Spatial Skills and Demographic Factors in CS1

Anna Ly
University of Toronto Mississauga
Canada
annahuynh.ly@mail.utoronto.ca

Jack Parkinson
University of Glasgow
UK
jack.parkinson@glasgow.ac.uk

Quintin Cutts
University of Glasgow
UK
quintin.cutts@glasgow.ac.uk

Michael Liut
University of Toronto Mississauga
Canada
michael.liut@utoronto.ca

Andrew Petersen
University of Toronto Mississauga
Canada
andrew.petersen@utoronto.ca

ABSTRACT

Motivation Prior studies have established that training spatial skills may improve outcomes in computing courses. Very few of these studies have, however, explored the impact of spatial skills training on women or examined its relationship with other factors commonly explored in the context of academic performance, such as socioeconomic background and self-efficacy.
Objectives In this study, we report on a spatial skills intervention deployed in a computer programming course (CS1) in the first year of a post-secondary program. We explore the relationship between various demographic factors, course performance, and spatial skills ability at both the beginning and end of the term.
Methods Data was collected using a combination of demographic surveys, existing self-efficacy and CS1 content instruments, and the Revised PVST:R spatial skills assessment. Spatial skills were evaluated both at the beginning of the term and at the end, after spatial skills training was provided.
Results While little evidence was found to link spatial skills to socioeconomic status or self-efficacy, both gender identity and previous experience in computing were found to be correlated to spatial skills ability at the start of the course. Women initially recorded lower spatial skills ability, but after training, the distribution of spatial skills scores for women approached that of men.
Discussion These findings suggest that, if offered early enough, spatial skills training may be able to remedy some differences in background that impact performance in computing courses.

CCS CONCEPTS

• **Social and professional topics → Computing education**.

KEYWORDS

spatial skills, gender, socioeconomic status, retention, CS1

Koli Calling '21, November 18–21, 2021, Joensuu, Finland

ACM ISBN 978-1-4503-8488-9/21/11…$15.00
https://doi.org/10.1145/3488042.3488049

ACM Reference Format:
Anna Ly, Jack Parkinson, Quintin Cutts, Michael Liut, and Andrew Petersen. 2021. Spatial Skills and Demographic Factors in CS1. In *21st Koli Calling International Conference on Computing Education Research (Koli Calling '21), November 18–21, 2021, Joensuu, Finland*. ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3488042.3488049

1 INTRODUCTION

Spatial skills have been associated with achievement in STEM subjects for decades [50]. The exact mechanisms by which spatial skills affect achievement are not well understood [30], but the evidence that improving spatial skills can, in turn, improve outcomes in STEM classrooms is well-established [45, 52]. Furthermore, there is evidence that spatial skills training transfers [52] and may be of particular value to under-represented populations [21].

Within computing specifically, recent studies have established a link between spatial skills and outcomes in computing courses [6, 13, 36]. However, very few studies have investigated the impact of spatial skills training on under-represented groups, such as women, or groups that are at risk of lower performance in computing courses, such as those from less affluent socioeconomic backgrounds. In this study, we report on an online training intervention deployed in a first year computer programming (CS1) course at a large, research-intensive university. We report on both the outcomes of the training and on relationships observed between spatial skills and various participant factors. The factors – socioeconomic status (including household income and parental education), self-efficacy, and gender – were selected because of their prominence in the literature. In particular, we seek to explore the following research questions:

- RQ1 Is online spatial skills training effective?
- RQ2 Do men and women both benefit from spatial skills training?
- RQ3 Are spatial skills related to other factors (i.e., socioeconomic status, self-efficacy, or previous experience) associated with differences in academic performance in computing?

In the next section, we introduce the prior work that informs these research questions, as we describe the current state of research in spatial skills and then discuss the factors that have been found to impact academic performance. Next, in Section 3, we introduce the educational context in which our study was conducted and describe how data was collected. We report on our observations in Section 4 and then discuss how that evidence addresses our research questions in Section 5. In the same section, we also explicitly identify threats to the validity of the study. Finally, we conclude with a few notes about future work.

2 RELATED WORK

"Spatial skills" is an umbrella term encapsulating a range of skills related to visual processing and internal visualisation, which can be described as "the ability to generate, retain, and manipulate abstract visual images" [12, 29]. Spatial skills have been associated with STEM achievement since the 1950s [50], with various studies over the years confirming and strengthening the relationship.

Study of the relationship and impact of spatial skills is particularly developed in engineering, where Sorby (see, e.g., [44–48]) has designed training materials for spatial skills and shown that training students on spatial skills improves academic outcomes and increases retention [46]. Sorby's training method consists of a set of exercises involving spatially loaded tasks such as isometric and orthographic drawing, rotation of 3D objects, projections of 2D shapes to 3D objects, flat pattern folding and more. These tasks have been shown to reliably develop spatial skills and have been developed into a workbook [48].

The spatial skills test typically used by Sorby is the Revised PSVT:R [59], a test of mental rotation. The participant must determine the rotation steps between two provided orientations of an object, then apply the same sequence of rotations to another given object and determine the correct result from a selection of five options. The test consists of 30 items ordered by increasing complexity (that is, later items require more rotations in more axes than previous items). The test is issued with a time limit of 20 minutes. Sorby uses a pass/marginal/fail system to determine whether students require training, where a pass is any score above 21, marginal is any score from 18 to 21 and a fail is 18 and below. Typically, all failing students are required to take additional training. Training is offered to marginally passing students but is not a requirement. The Revised PSVT:R (with the same pass/fail/marginal breakdowns applied) and an online implementation of Sorby's workbook are the instruments used throughout this study.

The relationship between computing achievement and spatial skills has seen a peak in interest, with multiple distinct research projects and publications examining different facets of the connection. Research has shown relationships between spatial skills and grades [22], aptitude for source code navigation [23], level of academic attainment in computing [34, 54], achievement in validated CS1 assessment tools [7, 13], computational thinking activities at the primary level [31], low-level expression evaluation [36], and (as a mediating factor) socioeconomic status and CS achievement [33].

Three independent studies have also demonstrated that training spatial skills can have a positive effect on CS outcomes, two of which focused on CS1 cohorts [6, 13, 36]. Bockmon et al. attempted an intervention with a CS1 cohort over two years: no intervention was used in the first year (control) and a paid voluntary spatial skills intervention was used in the second (treatment). Both cohorts completed a revised version of the SCS1 programming test (SCS1R) [5, 32] at the start and end of the academic year. While there was no significant difference between each cohort's pre-SCS1R scores, indicating that the groups were originally similar, the treatment group showed a significantly higher gain in the post-SCS1R scores than the control group.

Parkinson and Cutts attempted a similar intervention in the same year across a CS1 and a CS0 cohort [34]. Students were allocated into the treatment group based on a spatial skills test (Revised PSVT:R) at the start of the academic year, with students scoring below a passing threshold (18/30, as previously used in engineering by Sorby [45]) required to take a mandatory spatial skills course. The CS1 cohort had only limited course data collected and generally showed a weak correlation between spatial ability and CS assessment. The CS0 cohort, however, indicated a strong correlation between spatial skills and success in CS assessment, with those taking training improving at a much greater rate than their peers who scored marginal passes and were not required to take training.

The mechanism by which spatial skills impact computing achievement is also an area of active research interest. Parkinson and Cutts proposed a model in which spatial skills are related to code reading, the identification of key points in the code, and program comprehension [34]. Margulieux, taking a wider view, proposed Spatial Encoding Strategy Theory to explain the relationship between STEM fields, in general, and spatial skills [30]. Spatial Encoding Strategy Theory suggests that the process of developing spatial skills in turn develops generalizable strategies for encoding and orienting non-verbal information.

2.1 Academic Achievement in Computing

The impact of contextual factors on tertiary-level academic achievement is well-studied, with researchers in various fields exploring a wide range of cognitive, psychosocial (e.g., motivation, approaches to learning, personality traits), and contextual (e.g., socioeconomic status, secondary schooling, first-generation status) factors. In a large meta-analysis of research conducted between 1997 and 2010, Richardson et al. found that self-efficacy had the highest correlation with performance, followed by secondary school and standardized test performance, and finally various demographic and psychosocial factors [38]. The top correlate, self-efficacy, has strong connections to academic self-concept, with self-efficacy better predicting academic achievement and self-concept predicting and mediating motivation [8, 17]. Both are shaped by mastery experiences and social context, such as minority status within a group or socioeconomic status. Earlier work had found a strong relationship with family socioeconomic status (SES) [43], and a number of the factors considered by Richardson et al. might reflect SES.

Within computing, one focus has been on *predicting* retention and performance in the first programming course, often using measures from within the course [20], though some work has focused on *explaining* differences or on earlier experiences. For example, a 2018 ITiCSE working group explored early developmental activities – types of play – that might correlate with computing proficiency and found that early childhood experiences have some impact on success [16]. Within the first programming course, several factors have been considered including standardized tests and secondary school grades [39], mathematical ability [39, 56, 57], programming behaviours [55], abstraction ability [3], mindset [14, 56], self-efficacy [28, 57], the ability to articulate strategy [15], study strategies and time spent [28, 57], both formal and informal previous experiences [51, 57], and gender [28, 56, 57].

From this varied list, a few ideas stand out. First, many studies rely on grades (either in prior contexts or the context of interest) or engagement in course activities as data [20]. These may be observable indicators of underlying factors such as motivation, attitude, or SES, but these studies do not, in general, seek to identify the underlying factors which impact prior achievement or current engagement. Second, only a few studies have sought to link potentially *modifiable* attributes (e.g., mathematical ability, mindset, and study strategies) to performance.

Spatial skills, then, are important as they can be measured prior to the course and improved. However, a limited number of studies have investigated the relationship between spatial skills and other factors associated with academic performance in computing. Cooper et al. found that students at all SES levels (using the student's school as a proxy) benefited from spatial skills training [13]. No information about relationships between spatial skills and SES before training were reported. Later, Parker et al. provided evidence suggesting that spatial skills might act as a mediator in the relationship between SES and computing outcomes [33].

2.2 Gender and Spatial Skills

Due to persistent issues with under-representation in computing, gender and factors related to it have received particular attention [40, 41]. Under-represented ethnic groups are less well studied, largely because their under-representation is so pronounced as to make measurement difficult. Work investigating the impact of spatial skills training on under-represented groups in computing is sparse. In a small study, Cooper et al. found suggestive evidence that spatial skills training helped students (mostly women) from under-represented groups but suggested that a larger study was needed [13].

Outside of computing, several groups have argued that incorporating spatial skills into the curriculum is particularly important for supporting women [21, 25]. There is consistent evidence that women under-perform on some spatial skills measures [2, 9, 27, 53], though recent work has suggested that testing procedures may have magnified any actual difference [18]. As a result, the source of this difference – and its magnitude – remain a topic of debate [4, 18]. Nevertheless, one measure in which women under-perform, mental rotation (measured by the PSVT:R), has been argued to influence career and choice of study [37], making gender an important factor to consider when studying the impact of spatial skills in the context of computing. Fortunately, studies have consistently found that spatial skills interventions can reduce the measured gap [24, 52], suggesting that spatial skills interventions can help increase participation in subjects where women are currently under-represented [52].

3 METHOD

This study was conceived as a form of replication to examine prior claims that spatial skills training improve course outcomes in computing. We also seek to contribute to understanding of the phenomenon by specifically examining the impact of spatial skills training on an under-represented group (women) and the relationship of spatial skills to other factors correlated with performance.

This study was conducted during the first semester (September to December) of the 2020-21 academic year. Data was collected in a Python-based CS1 course at a research intensive, North American public university. All students enrolled were invited to participate in the study. 778 students were enrolled in the course after the deadline to add courses, and 670 completed the course. 78% of the students who started the course consented for their survey data to be used under a process approved by our IRB.

3.1 Contextual Information

The course is open to all students in the university but is mandatory for students in a computing program. It is also required for students in a mathematics or statistics program. No prior experience is assumed. About 60% of students enrolled in the course intend to major in a computing related field, and almost all of the students are in their first year of studies at the post-secondary (university) level. The gender breakdown of students in the course was heavily skewed, with self-reported men making up 74% of the population and self-reported women making up 25%. A very small number (around 1%) of students self-reported as another gender. The number is too small to be broken out in its own group, so statistics were performed on "men" and "women and other genders."

While a public institution, the university has a high number of students from other geographic regions. 49% of the students in the course resided in the university's region (think of the "region" as a county or a large metropolitan area). 13% were from outside the region but were "domestic" (residents of the same country). 37% were "international" (residents of another country). To add to this, the region features a relatively large number of recent immigrants, many of whom speak a language other than the language of instruction at home. Only 29% of the entire class reported speaking the language of instruction at their childhood home.

Due to COVID-19, the course was conducted entirely online. The university has previously offered a fully online CS1 offering [10] in addition to the more common inverted (flipped-classroom) offering [11], so a proven course structure and course materials were already prepared and ready. Additionally, the instructors were comfortable with the online environment due to an earlier term teaching online [60].

The course was structured as a series of one-week modules. In each week, the students reviewed preparatory material (videos with multiple choice and short answer questions to test comprehension), actively practiced the content and skills demonstrated during synchronous class and lab meetings, and then completed a set of coding exercises to gauge their comprehension of the material. Despite the expectation that students attend and participate in class meetings, all meetings were recorded to give students a chance to review or to attend asynchronously if needed. Labs were not recorded, but enough sessions were held that students could attend a meeting convenient to their timezone.

Spatial skills training was added to the course for this offering. The instructors were provided with published research on the relationship between spatial skills and computing and were consulted on the implementation of the intervention. The instructors decided, given evidence that spatial skills transfer to other contexts and that spatial skills may be more important to under-represented populations, that they would prefer that *all* students complete the spatial skills training regardless of initial spatial skills ability, so the training was added to the weekly preparation in the first two-thirds of the course as described in the next section. In addition, data from additional instruments was collected at the beginning and end of the course, as described in Section 3.3.

3.2 Spatial Skills Training

The spatial skills training was distributed across the first eight weeks of the course. As decided by the course instructors, these exercises were mandatory and accounted for 5% of their final mark. None of the course content outside of the spatial skills modules were directly related to the spatial syntax exercises, and prior to the midterm at week six, students were informed that they did not need to review the spatial skills-related material for tests.

The spatial skills training was delivered in a series of online modules. Sorby's original workbook [48] consists of 9 chapters of drilling exercises involving a range of spatially loaded skills, including grid-guided isometric and orthographic drawing, 3D rotation, 2D flat pattern "folding" to produce 3D objects, solids of revolution, reflection, symmetry and other similar activities. A bespoke online platform was developed to present the exercises. This was a necessary endeavour since many of the questions (particularly the drawing exercises) required complex interactions which could not be effectively recreated in any known existing visual learning environment or online delivery platform. The web application also featured automatic marking, generalised feedback and tutorial/example videos, centralising the entire spatial skills training experience into a single platform. Sorby's original 37 exercises, formulated of a total of 500 questions, were all included and spread across the 8 weekly modules.

After each weekly module, students were given an opportunity to voice feedback about their experience with the tool. This feedback caused a change in the planned delivery of the training. The modules were spread across eight weeks, rather than an originally planned six, due to student concerns about the amount of time that spatial skills training consumed. In addition, at the end of the term, the students were provided one more opportunity to provide feedback and were also asked directly about how often they guessed on or copied answers for the spatial skills training.

The same system was used to deploy two spatial skills assessments (the PSVT:R [59]): a pre-test in the first week and a post-test after training was completed. Two opportunities to complete the post-test were provided, as completion rates of the first offering were low due to competing deadlines. Some students decided to take both tests, and we recorded the first complete post-test score.

3.3 Research Data Collection

Based on our survey of prior work, we selected a number of potential factors to investigate in relation to spatial skills. Prior experience, socioeconomic status, self-efficacy, and gender were selected for investigation based on prior work within computing education (for prior experience, socioeconomic status, and gender) and prominence in the overall study of tertiary academic performance (for self-efficacy). Data about location of residence and native language(s) were also collected due to the nature of the university context. We did not collect secondary schooling data. Due to the high proportion of out of state and international students, comparisons between secondary contexts is difficult, and the university does not use a standardized test for admissions.

Students were asked to complete both a self-efficacy instrument [49] in the first and last weeks of the academic term, and they completed a CS1 knowledge assessment (the SCS1) [32] in the first week of the term. These are labeled as "pre-tests". The end of term assessment was completed in the last week of instruction but prior to the course's final exam and is the "post-test". Time controls for both instruments matched the original authors: no restrictions for self-efficacy and a one hour limit for the SCS1.

As directed by the original authors of the self-efficacy instrument [49], factor analysis was used to examine the factorization of the self-efficacy model obtained. It largely matched the model reported in the paper. To make comparisons about self-efficacy over the term, we assigned a score from 1 ("strongly disagree") to 7 ("strongly agree") to their responses per question. We computed the average response for each student to represent their self-efficacy. Some students omitted one or more questions on the self-efficacy instrument, and the averages computed for them simply disregard skipped items. This allows for comparisons between the pre- and post- tests for a single student and also for evaluation of the average self-efficacy of all students in the course. We expected an overall increase in self-efficacy over the term, as novice self-efficacy will drop upon introduction to the material and then gradually rebound [49].

The SCS1 [32] was deployed as a proxy for previous experience. Xie et al. raised concerns about the difficulty of the test and suggested that its use as a pre-test might encounter floor effects [58], but no better measures of experience have, as yet, been widely adopted by the community. Recent work has focused on shortening the SCS1 [5] and even the original authors have used a subset of the tool [33], but we deployed the full SCS1, as it remains the most studied version. The instrument was deployed using a Qualtrics survey provided by the original authors, and scores were computed automatically by the survey. One error was found in the survey items; that error was corrected before scores were collected.

In the third week of the term, after the deadline for adding the course, students were asked for consent to use their data and provided with an optional demographic survey. The questions on the survey are provided in Appendix A. The students are asked to self-report their gender, the location of their family's residence, and languages spoken in their home. The students were also asked a number of questions related to socioeconomic status. As a primary measure, we directly asked for their household income despite concerns that some students – and particularly lower-achieving students – may not be able to report accurately [43]. Due to concerns about low response rates to the household income measure, we also asked about parental education, as it represents another facet of socioeconomic status [43].

Since several separate instruments are being used, there were multiple opportunities for participants to not complete an instrument. We attempted to be resilient to data loss and only applied one blanket filter: any participant who failed to complete the spatial skills pre-test was removed. Unless otherwise stated, for all the data presented in the next section, results were computed using all of the participants who completed the relevant instruments. This leads to the number of participants varying slightly between tests.

Relationships between the various factors are tested using Spearman's rank correlation. Spearman's was chosen since the data contained ordinal elements. Effect sizes and p-values are reported, rather than reporting significance based on some alpha. When comparing distributions, a Wilcoxon rank sum test is used, since the distributions observed were not normal. p-values and a confidence interval are reported.

4 RESULTS

This section presents the data that was collected throughout the term. We begin by focusing on the relationship between spatial skills, both before and after training, and performance in the course. Then, we examine data corresponding to each of the factors described in Section 3.3: prior experience, household income and parental education, performance self-efficacy, and gender.

4.1 Spatial Skills Training

The median on the spatial skills pre-test (week 1 of the term) was 19 (n=630), in Sorby's "marginal" range. The median of the post-test score (after training) was 23 (n=537), in Sorby's "pass" range. The median gain between the pre- and post-tests was a modest 2, and the mean gain was approximately 1.81 (n=457, since not everyone completed both the pre- and post-tests). Parkinson and Cutts [36], who also provided training using Sorby's workbook, performed the training in-person and required their students to complete all of the exercises. If we focus on students who completed all of the training, the mean gain increases slightly (to 2.53, n=327) but the median gain (2) remains the same.

Next, we focus on students who under-performed on the pre-test. Sorby considers students who score above 21 on the PSVT:R to not need training. Considering only those students who scored 21 or below on the pre-test and who completed all of the exercises, the median gain increases to 4 and the mean gain increases to 4.34 (n=177). This data corroborates prior work that has demonstrated that spatial skills training in the context of CS1 can increase spatial skills performance [6, 35] – in this case in an online environment – and it suggests that students with less-developed initial spatial skills ability gain more from training.

	Grade	SCS1 (pre)	Self-Efficacy (first week)	Self-Efficacy (last week)	Spatial Skills (pre)	Spatial Skills (post)
Grade n = 624 median (IQR) = 2.3 (2.6)	1.00 (NA) n = 624	-	-	-	-	-
SCS1 (pre-course) n = 628 median (IQR) = 7 (6)	0.45 (< 0.001) n = 558	1.00 (NA) n = 628	-	-	-	-
Self-Efficacy (first week) n = 707 median (IQR) = 4.04 (2.71)	0.24 (< 0.001) n = 622	0.50 (< 0.001) n = 628	1.00 (NA) n = 707	-	-	-
Self-Efficacy (last week) n = 498 median (IQR) = 5.56 (1.04)	0.39 (< 0.001) n = 491	0.33 (< 0.001) n = 448	0.30 (< 0.001) n = 497	1.00 (NA) n = 498	-	-
Spatial Skills (pre-test) n = 630 median (IQR) = 19 (10)	0.27 (< 0.001) n = 545	0.26 (< 0.001) n = 569	0.12 (0.002) n = 628	0.12 (0.010) n = 430	1.00 (NA) n = 630	-
Spatial Skills (post-test) n = 537 median (IQR) = 23 (10)	0.26 (< 0.001) n = 530	0.25 (< 0.001) n = 482	0.07 (0.108) n = 536	0.19 (< 0.001) n = 472	0.64 (< 0.001) n = 457	1.00 (NA) n = 537

Table 1: Matrix of pairwise Spearman's correlations between the spatial skills and self-efficacy (pre- and post-) tests, the SCS1, and student grades. The top number in each cell is the correlation (r), and the numbers below are the p-value and sample size (n).

Prior work that demonstrated the effectiveness of training also evaluated the impact of training on performance in a computing

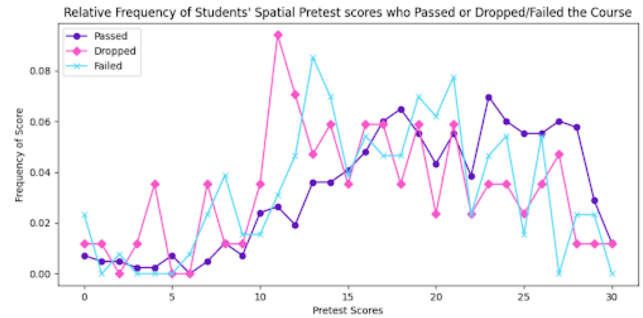

Figure 1: Line graph of pre-test spatial skills scores, separated into those who passed, failed, and did not complete the course. Students with lower initial spatial skills scores dropped the course at a higher rate.

course. [6, 36]. As we provided spatial skills training to all students, we cannot assess the impact of training on performance in the course, but we can investigate if there is a relationship between spatial skills ability and course performance. We use a Spearman's rank correlation to investigate the relationship between the assessed spatial skills scores and course performance measured by numeric grade in the course (on a 4.0 scale).

The results of these correlations (as well as correlations with other factors discussed later) are displayed in Table 1. Both the pre- and post-test scores on the spatial skills assessment are weakly correlated with final course grade (r=0.26-0.27, p<0.001).

The course grade measure ignores students who chose not to complete the course. Figure 1 displays the pre-test spatial skills scores for students who passed the course (n=416, median=20, IQR=9) versus those who failed (n=129, median=17, IQR=8) or dropped (n=85, median=16, IQR=10). Wilcoxon rank sum tests suggest that the distributions of students who drop and fail are not statistically different (p=0.282), but the distribution of those who pass is different from both those who fail (p<0.001) as well as those who drop (p<0.001).

Finally, we compared the distribution of spatial skills scores for both international and domestic students, as defined by the location of the residence of their parent(s). Domestic students scored a median of 19 (IQR=10, n=393) on the pre-test and a median of 22 (IQR=10, n=358) on the post-test. International students scored a median of 20 (IQR=9, n=162) on the pre-test and a median of 23 (IQR=9, n=141) on the post-test. A Wilcoxon rank-sum was used to compare the pre- and post- test distributions. For the pre-test distributions, no evidence was found for international students having different spatial abilities compared to domestic students (p=0.179). However, for the post-test distributions, there is marginal evidence of a difference between the two groups (p=0.050).

4.2 Prior Experience

We used an established language-independent measure of CS1 knowledge, the SCS1 [32], as a proxy for previous experience. The SCS1 was delivered as an online assignment in the first week of the term. The SCS1 consists of 27 items. The mean and median scores were 8.23 (SD=5.38) and 7 (IQR=6), respectively.

Table 1 displays correlations between the SCS1, course performance, and spatial skills assessments. The moderate correlation (r=0.447, p<0.001) between the SCS1 and course grade indicates that students with prior experience in the course have an advantage, as expected. There is also a weak correlation between the SCS1 and both the pre- (r=0.264, p<0.001) and post- (r=0.253, p<0.001) assessments of spatial skills. This suggests that students with greater prior experience tend to begin with – and to maintain – stronger spatial skills ability.

4.3 Socioeconomic Factors

We asked for both an estimate of household income and for the parent's education levels, as both represent facets of socioeconomic status. The response rate to the question about household income was low (only 230 respondents compared to 720 for other questions on the survey). This was not unexpected [43], since students may be either uncomfortable responding to a direct question about income and unsure of their family's income. We checked for correlations with grades (r=0.08, p=0.225), the spatial skills pre-test (r=0.00, p=0.610) and post-test (r=0.02, p=0.465), the SCS1 (r=-0.01, p=0.883), and self-efficacy pre-test (r=0.05, p=0.410) and post-test (r=0.18, p=0.018). Unfortunately, the lack of correlation with final grade, which has been identified in many other contexts [43], suggests that the data collected is not representative of the full population.

Response rates to the questions about parental education (n=481) were higher but still below the response rates for other questions, suggesting that students remain uncomfortable answering these questions or simply are not aware of their parents' educational achievements. As before, we explored correlations with grades, spatial skills, the SCS1, and self-efficacy and found negligible evidence. There was only a weak, negative correlation (r=-0.12, p=0.016) between a student's post-test spatial skills score and the educational attainment of their guardians. As before, the lack of correlation with final grade suggests that the data collected is not representative of the full population.

4.4 Self-Efficacy

Self-efficacy was evaluated in the first and last weeks of the semester. The median self-efficacy score in the first week was 4.04 of 7 (IQR=2.71), and was 5.56 (IQR=1.04) in the last week. As expected, the average self-efficacy increased over the term, as students encountered successes learning the material, and the IQR decreased, as prior experience matters less after a term of shared experiences.

Table 1 contains data on the correlations between the self-efficacy instrument deployments, grades, and spatial skills. There is only a weak (r=0.30, p<0.001) correlation between the self-efficacy of students measured in the first (pre-) and final (post-) weeks. This is expected: self-efficacy develops over time, as a result of mastery experiences, but it will not necessarily increase in a consistent manner for all students. Final week self-efficacy is moderately (r=0.39, p<0.001) correlated with the final grade in the course, while first week self-efficacy has a weaker correlation (r=0.24, p<0.001). Both of these results are also expected, as student self-efficacy at the end of the term, after several weeks of relevant experiences, should be more related to actual ability.

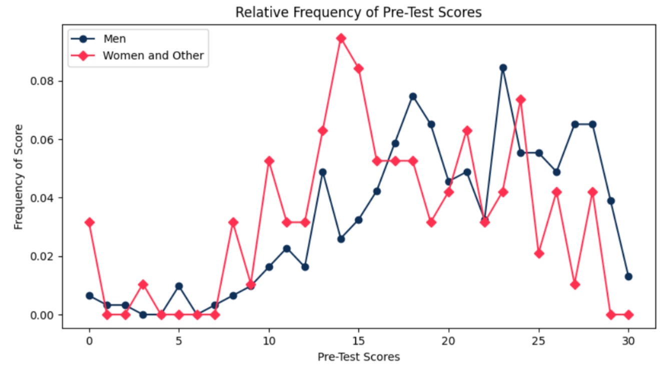

Figure 2: Line graph of pre-test scores by gender. The distribution of scores for women is lower (shifted to the left) relative to the distribution for men.

Finally, we examined the relationship between self-efficacy and spatial skills ability. The relationship between the pre-tests of self-efficacy and spatial skills is very weak (r=0.12, p<0.001). The relationship between the post-tests of self-efficacy and spatial skills is slightly stronger (r=0.19, p<0.001) but remains negligible.

4.5 Gender

We asked students to self-identify their gender. As noted in Section 3.1, a small number of students reported a gender other than man or woman. We chose to include these students with women, as there were too few to analyze separately and as we felt they would, like the women in the course, be experiencing effects of under-representation.

We observed a difference between the performance of men and women on both the pre- and post- tests of spatial skills. Considering only the students who completed the pre-test and a post-test, we find that men (n=307) scored a mean of 20.27 (SD=6.07) and median of 21 (IQR=8) on the pre-test. Women (n=95) scored an average of 17.01 (SD=6.32) with a median of 17 (IQR=9). Figure 2 shows the distribution of scores for men and women on the pre-test. Since the distribution of pre-test scores is left skewed, we used the Wilcoxon rank-sum two-sided test (p<0.001) and obtained a 95% confidence interval of (2, 5), suggesting that the average difference between men's and women's pre-test scores is between 2 and 5 points.

After training, the difference in spatial skills scores declines but does not disappear. Figure 3 shows the distributions of scores for men and women, again only considering students who completed both the pre-test and a post-test. The distributions are more similar, but the distribution of women's scores remains lower. Men (n=307) scored an average of 21.91 (SD=6.54) on the post-test, with a median of 23 (IQR=9). Women (n=95) scored a a mean of 19.93 (SD=6.84), with a median of 21 (IQR=9.5). Using the Wilcoxon rank-sum two sided test (p=0.007), the 95% confidence interval for the difference between the two average scores is (1, 3), suggesting that the difference in scores has decreased but has not been eliminated.

The counts of men and women in the analysis above are low because we considered only students who completed both the pre-test and post-test. 173 students completed the pre-test but did not

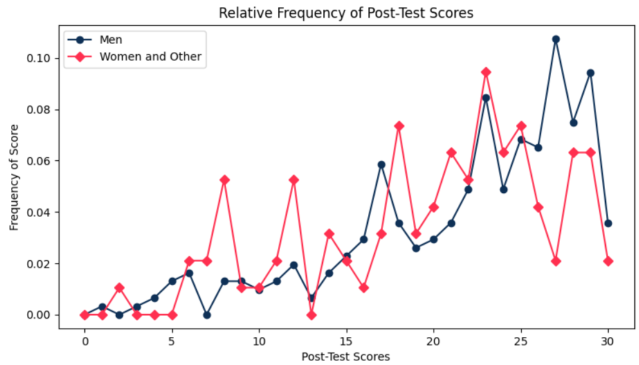

Figure 3: Line graph of post-test scores by gender. The distribution of scores for women is more similar to that of the men in comparison to the pre-test.

complete the post-test. If we consider all students who completed the spatial skills pre-test, then men (n=390) scored a mean of 19.97 (SD=6.23) with a median of 21 on the pre-test. The mean declined by 0.3 and the median did not change. Women (n=127) scored a mean of 16.09 (SD=6.41) with a median of 15 – a drop of 0.92 in the mean and 2 on the median. This larger shift in the mean and median suggests that women with low initial spatial skills are at higher risk of dropping the course.

To investigate this result, we examined the distribution of final grades and drop rates for both men and women. While a number of women perform very well, a larger proportion dropped the course than men. Specifically, 12% (n = 156) of women drop the course compared to only 8% (n = 459) of men. Furthermore, 26% of women fail the course, compared to 20% of men.

These trends are reflected in the final grades in the course. On average, men obtain a mean final grade of 2.14 on a 4 point scale (SD=1.4) with a median grade of 2.7 (IQR=2.3). Women had a mean grade of 1.87 (SD=1.4), with a median of 2.3 (IQR=3.0). A Wilcoxon rank-sum two sided test (p=0.042) yields a 95% confidence interval of the difference between the two average final grades of (0, 0.4). The p-value suggests that the evidence of a difference is marginal, with men appearing to continue to hold a slight advantage.

5 DISCUSSION

5.1 RQ1: Spatial Skills Training

The data presented in Section 4.1 suggests that training spatial skills using online resources is effective. We saw a median 2 point (out of 30) increase in spatial skills across all students who completed both the pre- and a post- assessment. However, we saw a significantly larger 4 point median gain when considering only the students the Sorby believed would benefit from training. These gains are comparable to those observed in in-person spatial training reported previously [6, 36]. The latter observation also suggests that spatial skills training is particularly important for students who enter the course with less preparation.

Based on Figure 1 and the drop rate data we observed, we believe there is a need to accelerate the training or to provide it prior to the course. As discussed in Section 4.5, including the students who did not complete the course lowered the average spatial pre-test scores, indicating that students who dropped the course are more likely to be facing a spatial skills deficit compared to their peers who remain in the course. In our case, spatial skills training took place over the first two-thirds of the course. If spatial skills do have an impact on computing outcomes, those skills may have been developed too late to prevent students with a deficit from falling behind.

Efforts to accelerate the training, however, may be challenging due to its perceived difficulty. After each spatial skills assessment and training module, the students were provided with an option to provide feedback and to report any issues they encountered. In addition to the weekly feedback, the students were asked three relevant questions in the final course survey. The first question asked how often they guessed the answer to exercises in the spatial skills training. The second question asked how often they directly copied answers to exercises in the spatial skills training. The third question provided a final opportunity to provide feedback.

The free response feedback indicates that students found the spatial skills content challenging. Feedback in the early weeks suggested that students were spending more time (2-3 hours per week) on training than we expected given what we had been told about in-person training (1-2 hours per week), so we reduced the number of modules per week to deliver the training in eight, rather than the original six, weeks. Students also voiced concerns that they did not see how spatial skills helped them program. We responded with a summary of the research on spatial skills and reiterated a message from earlier in the term that we believed that the training would help not only in computing but also in other fields. This explanation reduced the number of concerns voiced in the weekly feedback, but in the final survey, 20 students (of 478 respondents) explicitly noted that they disliked or hated the spatial skills training, while only 4 explicitly noted that they enjoyed it. This feedback suggests that instructors will need to be particularly careful to provide support resources and to explain (and reinforce) the reason why spatial skills exercises are included in the course.

Despite the difficulty, students reported that they engaged honestly with the material. Table 2 shows the responses to two questions about guessing and copying that were on the final survey. This is self-reported data, so students may not have been comfortable disclosing this behaviour. As a result, the distributions may under-report guessing and copying behaviour. However, students were given the opportunity to not respond, and relatively few exercised this option. The results support the idea that the students found the spatial skills exercises challenging, with the majority reporting that they needed to guess at least "sometimes", and 18% of the class reporting that they did so often. Very few reported copying answers, however, suggesting that the majority attempted the training.

5.2 RQ2: Spatial Skills Training and Gender

In Section 4.5, we observed that women entered the course with lower spatial skills than men and that they completed the course at a lower rate than men. While the reasons for a spatial skills deficit are debated, it has been documented in other fields [4, 27]. Our observations confirm that this effect is present in computing, as it is in other STEM disciplines.

	Almost Always	Often	Sometimes	Rarely	Never	Do Not Wish to Respond
On spatial skills exercises, how often did you guess answers?	24 (4%)	68 (14%)	175 (36%)	109 (29%)	91 (18%)	17 (3%)
On spatial skills exercises, how often did you copy answers?	4 (0%)	4 (0%)	14 (2%)	21 (4%)	425 (87%)	20 (4%)

Table 2: Responses to the the end-of-term survey. Many students reported needing to guess but very few reported copying.

Fortunately, researchers in other fields have also reported that the skills deficit can also be addressed through training [24, 52], and we observed a similar trend. When we analysed the individual differences between pre- and post- test scores for both men and women, women improved an average of 2.92 points (of 30) between the two tests, in contrast to an average gain of only 1.64 for men. This result aligns with our previous analysis showing that students with lower starting spatial skills scores improved more, and it emphasizes that the narrowing gap in spatial skills we saw in the post-test is not driven by students dropping the course.

Jones and Burnett have argued that incorporating spatial skills into a curriculum is particularly important for supporting women [21]. Our results support their argument. We have demonstrated evidence that many women enter the course with a disadvantage in their spatial skills preparation and that training in the course can reduce – though not, in our case, totally eliminate – that gap.

Unfortunately, due to the demographics of our course, we were unable to collect evidence to evaluate the impact on other under-represented groups. This should be the focus of future work.

5.3 RQ3: Spatial Skills and Other Factors

We observed evidence suggesting that spatial skills (measured in both the pre- and post- tests) are weakly correlated with both prior experience with computing (as estimated by the SCS1) and demonstrated ability in the course (as estimated by final grades). We also saw a very weak relationship between spatial skills and self-efficacy, but did not obtain any evidence of a relationship between spatial skills and factors related to socioeconomic status (SES).

Previous Experience and Self-Efficacy: The relationship between spatial skills and computing experiences is expected, given prior work that has demonstrated relationships between spatial skills ability and performance in computing courses [34, 54] and achievement in validated CS1 assessment tools [7, 13]. Given the moderate correlation between self-efficacy and performance, the relationship between spatial skills and self-efficacy might be viewed as a residual relationship. We suggest that mastery experiences related to spatial skills do not naturally contribute to a student's self-efficacy in computing. The students in our context did not perceive these experiences as contributing to their computing ability. This result contrasts with increases in confidence observed by Cooper et al. [13]. It is possible that the students in that study may have been able to compare their computing-related mastery experiences with peers who did not receive spatial skills training. It's also possible that the measurements of confidence in Cooper et al.'s study and self-efficacy in ours are not compatible. Further work will be required to further investigate this question, but we believe our observations at least emphasize the need to explain and motivate spatial skills training for students.

Socioeconomic Status: The lack of any evidence of a relationship between spatial skills and factors related to socioeconomic status (SES) was unexpected, given the prior work by Parker et al. [33] that proposed that spatial skills might act as a mediator between SES and computing outcomes. However, we also failed to see expected links between SES and performance in the course, which casts doubt on the quality of our SES data. Collecting SES data is difficult, and the various measures used to evaluate it can be unreliable [42], so we believe it is more likely that the link observed by Parker et al. exists and that additional measurement is required. parental educational level and household income, which we attempted to measure, are commonly used measures of SES, but we observed lower response rates to both of these questions. In the future, we will explore other factors, such as parental occupation [19] or the family affluence scale [26].

5.4 Threats to Validity

During this study, a number of threats were identified related to the global environment and the deployment of the intervention and data collection instruments. The first and most obvious issue is the COVID-19 pandemic. Originally, we planned to compare learning outcomes and course completion rates across years to provide additional evidence for questions about the impact of spatial skills training on learning outcomes in computing courses. However, with COVID-19 forcing changes to delivery and, in general, increasing the stress on students, this comparison has had to be postponed. While we believe the course grades reported are as valid a measure of student performance in the course as they are in "normal" terms, the distribution of grades is not comparable to other years. In particular, we believe that a larger number of students did not complete the course, in comparison to prior years, due to stresses imposed by COVID-19 remote learning. We anticipate collecting further data to address this question in the future, as we and the course instructors are convinced that spatial skills training has the potential to have an out-sized impact on under-represented communities.

The second issue relates to the deployment of the spatial skills post-test. Participation in the post-test was low, so a second opportunity to complete the test was provided. Some students chose to do the test twice, despite being told that they only needed to complete it once. We decided to use the first complete post-test score per student for our analysis.

Finally, we must be careful about drawing conclusions that are too strong from a single study. This work explicitly aimed to confirm and then extend prior work in the area, to build up a corpus of literature from multiple contexts to allow for stronger conclusions to be drawn by a later meta-analysis. While we did observe trends that align with prior work – in particular, in the relationship between spatial skills and academic achievement, and in the larger impact of spatial skills training on women – we were unable to confirm links between socioeconomic status and spatial skills that have been reported in prior work. This study does not invalidate that work, as noted earlier, but it does highlight the need for replication in computing education [1].

6 CONCLUSIONS

We deployed an online spatial skills training intervention in a CS1 context, evaluated its impact, and collected data on potentially related factors that are linked to academic performance in computing at the post-secondary level. We found strong evidence that online training is effective at improving spatial skills, and is particularly effective for students with a spatial skills deficit identified by the initial spatial skills assessment. This result is particularly relevant to our efforts to decrease the under-representation of women in computing, as we observed that women entered the course with deficits in spatial skills at higher rates than men.

We also documented links between spatial skills factors linked to academic performance in computing, such as self-efficacy and prior experience with CS1 content. Our observations suggest that while students with more computing experience tend to exhibit higher spatial skills, students will not naturally connect spatial skills training to their computing abilities, so instructors deploying such interventions need to explain and motivate that link.

We were unable to gather evidence to determine if increasing spatial skills with training also increased performance in the course, but we have concerns, based on observed drop rates, that the training may have been implemented too late in the term to be fully effective. We anticipate collecting further data in future terms to determine if the spatial skills intervention is effective at improving course performance in our context, and we encourage those deploying spatial skills interventions to do so as early as possible.

We also encourage other instructors and researchers to consider deploying a spatial skills assessment in their context. Additional evidence related to our findings about the spatial skills - gender link in computing is needed, and this link should also be explored in other under-represented groups. With the mounting evidence in computing that improved spatial skills are related to improved learning outcomes, it is particularly important to deploy and study interventions, like spatial skills training, that have the potential to support the retention of students in under-represented groups.

ACKNOWLEDGMENTS

We gratefully acknowledge the constructive criticism offered by the reviewing community of Koli Calling. Their support and suggestions for improvement have significantly improved our own understanding as well as the quality of this work.

A DEMOGRAPHIC SURVEY

Please answer the following questions to the best of your ability. If you do not wish to answer any question, you may skip it.

(1) What [region] do your parent(s) reside in? If your parent(s) live outside of [country], please provide their country of residence.
(2) Do you receive [government financial aid program]? (Yes, No, Unsure)
(3) What is your household annual income (estimated)? (5 bins, "I do not know", "I prefer not to answer.")
(4) What is the highest level of education completed by one of your parents? (Primary or none, Secondary, Certificate or 2-year (College) Degree, 4-year (University) Degree, Post-graduate Degree)
(5) What is the highest level of education completed by another of your parents? (Primary or none, Secondary, Certificate or 2-year (College) Degree, 4-year (University) Degree, Post-graduate Degree)
(6) What language(s) were spoken in your childhood home?
(7) If you are currently employed, how many hours do you work in a typical week? (4 bins)
(8) What is your self-identified gender? Please leave this blank if you prefer not to answer.

REFERENCES

[1] Alireza Ahadi, Arto Hellas, Petri Ihantola, Ari Korhonen, and Andrew Petersen. 2016. Replication in Computing Education Research: Researcher Attitudes and Experiences. In *Proceedings of the 16th Koli Calling International Conference on Computing Education Research* (Koli, Finland). 2–11.
[2] Maryann Baenninger and Nora Newcombe. 1989. The role of experience in spatial test performance: A meta-analysis. *Sex roles* 20, 5-6 (1989), 327–344.
[3] Jens Bennedsen and Michael E Caspersen. 2006. Abstraction ability as an indicator of success for learning object-oriented programming? *ACM Sigcse Bulletin* 38, 2 (2006), 39–43.
[4] Vanessa Bentley, Annelies Kleinherenbrink, Gina Rippon, Diana Schellenberg, and Sigrid Schmitz. 2019. Improving practices for investigating spatial "stuff": Part I: Critical gender perspectives on current research Practices. *The Scholar and Feminist Online* 15, 2 (2019).
[5] Ryan Bockmon, Stephen Cooper, Jonathan Gratch, and Mohsen Dorodchi. 2019. (Re)Validating Cognitive Introductory Computing Instruments. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*. Association for Computing Machinery, New York, NY, USA, 552–557.
[6] Ryan Bockmon, Stephen Cooper, Jonathan Gratch, Jian Zhang, and Mohsen Dorodchi. 2020. Can Students' Spatial Skills Predict Their Programming Abilities?. In *Proceedings of the 2020 ACM Conference on Innovation and Technology in Computer Science Education* (Trondheim, Norway) *(ITiCSE '20)*. Association for Computing Machinery, New York, NY, USA, 446–451.
[7] Ryan Bockmon, Stephen Cooper, William Koperski, Jonathan Gratch, Sheryl Sorby, and Mohsen Dorodchi. 2020. A CS1 Spatial Skills Intervention and the Impact on Introductory Programming Abilities. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education* (Portland, OR, USA) *(SIGCSE '20)*. Association for Computing Machinery, New York, NY, USA, 766–772.
[8] Mimi Bong and Einar M Skaalvik. 2003. Academic self-concept and self-efficacy: How different are they really? *Educational psychology review* 15, 1 (2003), 1–40.
[9] Andrea Bosco, Anna M Longoni, and Tomaso Vecchi. 2004. Gender effects in spatial orientation: Cognitive profiles and mental strategies. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition* 18, 5 (2004), 519–532.
[10] Jennifer Campbell, Diane Horton, and Michelle Craig. 2016. Factors for success in online CS1. In *Proceedings of the 2016 ACM conference on innovation and technology in computer science education*. 320–325.
[11] Jennifer Campbell, Diane Horton, Michelle Craig, and Paul Gries. 2014. Evaluating an inverted CS1. In *Proceedings of the 45th ACM technical symposium on Computer science education*. 307–312.
[12] John B Carroll et al. 1993. *Human cognitive abilities: A survey of factor-analytic studies*. Number 1. Cambridge University Press.
[13] 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*. 13–20.
[14] Quintin Cutts, Emily Cutts, Stephen Draper, Patrick O'Donnell, and Peter Saffrey. 2010. Manipulating mindset to positively influence introductory programming performance. In *Proceedings of the 41st ACM technical symposium on Computer science education*. 431–435.
[15] Quintin Cutts, Patricia Haden, Ken Sutton, Ilona Box, John Hamer, Raymond Lister, Denise Tolhurst, Sally Fincher, Anthony Robins, Bob Baker, et al. 2006. The ability to articulate strategy as a predictor of programming skill. In *Conferences in Research and Practice in Information Technology Series*.
[16] Quintin Cutts, Elizabeth Patitsas, Elizabeth Cole, Peter Donaldson, Bedour Alshaigy, Mirela Gutica, Arto Hellas, Edurne Larraza-Mendiluze, Robert McCartney,

and Charles Riedesel. 2018. Early Developmental Activities and Computing Proficiency. In *Proceedings of the 2017 ITiCSE Conference on Working Group Reports* (Bologna, Italy) *(ITiCSE-WGR '17)*. Association for Computing Machinery, New York, NY, USA, 140–157.

[17] Johan Ferla, Martin Valcke, and Yonghong Cai. 2009. Academic self-efficacy and academic self-concept: Reconsidering structural relationships. *Learning and individual differences* 19, 4 (2009), 499–505.

[18] Maryanne L Fisher, Tami Meredith, and Melissa Gray. 2018. Sex differences in mental rotation ability are a consequence of procedure and artificiality of stimuli. *Evolutionary Psychological Science* 4, 2 (2018), 124–133.

[19] Bruna Galobardes, Mary Shaw, Debbie A Lawlor, John W Lynch, and George Davey Smith. 2006. Indicators of socioeconomic position (part 1). *Journal of Epidemiology & Community Health* 60, 1 (2006), 7–12.

[20] Arto Hellas, Petri Ihantola, Andrew Petersen, Vangel V. Ajanovski, Mirela Gutica, Timo Hynninen, Antti Knutas, Juho Leinonen, Chris Messom, and Soohyun Nam Liao. 2018. Predicting Academic Performance: A Systematic Literature Review. In *Proceedings Companion of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education* (Larnaca, Cyprus) *(ITiCSE 2018 Companion)*. Association for Computing Machinery, New York, NY, USA, 175–199.

[21] S Jones and G Burnett. 2006. Give the girls a chance: should spatial skills training be incorporated into the curriculum. *The Internet Society II: Advances in education, commerce & governance* (2006).

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

[23] Sue Jane Jones and Gary E Burnett. 2007. Spatial skills and navigation of source code. *ACM SIGCSE Bulletin* 39, 3 (2007), 231–235.

[24] Steven J Kass, Robert H Ahlers, and Melissa Dugger. 1998. Eliminating gender differences through practice in an applied visual spatial task. *Human Performance* 11, 4 (1998), 337–349.

[25] Lynn S Liben. 2015. The STEM gender gap: The case for spatial interventions. *International Journal of Gender, Science and Technology* 7, 2 (2015), 133–150.

[26] Yu-Chen Lin. 2011. Assessing the use of the Family Affluence Scale as socioeconomic indicators for researching health inequalities in Taiwan adolescents. *Social Indicators Research* 102, 3 (2011), 463–475.

[27] Marcia C Linn and Anne C Petersen. 1985. Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child development* (1985), 1479–1498.

[28] Alex Lishinski, Aman Yadav, Jon Good, and Richard Enbody. 2016. Learning to Program: Gender Differences and Interactive Effects of Students' Motivation, Goals, and Self-Efficacy on Performance. In *Proceedings of the 2016 ACM Conference on International Computing Education Research* (Melbourne, VIC, Australia) *(ICER '16)*. Association for Computing Machinery, New York, NY, USA, 211–220.

[29] David F Lohman. 1988. Spatial abilities as traits, processes, and knowledge. (1988).

[30] Lauren E. Margulieux. 2020. Spatial Encoding Strategy Theory: The Relationship between Spatial Skill and STEM Achievement. *ACM Inroads* 11, 1 (Feb. 2020), 65–75.

[31] Marialaura Moschella and Demis Basso. 2020. Computational Thinking, spatial and logical skills. An investigation at primary school. *Ricerche di Pedagogia e Didattica. Journal of Theories and Research in Education* 15, 2 (2020), 69–89.

[32] 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* (Melbourne, VIC, Australia) *(ICER '16)*. Association for Computing Machinery, New York, NY, USA, 93–101.

[33] Miranda C. Parker, Amber Solomon, Brianna Pritchett, David A. Illingworth, Lauren E. Marguilieux, and Mark Guzdial. 2018. Socioeconomic Status and Computer Science Achievement: Spatial Ability as a Mediating Variable in a Novel Model of Understanding. In *Proceedings of the 2018 ACM Conference on International Computing Education Research* (Espoo, Finland) *(ICER '18)*. Association for Computing Machinery, New York, NY, USA, 97–105.

[34] Jack Parkinson and Quintin Cutts. 2018. Investigating the Relationship Between Spatial Skills and Computer Science. In *Proceedings of the 2018 ACM Conference on International Computing Education Research* (Espoo, Finland) *(ICER '18)*. Association for Computing Machinery, New York, NY, USA, 106–114.

[35] Jack Parkinson and Quintin Cutts. 2020. The Effect of a Spatial Skills Training Course in Introductory Computing. In *Proceedings of the 2020 ACM Conference on Innovation and Technology in Computer Science Education* (Trondheim, Norway) *(ITiCSE '20)*. Association for Computing Machinery, New York, NY, USA, 439–445.

[36] Jack Parkinson, Quintin Cutts, and Steve Draper. 2020. Relating Spatial Skills and Expression Evaluation. In *United Kingdom & Ireland Computing Education Research conference.* 17–23.

[37] Claudia Quaiser-Pohl and Wolfgang Lehmann. 2002. Girls' spatial abilities: Charting the contributions of experiences and attitudes in different academic groups. *British Journal of Educational Psychology* 72, 2 (2002), 245–260.

[38] Michelle Richardson, Charles Abraham, and Rod Bond. 2012. Psychological correlates of university students' academic performance: a systematic review and meta-analysis. *Psychological bulletin* 138, 2 (2012), 353.

[39] Nathan Rountree, Janet Rountree, Anthony Robins, and Robert Hannah. 2004. Interacting Factors That Predict Success and Failure in a CS1 Course. *SIGCSE Bull.* 36, 4 (June 2004), 101–104.

[40] Linda J Sax, Kathleen J Lehman, Jerry A Jacobs, M Allison Kanny, Gloria Lim, Laura Monje-Paulson, and Hilary B Zimmerman. 2017. Anatomy of an enduring gender gap: The evolution of women's participation in computer science. *The Journal of Higher Education* 88, 2 (2017), 258–293.

[41] Linda J Sax, Hilary B Zimmerman, Jennifer M Blaney, Brit Toven-Lindsey, and Kathleen Lehman. 2017. Diversifying undergraduate computer science: The role of department chairs in promoting gender and racial diversity. *Journal of Women and Minorities in Science and Engineering* 23, 2 (2017).

[42] Vickie L Shavers. 2007. Measurement of socioeconomic status in health disparities research. *Journal of the national medical association* 99, 9 (2007), 1013.

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

[44] Sheryl Sorby, Beth Casey, Norma Veurink, and Alana Dulaney. 2013. The role of spatial training in improving spatial and calculus performance in engineering students. *Learning and Individual Differences* 26 (2013), 20–29.

[45] Sheryl Sorby, Norma Veurink, and Scott Streiner. 2018. Does spatial skills instruction improve STEM outcomes? The answer is 'yes'. *Learning and Individual Differences* 67 (2018), 209–222.

[46] 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.

[47] Sheryl A Sorby and Beverly J Baartmans. 1996. A Course for the Development of 3-D Spatial Visualization Skills. *Engineering Design Graphics Journal* 60, 1 (1996), 13–20.

[48] Sheryl Ann Sorby, Anne Frances Wysocki, and Beverly Gimmestad Baartmans. 2003. *Introduction to 3D Spatial Visualization: an active approach.* Cengage Learning.

[49] Phil Steinhorst, Andrew Petersen, and Jan Vahrenhold. 2020. Revisiting Self-Efficacy in Introductory Programming. In *Proceedings of the 2020 ACM Conference on International Computing Education Research* (Virtual Event, New Zealand) *(ICER '20)*. Association for Computing Machinery, New York, NY, USA, 158–169.

[50] Donald E Super and Paul B Bachrach. 1957. Scientific careers and vocational development theory: A review, a critique and some recommendations. (1957).

[51] Anya Tafliovich, Jennifer Campbell, and Andrew Petersen. 2013. A student perspective on prior experience in CS1. In *Proceeding of the 44th ACM technical symposium on Computer science education.* 239–244.

[52] 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.

[53] Daniel Voyer, Carla Nolan, and Susan Voyer. 2000. The relation between experience and spatial performance in men and women. *Sex Roles* 43, 11 (2000), 891–915.

[54] 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.

[55] Christopher Watson, Frederick W.B. Li, and Jamie L. Godwin. 2014. No Tests Required: Comparing Traditional and Dynamic Predictors of Programming Success. In *Proceedings of the 45th ACM Technical Symposium on Computer Science Education* (Atlanta, Georgia, USA) *(SIGCSE '14)*. Association for Computing Machinery, New York, NY, USA, 469–474.

[56] Brenda Cantwell Wilson. 2002. A study of factors promoting success in computer science including gender differences. *Computer Science Education* 12, 1-2 (2002), 141–164.

[57] Brenda Cantwell Wilson and Sharon Shrock. 2001. Contributing to Success in an Introductory Computer Science Course: A Study of Twelve Factors. *SIGCSE Bull.* 33, 1 (Feb. 2001), 184–188.

[58] Benjamin Xie, Matthew J. Davidson, Min Li, and Amy J. Ko. 2019. An Item Response Theory Evaluation of a Language-Independent CS1 Knowledge Assessment. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education* (Minneapolis, MN, USA) *(SIGCSE '19)*. Association for Computing Machinery, New York, NY, USA, 699–705.

[59] So Yoon Yoon. 2011. *Psychometric properties of the revised purdue spatial visualization tests: visualization of rotations (The Revised PSVT: R).* Purdue University.

[60] Larry Yueli Zhang, Andrew K. Petersen, Michael Liut, Bogdan Simion, and Furkan Alaca. 2021. A Multi-Course Report on the Experience of Unplanned Online Exams. In *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education* (Virtual Event, USA) *(SIGCSE '21)*. Association for Computing Machinery, New York, NY, USA, 17–23.