. In Casey et al., spatial skills were found to relate to academic performance in high-SES communities, but not low-SES ones. They found that children in low-income communities were less likely than their affluent peers to benefit, in terms of academic performance, from good spatial skills. This is an extension of work that studies low-SES students with high-level computational skills that are unable to transfer their skills into a mathematics classroom [31, 32].

These results do not give us a clear answer to the question of the role of spatial reasoning and academic performance. Spatial reasoning seems to be playing a role for high-SES students, but not for low-SES students. We suggest that these prior studies are studying students where they are. Perhaps low-SES students take classes that do not tax their spatial reasoning. We can imagine teaching (for example) algebra in ways that make demands on spatial reasoning (e.g., that emphasize the *sides* of an equals sign and the need to *balance* the two sides [1]), or in ways that may not make as many demands (e.g., with a greater focus on mathematical calculation). Because of the differences in classes, we cannot clearly say if the need for spatial reasoning ability is inherent to the demands of these academic disciplines, or if spatial reasoning plays a role only when specific teaching methods are used or only for specific learning outcomes. Our study uses one measure of CS achievement, for students with varying SES backgrounds, in the same introductory classes.

2.1.2 Spatial Reasoning and CS Achievement. Wai et al. have solidified the importance of spatial reasoning in developing expertise in the STEM fields at large [43]. Looking within CS Education, several studies have explored the connection between spatial reasoning and CS achievement [9, 11, 12, 19, 20, 25, 46]. Measuring CS achievement through grades and spatial ability through a visualization task, the correlation is small, but positive [11]. Using the completion time of code comprehension exercises and mental rotation skills, there is a strong correlation [19, 20]. Spatial ability has been found to be a predictor of scores on a Logo program, BASIC exams, and adapted AP CS tests [9, 25, 46]. Other research has found a link between spatial cognition and source code navigation and program comprehension [12].

These studies have measured CS achievement in different ways and defined spatial ability through different measures, which makes it difficult to make a general statement about the relationship. Still, there are consistent positive correlations between spatial reasoning and CS achievement over a variety of measures, definitions, and curricula. These multiple studies showing a positive relationship create a kind of replication of the results across different contexts, strengthening the argument for the relationship between these two variables. In our study, we have a single definition of CS achievement and of spatial ability that we based on our survey instruments, which gives us the opportunity for a clearer statement about the relationship between the two variables.

2.2 Access to Computing

In our study, we hypothesized a path from *socioeconomic status* to *access to computing*, and then *access* to *computer science achievement*. This path was motivated by the existing literature that connected these variables.

We define access to computing across four different dimensions: formal exposure to computing, informal exposure to technology, perceptions of computing, and encouragement to pursue computing. This definition is further explored in Section 3.2.1.

2.2.1 SES and Access. Previous reports have seen that not only are higher-income households more likely to have computers in the home [2], the way in which these computers are used varies by SES. Based on reports from the National Telecommunication and Information Administration [38], SES also impacts the speed of internet connection in the home, number of computers per household, and the quality of those computers. In addition to technical factors, SES can also impact various social factors that relate to access – for instance, having peers [24] and family members [3] who are sophisticated users of technology can impact your own understanding of it, and these can also be affected by SES. Outside of the home, school-level SES can impact how computers are used. For instance, lower-SES teachers often have less technical support for their computers in the classroom [45] so they use them less often. Additionally, because they often can't assume that students will have home access to computers, they spend a large portion of their time teaching basic computer skills and are hesitant to send children home with computer assignments [44]. There are even broader differences in the ways in which access is provided to students in different SES schools – for instance, low-SES schools are more likely to use computers for "remediation of skills" and review, while higher-SES schools are more likely to use computers for creative expression [4].

We know that SES can be a determining variable as to whether students' perceptions of software are more affected by home computer or by in-classroom exposure [30], though how SES was measured in that case is unclear. One study found that SES (measured according to parents' occupation(s)) does not predict computer ownership but does affect attitudes, use, and competencies [40]. We also have evidence that students without prior access, exposure, and opportunities to use technology fall behind in college due to simply not knowing how to use the technological tools that colleges depend on in this digital age [13]. As evidenced above, these studies have operationalized SES differently and cover three different types of technology–software, computer ownership, and college technology use.

2.2.2 Access and CS Achievement. Past literature has also shown a link between home access to computing and achievement, although this research has focused on math and science achievement rather than computer science achievement in particular. For instance, Attewell and Battle [2] find that eighth grade students with home access to computers score an average of 5 points higher on math than those without home computers, even when controlling for various SES predictors. This relationship is found elsewhere as well, with home access to "information and communication technologies" related to math and science achievement in 9th graders in Turkey [10].

Table 1: Participant demographics

Gender	Race/Ethnicity	Major
71% Female	23% Asian	91% Non-Computing Major
28% Male	5% Black	8% Computing area Major
1% Did not disclose	64% White	<1% Undeclared
	8% Two or more	

3 METHODS

3.1 Student Population

Participants in this study were asked to complete four surveys: (1) *SES survey*, (2) *Access to Computer Science survey*, (3) *Spatial Reasoning Skills survey*, and (4) *a CS Achievement assessment*. These surveys were completed near the end of the participant's first college computing course, which was intentionally placed to prevent a floor-effect on the achievement assessment. Participants were recruited from two universities in the southeastern United States and given three weeks to complete this study for extra credit in their course. Alternatively, students could elect to complete a set of programming problems for the same extra credit. Most of our study population was recruited from an Introduction to Media Computation course, a media-centric computing course designed to interest students that are typically disinterested in traditional computing courses. Six students came from another introductory course designed for computer science majors using the C programming language.

195 responses were recorded to our consent form. However, some participants submitted multiple times or did not complete all the forms. If a student repeated a submission, they were excluded from the final pool if their results were not consistent across submissions. 163 participants met all requirements for inclusion into the study. A summary of participant demographics can be found in Table 1. In addition to those demographics, our participants had a mean SES value of 11.18 on a 14-point scale, with a standard deviation of 1.94.

3.2 Assessment Instruments

Four survey instruments were used in our study to measure each of our variables. We used the validated Family Affluence Scale (FAS III) as our SES survey instrument. This survey consists of five multiple choice questions that reflect "market forces, economic trends, technological advances, as well as cultural, social and geographical norms in consumption across Europe and North America" [16]. We used the Revised Purdue Spatial Visualization Test (Revised PSVT:R) as an indicator of spatial ability. This assessment measures the ability to complete mental rotations, an indication of spatial ability [49]. We developed a survey to measure access to learning computing opportunities. This survey is described in more detail in Section 3.2.1. We measured CS Achievement with an adapted

version of the validated psuedocode-based CS1 assessment, SCS1 [29]. We drew 12 questions from the 27 question test, based on their previous difficulty and discrimination values from an item response theory analysis. We chose to maintain the make-up of the exam by including four questions from each question type (definitional, tracing, and code completion) and focused on four content areas (if statements, for loops, while loops, and logical operators). We collected demographic information following the completion of the four surveys to reduce stereotype threat [36].

3.2.1 Access Survey Construction and Quality.
We defined access using four constructs: formal access through school and summer camps; informal access through proximity to internet, technology, and media; perceptions of self and the field; and encouragement to pursue the field. An example of questions in each category can be found in Table 2.

Formal access is defined by access to learning opportunities with a curriculum. We have seen a similar distinction made in computing education research literature when considering the role of curriculum versus exploratory learning opportunities online [22]. Formal access includes traditional education pathways, such as whether students had opportunities to access CS teachers and computing courses in elementary, middle, or high school. We also included in this category outside-of-school learning opportunities with curricula, such as computing clubs or coding camps.

Informal access includes access to technologies that do not have a curriculum, but could be used for exploratory learning. This includes questions on access to the Internet, technology, and media. Questions on access to the Internet included whether they had an internet connection at home, but also what kind of internet access (DSL, dial up, etc.), and whether internet access was limited to a mobile phone. Questions on technology assessed if students had a personal or family computer growing up, where they had access to a computer (at home or a library) and whether they had someone in their household to fix their computer. Questions on media asked if students watched television or movies that were related to programming or computer science, and if they played video games.

Encouragement questions asked if anyone had encouraged the participant to pursue computing and, if so, who. We know from prior work that access can improve perceptions of ability, but perceptions of ability alone do not predict higher self-efficacy or intent to pursue computing without encouragement for females and members of under-represented minority groups [8, 14]. Perceptions of ability alone predicts self-efficacy and intent to persist for white or Asian males [14]. Because access alone is unlikely to be effective for underrepresented groups and women in computing without encouragement, we included encouragement as part of our access construct.

The perceptions portion of the survey included questions on whether students saw themselves represented in computing fields and if they felt they were computer scientists. We also asked questions about what the participants perceived the field to be, if they thought computing would be a viable career to them, and their values and whether computing met those values. We included these all in the perceptions variable, and included them in our hypothesized access construct because we saw the perceptions to be the desirable

outcomes of access. As we found in the Principal Component Analysis (Section 4.1), the perceptions variable does not represent the same underlying construct as the other three access components, and thus adds a new dimension to our access survey.

To grade the survey, three coders, including two authors, graded 11% of the questions and 16% of the participants. Each person graded individually. Each question was ranked on a 5-point scale. We performed an inter-rater reliability analysis using Cohen's Kappa to determine consistency among raters. This process was repeated, discussing the questions and codings, until a Cohen's Kappa of 0.7 was achieved. Then one coder graded all the surveys, according to the developed coding scheme, to construct the final access survey score, with accompanying categorical scores.

Since we constructed a survey to assess variables of access to computing, we assessed the reliability of the survey. We found the survey to have a Cronbach's alpha of 0.84, which indicates acceptable reliability [28].

3.3 Procedure
We distributed the study at the end of the semester after students had learned the concepts that were included in the CS achievement assessment. The entire study took place online. After obtaining participants' consent, participants were directed to the next survey. This process of linking surveys continued until all surveys were completed. Participants were directed, in this manner, to complete surveys in the following order: SES, spatial reasoning, access to computing, and CS achievement. We acknowledge that the Revised PSVT:R is supposed to be a timed test; takers are typically only allowed 25 minutes to complete. However, due to technological constraints, we could not make the spatial reasoning assessment timed. For completion of all the surveys and assessments, students received extra credit on one of their homework assignments in their computing course.

3.4 Structural Equation Modeling
In order to determine the relationship between our four constructed variables (SES, access to computing, spatial ability, and CS achievement), structural equation modeling (SEM) was employed. SEM can be thought of as a combination of exploratory factor analysis and multiple regression [41]. This method creates a series of regression equations to represent the hypothesized relationships being studied, and organizes those relations visually to create a clear conceptualization of the theory being explored [6]. SEM allows researchers to explore and test theory regarding how constructs are linked and the directionality of relationships [33]. SEM is the most appropriate method to answer our research question regarding mediating variables between SES and CS achievement.

SEM is confirmatory by nature, because of the emphasis on building models grounded in theory and literature [33]. SEM is not the same as Confirmatory Factor Analysis (CFA) modeling. CFA is a type of SEM, along with path analysis, structural regression models, and latent change models [39]. SEM can be exploratory when building structural regression models to test or disconfirm proposed theories involving explanatory relationships among various latent variables [39]. We withhold discussion on our individual models until Section 5 due to the importance of these models being

Table 2: Participant demographics

Category	Example question	Answer choices
Formal access	Did your school have a computer science teacher in elementary, middle, or high school?	Elementary School, Middle School, High School (Select many)
Informal access	If there was a computer in your family house, who was most likely to fix the computer when something went awry?	Dad, Mom , Brother, Sister, Me, Nobody in my house took care of it, Other (Select many)
Perceptions	Please rate how much you agree or disagree with the following statements: I can picture myself as a computer scientist.	Strongly Agree, Agree, Neutral, Disagree, Strongly Disagree (Select one)
Encouragement	Has anyone offered you personal advice on how to succeed in computing? Select all that apply.	HS Teachers, HS Advisors, College/Univ. Faculty, College/Univ. Advisors, Parents, Friends/Peers, Other (Please explain) (Select many)

considered in comparison with each other, rather than on their own.

There are five steps to build any SEM: model specification, identification, estimation, evaluation, and modifications [39]. Model specification is the step of gathering existing theories to formally state the hypothesized relationships among the variables. Hence, our models are built from the relationships explored in the literature described in Section 2. Model identification involves applying data to the variables in the hypothesized model. This data for our study is described in Section 3.2. Model estimation is using software to determine path coefficients between variables. In our study, we use the EQS software [5] to determine the impact of one variable on another. The scale of impact is described as a path coefficient, which is analogous to β in a regression equation [39]. These numbers are standardized, and typically fall in the range of -1 to 1. Model evaluation is using model fit indexes to determine how well the data fit the model. While there are dozens of fit indexes, we focus on Chi-square difference tests, Root mean square error of approximation (RMSEA), Comparative Fit Index (CFI), and Bayes Information Criterion (BIC). The last step of building a SEM is model modification, which involves adding or removing parameters to improve the fit. One of our models is a modified version of our original model, which lends itself to Chi-square difference tests to compare models. Another one of our models is not a modification of the original model, which necessitates the use of the BIC measure to compare model fit.

A brief history of SEM and a primer for its role in education research and practice can be found in Khine's book on the topic [21].

4 ANALYSIS

4.1 Principal Component Analysis

We used Principal Component Analysis to test whether our measurements of *access*, *spatial reasoning*, and *SES* were reasonable. First, we tested whether all the components of *access* belonged together. Then, we tested whether SES was really part of the same underlying construct as the others.

Formal access, informal access, encouragement, and perceptions were all expected to represent an underlying latent variable: access to computing. To test this expectation, participants' scores on these four components of the access survey were run through a principal component analysis, which tests whether scores on different measurements co-vary and, therefore, represent the same latent variable. The principal component analysis yielded one reliable component, Eigenvalue = 1.47, accounting for 37% of the variance, with an elbow in the scree plot at the second component, Eigenvalue = 0.97 and accounting for 24% of the variance. The loadings onto this variable were acceptable for formal access (0.69), informal access (0.65), and encouragement (0.67), but the loading for perceptions (0.36) was too low to include. Therefore, the latent variable represented by formal access, informal access, and encouragement is likely highly related to access to computing, and perceptions does not reliably represent the *access* variable.

After analyzing these four components in a separate analysis, participants' SES and spatial grade were added to an omnibus principal components analysis to ensure that they did not represent the same latent variable and contributed unique variance to the model. This analysis yielded two reliable components, Eigenvalues = 1.78 and 1.12 and accounting for 30% and 19% of the variance. The third component had an Eigenvalue of 0.93 and accounted for 16% of the variance. The resultant scree plot had no clear elbow, which makes sense given the loadings. The first component had sufficient loadings for formal access (0.59), informal access (0.66), and encouragement (0.56), similar to the previous analysis. The second component had sufficient loadings for spatial grade (0.63) and perceptions (0.52). SES did not load well onto either component (first = 0.49 and second = -0.48). Therefore, SES is likely somewhat related to each component, but independent enough to be its own component, hence the ambiguous third component.

The Principal Components Analysis gave us support for doing our analysis with access, spatial grade, and SES as separate and distinct variables. Although the perceptions variable does not represent the same underlying construct as the other three access components, we kept the variable in our models to represent a different dimension of our access survey.

4.2 Structural Equation Modeling

We present our three models and provide estimation and evaluation statistics for each. As mentioned in Section 3.4, we chose the model fit indexes of Chi-square difference tests, Root mean square error of approximation (RMSEA), and Comparative Fit Index (CFI). When
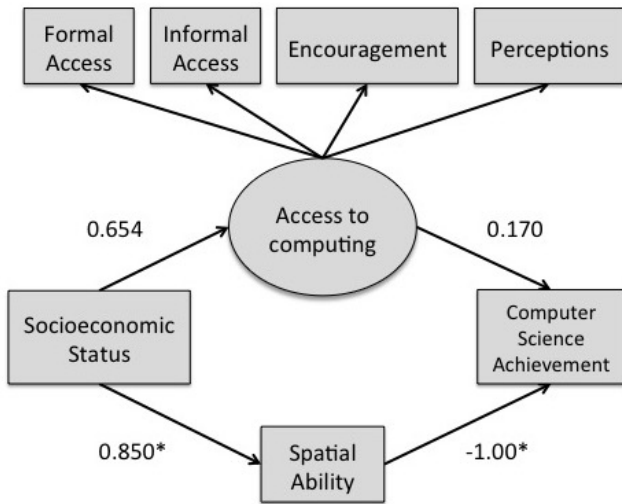
Figure 1: Our Model 1 includes the mediating variables of access to computing and spatial ability such that socioeconomic status affects access and spatial ability and access and spatial ability affects CS achievement.
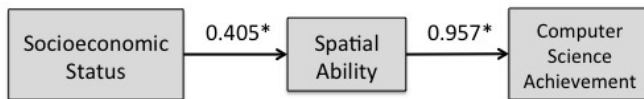

Figure 2: Our Model 2 includes the mediating variables of spatial ability such that socioeconomic status affects spatial ability and spatial ability affects CS achievement.

comparing models during an exploratory analysis, the higher CFI value is better and the smaller RMSEA and BIC values are better [33]. A summary of our analysis is presented in Table 3.

4.2.1 Model 1: Access and Spatial Ability. We began by testing the overall fit of our original hypothesized model, as seen in Figure 1. Model 1 stated that SES would have an impact on both access to computing and spatial ability, which in turn would each have an impact on CS achievement. Access to computing was a latent factor which was indirectly measured via scales of formal access, informal access, encouragement, and perceptions, as described in Section 3.2.1. In this model, spatial ability had a strong effect on CS achievement (β = -1, p <0.05) and SES on spatial ability (β = 0.850, p <0.05).

4.2.2 Model 2: Spatial Ability. We then tested Model 2, which was a modified version of our Model 1. Model 2, as seen in Figure 2, represented a hypothesis that access to computing did not play a role in affecting CS achievement. Rather, spatial ability is the only variable included to mediate the effect of SES on CS achievement. Thus, we removed all variables of access to computing, leaving a simplified model of SES having an impact on spatial ability, which in turn had an impact on CS achievement. This model again found a strong relationship between spatial ability and CS achievement (β
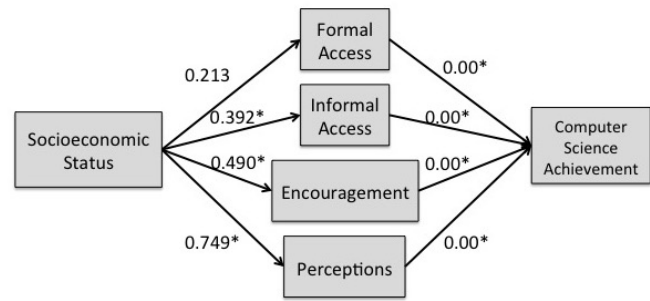

Figure 3: Our Model 3 includes the mediating variables of access to computing such that socioeconomic status affects access and access affects CS achievement.

= 0.957, p <0.05) and a relationship between SES and spatial ability (β = 0.405, p <0.05).

4.2.3 Model 3: Access. We created Model 3 from changing the variables explored in Model 1. Model 3, as seen in Figure 3, isolated access to computing, testing the impact that the components of access would have on CS achievement if spatial ability were not a factor. In contrast to Model 2, we removed the spatial ability variable and allowed for each aspect of access to be a separate, observed variable with a path from SES and to CS achievement. It should be noted that, since we changed the variables within the model, this model is not considered a modification of Model 1. However, this model included a significant relationship between SES and each variable of access, except for formal access. However, we found no relationship between the different variables of access and CS achievement (β = 0, p <0.05).

5 DISCUSSION

None of our models have statistical measurements that meet individual fit index thresholds for a good model (CFI > 0.95 for acceptance, RMSEA < 0.08) [33]. However, our study and use of SEM is an exploratory one, not confirmatory. Hence, we can compare the models against each other to determine which is the best fit.

5.1 Access and Spatial Ability as Mediating Variables

Model 1 is not as good as Model 2, because of the higher RMSEA and lower CFI. In other words, access and spatial ability are not likely to both be mediating variables. We believed this would be the case, given the evidence in the literature to support each path between SES, access, spatial ability, and computer science achievement. However, the literature support does not necessitate both access and spatial ability being mediating variables, which is mirrored in our model fit index comparisons.

5.2 Access as a Mediating Variable

Model 3 is not as good as the other models based on CFI and RMSEA values. CFI is lower and the RMSEA is higher in Model 3 than in Model 2 or Model 1. Additionally, because our Model 3 is not nested into Model 1, it is appropriate to compare BIC values between Model

Table 3: A summary of the model variables and fit indexes

Model	Intervening Variable(s)	Standardized Path Coefficient from SES [a]	Standardized Path Coefficient to CS Achievement [a]	χ^2	$\Delta\chi^2$	df	RMSEA [b]	CFI [c]	BIC [d]
1	Access	.654	.170	5712.967	–	21	1.2993	0.006	14386.361
	Spatial ability	.850*	-1.00*						
2	Spatial ability	0.405*	0.957*	5701.378	11.589	28	1.118	0.009	14385.721
3	Formal access	0.213	0.00*	5707.075	5.892	16	1.482	0.006	14452.543
	Informal access	0.392*	0.00*						
	Perceptions	0.749*	0.00*						
	Encouragement	0.490*	0.00*						

[a] * represents significance at the 5% level
[b] RMSEA: Root Mean Square Error of Approximation
[c] CFI: Comparative Fit Index
[d] BIC: Bayes Information Criterion

1 and Model 3. The BIC value is larger for Model 3, which again implies that Model 3 is not as good of a model. All of this points to Model 3 not being the best model that we explored. This means that the variables that we used to approximate access (formal access, informal access, perceptions, and encouragement) are not mediating variables between SES and CS achievement, for our sample.

Within the model, SES has a statistically significant impact on the variables that approximate access, except for formal access to CS. However, none of the variables had an impact on CS achievement. It is also interesting to note that the variable with the highest path coefficient with SES is perceptions, which we found in Section 4.1 to be representing a different construct than the other three access variables. We add support to the literature on SES affecting access, which is often assumed rather than shown. However, we encourage further research on the connections between access to computing and CS achievement. Perhaps *when* a child is exposed to computer science is not nearly as important as other factors, and thus the effects are mitigated by an introductory computing course.

5.3 Spatial Ability as a Mediating Variable

Model 2 is the best model among the three tested, implicating spatial ability as a mediating variable between SES and CS achievement. Model 2 has the highest CFI value and the lowest RMSEA value among the models tested. Since this model is nested, we can use a Chi-square difference test to compare Model 2 with Model 3. This test value is insignificant (p>0.05) in our case, which means both models fit equally well and the smaller model (Model 2) can be accepted just as well [47]. Furthermore, each path within the model is significant. SES has a medium effect on spatial ability, and spatial ability has a large effect on CS achievement.

This finding extends the literature on SES, spatial ability, and CS achievement. It means there is a connection between these three variables, more so than access to computing. Spatial ability is a better mediating variable for SES and CS achievement than access, or in addition to access to computing. We can begin to answer our research question with support that SES affects spatial ability and spatial ability affects CS achievement.

As we mentioned, this model, as is true for the others, did not meet thresholds of individual fit indexes for a "good" model. However, we were not using SEM to confirm a model by fitting it to data. Rather, we were trying to build a novel model for CS education, where there is a lack of theory to account for observed connections between SES and CS achievement. This model can serve as a foundation for continued study to understand how SES affects CS achievement.

6 LIMITATIONS

We identify our study's shortcomings here as we currently see them, in recognition of the need for future research to replicate and extend this work in broader contexts. Our study limitations include a privileged and biased sample, and the use of exploratory rather than confirmatory statistics.

It is important to note that a majority of the students came from an institution with a high average on standardized tests among students, including SAT (1450 out of 1600) and ACT (32 out of 36) scores. Additionally, the median family income is $130,000, which is among the highest among highly selective public colleges. This means that the population that we drew from is skewed towards high-SES and, likely, high-spatial ability participants.

Additionally, all of our participants came from large, public universities in the United States. Our model is thus based on a sample of individuals that are not representative of the whole. We encourage further research into whether our model holds in K-12 settings, private and community colleges, and non-American schools and universities.

Our study was exploratory in nature, rather than confirmatory. While we are not contributing indisputable evidence that spatial ability is a mediating factor between SES and CS achievement, we have compared models and selected the best one to conduct confirmatory analyses with different populations. Our model reduces uncertainty in the community about these variables, and allows for more rigor and systematicity than starting future studies only at the exploratory stage.

7 CONCLUSION

We started this exploration with a hypothesis that socioeconomic status (SES) likely influenced CS achievement through the intervening variable of access. We thought that high-SES students likely had more positive access to computing education before they entered their first CS classes, and that's what led to higher achievement. However, our results do not support that hypothesis.

Instead, we find that spatial ability is a more powerful intervening variable than access. We had prior evidence from Cooper et al. that the impact of SES on CS achievement was mediated by spatial ability [9]. Our study specifically looked at that relationship, and our findings support it. Our results suggest that high-SES students tend to have higher spatial ability, and that higher spatial ability, or the better ability to make use of spatial reasoning, thus predicts greater CS achievement. Students from low-SES backgrounds tend to have lower spatial ability, or are less able to make use of spatial reasoning, which may be inhibiting their success in CS classes.

While surprising, the result is a positive one. Spatial ability can be taught [37]. David Uttal and his colleagues developed an approach to teaching spatial ability that measurably led to improved spatial ability that transfered outside the original testing context and was retained for months later [42]. Sheryl Sorby successfully taught spatial ability to Engineering students, which resulted in better performance in Engineering classes [35]. Spatial ability is an intervening variable that we can manipulate without changing students' SES.

We are not claiming that we have made an exhaustive search for intervening variables. We certainly should explore more. SES, spatial ability, and access do not explain all of CS performance. The more we understand the relationship between SES and CS performance, the more we might be able to mitigate the effects of low-SES background in students.

While we have support for the model explaining SES impact on CS performance with mediation from spatial ability, we are not convinced that this model is complete and exhaustive. Because we gathered data only at the post-secondary school level, we are working from a biased sample. All of the students we studied already made it to post-secondary school. Any low-SES students in our sample already overcame odds to make it to this level. We do not know much about low-SES students who tried CS before the post-secondary level.

There may be different models at play between SES and CS performance at the elementary and secondary school level. In particular, access may play a more critical role in primary or secondary school achievement. Access is likely an important variable in broadening participation in computing, but its impact may not be on CS achievement. For example, a lack of access may lead to higher attrition, so we do not even see the students without access in our sample populations.

Our current model gives us a lever. We now have an explanation for why SES impacts CS performance, and that explanation suggests a possible intervention. That is a useful contribution, both for understanding CS performance and for finding ways to mitigate low-SES conditions.

REFERENCES

[1] Abraham Arcavi. 2003. The role of visual representations in the learning of mathematics. *Educational studies in mathematics* 52, 3 (2003), 215–241.

[2] Paul Attewell and Juan Battle. 1999. Home Computers and School Performance. *The Information Society* 15, 1 (1999), 1–10. https://doi.org/10.1080/019722499128628 arXiv:https://doi.org/10.1080/019722499128628

[3] Brigid Barron, Caitlin Kennedy Martin, Lori Takeuchi, and Rachel Fithian. 2009. Parents as learning partners in the development of technological fluency. (2009).

[4] Henry Jay Becker. 2000. Who's wired and who's not: Children's access to and use of computer technology. *The future of children* (2000), 44–75.

[5] Peter M Bentler. 1995. *EQS structural equations program manual*. Multivariate software.

[6] Barbara M Byrne. 1994. *Structural equation modeling with EQS and EQS/Windows: Basic concepts, applications, and programming*. Sage.

[7] Beth M Casey, Eric Dearing, Marina Vasilyeva, Colleen M Ganley, and Michele Tine. 2011. Spatial and numerical predictors of measurement performance: The moderating effects of community income and gender. *Journal of Educational Psychology* 103, 2 (2011), 296.

[8] J McGrath Cohoon and Katharine M Baylor. 2003. Female graduate students and program quality. *IEEE Technology and Society Magazine* 22, 3 (2003), 28–35.

[9] Stephen Cooper, Karen Wang, Maya Israni, and Sheryl Sorby. 2015. Spatial skills training in introductory computing. In *Proceedings of the eleventh annual International Conference on International Computing Education Research*. ACM, 13–20.

[10] Erhan Delen and Okan Bulut. 2011. The relationship between students' exposure to technology and their achievement in science and math. *TOJET: The Turkish Online Journal of Educational Technology* 10, 3 (2011).

[11] Sally Fincher, Bob Baker, Ilona Box, Quintin Cutts, Michael de Raadt, Patricia Haden, John Hamer, Margaret Hamilton, Raymond Lister, and Marian Petre. 2005. Programmed to succeed?: A multi-national, multi-institutional study of introductory programming courses. (2005).

[12] Maryanne Fisher, Anthony Cox, and Lin Zhao. 2006. Using sex differences to link spatial cognition and program comprehension. In *Software Maintenance, 2006. ICSM'06. 22nd IEEE International Conference on*. IEEE, 289–298.

[13] Joanna Goode. 2010. Mind the Gap: The Digital Dimension of College Access. *The Journal of Higher Education* 81, 5 (2010), 583–618. https://doi.org/10.1353/jhe.2010.0005

[14] Mark Guzdial, Barbara Ericson, Tom McKlin, and Shelly Engelman. 2012. A statewide survey on computing education pathways and influences: factors in broadening participation in computing. In *Proceedings of the Ninth Annual International Conference on Computing Education Research (ICER'12)*. ACM, New York, NY, 143–150.

[15] Daniel A Hackman and Martha J Farah. 2009. Socioeconomic status and the developing brain. *Trends in cognitive sciences* 13, 2 (2009), 65–73.

[16] Jane E. K. Hartley, Kate Levin, and Candace Currie. 2016. A new version of the HBSC Family Affluence Scale - FAS III: Scottish Qualitative Findings from the International FAS Development Study. *Child Indicators Research* 9, 1 (01 Mar 2016), 233–245. https://doi.org/10.1007/s12187-015-9325-3

[17] John Hattie. 2008. *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. routledge.

[18] Thomas B Hoffer et al. 1995. Social Background Differences in High School Mathematics and Science Coursetaking and Achievement. Statistics in Brief. (1995).

[19] Sue Jones and Gary Burnett. 2008. Spatial ability and learning to program. *Human Technology: An Interdisciplinary Journal on Humans in ICT Environments* (2008).

[20] Sue Jane Jones and Gary E Burnett. 2007. Spatial skills and navigation of source code. In *ACM SIGCSE Bulletin*, Vol. 39. ACM, 231–235.

[21] Myint Swe Khine. 2013. *Application of structural equation modeling in educational research and practice*. Springer.

[22] Michael J Lee and Andrew J Ko. 2015. Comparing the effectiveness of online learning approaches on CS1 learning outcomes. In *Proceedings of the eleventh annual International Conference on International Computing Education Research*.

ACM, 237–246.

[23] Susan C Levine, Marina Vasilyeva, Stella F Lourenco, Nora S Newcombe, and Janellen Huttenlocher. 2005. Socioeconomic status modifies the sex difference in spatial skill. *Psychological science* 16, 11 (2005), 841–845.

[24] Jane Margolis, Rachel Estrella, Joanna Goode, Jennifer Jellison Holme, and Kim Nao. 2010. *Stuck in the shallow end: Education, race, and computing.* MIT Press.

[25] Richard E Mayer, Jennifer L Dyck, and William Vilberg. 1986. Learning to program and learning to think: what's the connection? *Commun. ACM* 29, 7 (1986), 605–610.

[26] Nora S Newcombe. 2010. Picture this: Increasing math and science learning by improving spatial thinking. *American Educator* 34, 2 (2010), 29.

[27] Kimberly G Noble, M Frank Norman, and Martha J Farah. 2005. Neurocognitive correlates of socioeconomic status in kindergarten children. *Developmental science* 8, 1 (2005), 74–87.

[28] Jum C Nunnally and Ira H Bernstein. 1978. Psychometric theory. (1978).

[29] Miranda C Parker, Mark Guzdial, and Shelly Engleman. 2016. Replication, validation, and use of a language independent CS1 knowledge assessment. In *Proceedings of the 2016 ACM conference on international computing education research.* ACM, 93–101.

[30] Nichole Pinkard. 2005. How the Perceived Masculinity and/or Feminity of Software Applications Influences Students' Software Preferences. *Journal of Educational Computing Research* 32, 1 (2005), 57–78.

[31] Geoffrey B Saxe. 2002. Candy selling and math learning. *Teaching and learning: The essential readings* (2002), 86–106.

[32] Geoffrey B Saxe, Venus Dawson, Randy Fall, and Sharon Howard. 1996. Culture and childrenâĂŹs mathematical thinking. *The nature of mathematical thinking* (1996), 119–144.

[33] James B Schreiber, Amaury Nora, Frances K Stage, Elizabeth A Barlow, and Jamie King. 2006. Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of educational research* 99, 6 (2006), 323–338.

[34] Selcuk R Sirin. 2005. Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of educational research* 75, 3 (2005), 417–453.

[35] Sheryl A Sorby. 2009. Educational research in developing 3-D spatial skills for engineering students. *International Journal of Science Education* 31, 3 (2009), 459–480.

[36] Claude M Steele and Joshua Aronson. 1995. Stereotype threat and the intellectual test performance of African Americans. *Journal of personality and social psychology* 69, 5 (1995), 797.

[37] Mike Stieff and David Uttal. 2015. How much can spatial training improve STEM achievement? *Educational Psychology Review* 27, 4 (2015), 607–615.

[38] National Telecommunications and Information Administration. 2008. Networked Nation: Broadband in America, 2007. (2008).

[39] Timothy Teo, Liang Ting Tsai, and Chih-Chien Yang. 2013. Applying Structural Equation Modeling (SEM) in Educational Research. In *Application of structural equation modeling in educational research and practice.* Springer, 3–21.

[40] J. Tondeur, I. Sinnaeve, M. van Houtte, and J. van Braak. 2011. ICT as cultural capital: The relationship between socioeconomic status and the computer-use profile of young people. *New Media & Society* 13, 1 (2011), 151–168. https://doi.org/10.1177/1461444810369245

[41] Jodie B Ullman and Peter M Bentler. [n. d.]. *Structural equation modeling.* Wiley Online Library.

[42] David H Uttal, Nathaniel G Meadow, Elizabeth Tipton, Linda L Hand, Alison R Alden, Christopher Warren, and Nora S Newcombe. 2013. The malleability of spatial skills: A meta-analysis of training studies. *Psychological bulletin* 139, 2 (2013), 352.

[43] Jonathan Wai, David Lubinski, and Camilla P Benbow. 2009. Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology* 101, 4 (2009), 817.

[44] Mark Warschauer, David Grant, Gabriel Del Real, and Michele Rousseau. 2004. Promoting academic literacy with technology: Successful laptop programs in K-12 schools. *System* 32, 4 (2004), 525–537.

[45] Mark Warschauer, Michele Knobel, and Leeann Stone. 2004. Technology and equity in schooling: Deconstructing the digital divide. *Educational policy* 18, 4 (2004), 562–588.

[46] Noreen M Webb. 1984. Microcomputer learning in small groups: Cognitive requirements and group processes. *Journal of Educational Psychology* 76, 6 (1984), 1076.

[47] Christina Werner and Karin Schermelleh-Engel. 2010. Deciding between competing models: Chi-square difference tests. *Goethe University. Available online: https://perma. cc/2RTR-8XPZ (accessed on 21 July 2017)* (2010).

[48] Karl R White. 1982. The relation between socioeconomic status and academic achievement. *Psychological bulletin* 91, 3 (1982), 461.

[49] SY Yoon. 2011. Revised Purdue Spatial Visualization Test: Visualization of Rotations (Revised PSVT: R)[Psychometric Instrument]. (2011).