Exploring Models and Theories of Spatial Skills in CS through a Multi-National Study

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

Sebastian Dziallas
sdziallas@pacific.edu
University of the Pacific
Stockton, CA, United States

Fiona McNeill
f.j.mcneill@ed.ac.uk
University of Edinburgh
Edinburgh, Scotland, UK

Jim Williams
jimw@cs.wisc.edu
University of Wisconsin
Madison, WI, United States

ABSTRACT

Background and Context. The relationship between spatial skills and computing science success has been demonstrated at multiple institutions. ICER has reacted positively to two theories for why this relationship exists, by both Parkinson & Cutts and Margulieux. However, only limited work has been done to validate these theories, and more confirmatory research about the relationship between spatial skills and module grades in CS is necessary.

Objectives. We wish to validate two dimensions of existing theories for the relationship between spatial skills and CS: does CS learning improve spatial skills (as has been observed in other domains, such as physics) as Parkinson & Cutts propose, and does the relationship with grades predominantly apply to students with less prior programming fluency when they begin their learning, as Margulieux proposes. We also wish to contribute more data to the existing set of correlations between spatial skills and measures of CS success (replication).

Method. We conducted a multi-institutional, multi-national project to capture prior programming experience and module grades in CS at three institutions, as well as conducting spatial skills tests at three points during the academic year. We compare spatial skills results with module grades, we examine changes in spatial skills over a period of CS learning and we explore whether the correlations between spatial skills and module grades apply for students at all levels of prior programming fluency.

Findings. We found that spatial skills correlated with module grades at each institution, spatial skills improved over the first semester of teaching (though not the second semester, and at different rates in different institutions) and students with lower self-reported prior programming fluency exhibited much stronger correlations between spatial skills and grades than students with greater programming fluency.

Implications. This work provides additional evidence that spatial skills are correlated with introductory CS outcomes. It also takes steps towards validating existing theories for the relationship by demonstrating that spatial skills can be trained through CS learning and students with lower levels of prior programming fluency are more likely to rely on spatial skills in their CS learning.

CCS CONCEPTS

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

KEYWORDS

spatial skills, introductory computing, cognition

1 INTRODUCTION

We are gradually beginning to understand more about the relationship between spatial skills and computing science (CS). The bottom line is that spatial skills appear to be valuable in computing, though often with caveats: different contexts and different cohorts sometimes don't see the same effects, and the effect sizes of correlations and other statistical tests are often different. Parkinson & Cutts began their work in this area by presenting a model for the relationship, suggesting that there is an underlying skill that ties spatial skills and all STEM learning together: by practising spatial skills, we may improve our STEM outcomes, but by learning STEM subjects we may also be able to develop our spatial skills [28].

This paper presents three major findings: the first bolsters research indicating that spatial skills are related to success in introductory CS. The second contribution is the discovery that spatial skills can be improved by learning introductory CS, as suspected by Parkinson & Cutts, though there are differences in how much these skills can be improved between institutions and over time. The third contribution is an analysis of how prior programming fluency affects this relationship when examining introductory programming cohorts, as this has been theorised as a confound in research.

2 BACKGROUND

2.1 What are Spatial Skills?

Spatial skills are cognitive skills associated with understanding and internalising spatial elements and concepts. Spatial skills include mental rotation, orientation, and manipulation [7]. There are multiple factors of spatial skills which can be tested independently, but are all related and connected [6]. Parkinson & Cutts provide a breakdown of the different factors of spatial skills in their 2018 work [28].

The factor of the most interest is *spatial visualisation*, which can be tested through mental rotation tasks. While it is possible that many factors of spatial skills may influence aptitude in a range of subjects, there is lots of evidence that spatial visualisation specifically is connected to success in STEM domains, including computing science. In most instances throughout this paper when we refer to “spatial skills”, the specific factor being measured is spatial visualisation.

2.2 Spatial Skills are Related to Computing Science Success

Some early work relating spatial skills and success in CS was conducted by Mayer, who discovered that performance in a paper folding task (a classic spatial visualisation test [17, 44]) correlated with beginners’ CS assessment scores after learning BASIC for the first time [23]. Work conducted by Fincher *et al.* in a large, multi-institutional project discovered that students’ ability to draw more complex forms of maps (closely connected to spatial orientation [32, 34], related to spatial visualisation [6]) was a predictor for success in an introductory CS module [11, 35, 42].

Jones & Burnett conducted two separate studies connecting spatial skills with computing science success. In their first study, they designed an activity to track the speed and, in a sense, “distance travelled” in source code navigation. Students with better spatial skills completed the tasks more quickly and made larger, more useful jumps around the code than their peers with lower spatial skills, demonstrating that people with better spatial skills are better at source code navigation [13]. The findings partially affirm predictions by Cox, Fisher & O’Brien, who theorised that spatial skills could be valuable for people navigating code as they would navigate real space, naming the internal, navigable representation of code a “codespace” [9]. Following this, Jones & Burnett examined the spatial skills of a master’s conversion cohort – that is, a cohort of CS master’s students who did not have a CS undergraduate degree – and discovered that spatial skills correlated with module grades in their programming modules, but not their non-programming modules [14].

Other correlations between spatial skills and measures of CS success have been observed since Jones & Burnett’s work. Cooper *et al.* discovered that there was a correlation between spatial skills and a standardised CS assessment taken at the end of a 2-week pre-college programming summer school [8]. Bockmon *et al.* discovered that there was a correlation between spatial skills and scores in a reduced set of questions from a standardised, language independent CS1 concept inventory [25], the SCS1R [2], across three institutions in the USA [4]. In a separate study across the same institutions, Bockmon *et al.* also found that of a range of predictive measures taken at the start of a semester – spatial skills, socioeconomic status, attitudes as measured by Dorn & Tew’s Computing Attitudes Survey [10] and prior CS experience as measured by the SCS1R – spatial skills had the highest individual accuracy in predicting students’ scores in a post-test of the SCS1R at the end of the semester [3]. This is similar to the previous findings of Parker *et al.*, who discovered through principal component analysis that spatial skills were a more powerful intervening variable for CS1 assessment performance than access to computers [26].

Parkinson & Cutts discovered that there was a correlation between spatial skills and exam scores in a CS0 cohort [29] and separately with a test of expression evaluation [31]. Ly *et al.* also observed correlations between spatial skills and final module grades in a CS1 cohort [19]. The correlation between spatial skills and various forms of CS assessment and success has, therefore, been explored in many areas. It is important to note, however, that there are some distinct differences in the correlations observed. From the studies mentioned, Jones & Burnett observed comparatively very high correlations with module marks (r=.62 and r=.71), while Parkinson & Cutts observed a medium to strong correlation for the CS0 students (r=.50 for their final exam). Bockmon *et al.* observed a smaller correlation (r=.41) and Ly *et al.*’s finding was even smaller (r=.25). Clearly, there is a difference in how much spatial skills play a role in different computing contexts and with different measures of success. While the relationship does appear to be reliably present, it is not reliably consistent, so the amount of attention it should be paid in the broader context of CS education and progression has yet to be clearly determined.

2.3 Why are Spatial Skills Valuable in Computing Science?

Marguliuex presented a theory for the relationship between spatial skills and success in STEM called Spatial Encoding Strategy theory (SpES) [22]. The theory states that we can use grid and place cells in the hippocampus – which we originally evolved for navigation – for encoding *all* non-verbal information and representations. Those with better spatial skills are capable of more efficient encoding of non-verbal information using these neuro-structures, and therefore are more likely to be able to hold complex and overlapping conceptual models and procedures in their heads. In CS, these may be models constructed on the fly, like a model of a program, a problem domain or a data/object relationship, or they may be more persistent models such as the syntax structure of a programming language or a complex computer system or process. Margulieux’s two broad examples when presenting the theory are “running mental models” and “building robust notional machines”.

Margulieux’s theory comes after the work of Parkinson & Cutts, who proposed that spatial skills *themselves* do not necessarily help with learning in computing, but rather they expose an abstract set of skills which they called an “underlying cognitive ability” (see figure 1) [28]. They theorised that as a result, spatial skills training could improve outcomes in CS and other STEM domains, which was already known to be true of engineering: Sheryl Sorby has been developing and applying spatial skills training exercises programmes for nearly 30 years with consistently positive outcomes

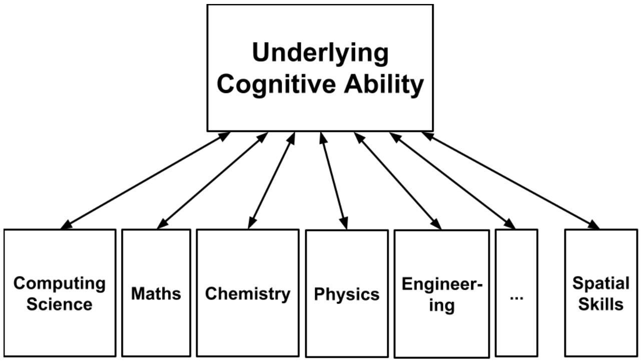

Figure 1: The model for the relationship between spatial skills and STEM subjects, presented by Parkinson and Cutts [28]

for engineering students [37, 38, 40]. Cooper *et al.*, Bockmon *et al.* and Parkinson & Cutts have each conducted Sorby's spatial skills training in introductory computing and have observed positive results for CS outcomes, demonstrating that Parkinson & Cutts' original theory is likely to be true. (Note that Ly *et al.* have also delivered a spatial skills training programme in a CS1 setting, but did not conduct appropriate pre- and post-testing with a control group to determine if the training itself was effective [19].)

However, Parkinson & Cutts' model also suggests that if this set of skills can be trained in other domains, we would see spatial skills increase over time spent in STEM study. The arrows go both ways: spatial skills training may lead to improvements in CS, but CS learning may also lead to improvements in spatial skills [28].

This was found to be true in college physics education. In a study by Pallrand and Seeber [24], students from four cohorts had their spatial skills tested before and after a period of different forms of teaching:

(1) Students studying introductory physics alongside a spatial skills intervention (experimental)
(2) Students studying introductory physics with lectures on the history of physics equal to the time of the spatial skills intervention (placebo)
(3) Students studying introductory physics with no additional instruction (control)
(4) Students studying liberal arts (eliminating test-retest effects)

Across several tests of spatial skills taken before and after the instruction period, the experimental group showed highly significant, large gains in spatial skills, the placebo and control groups both showed some significant improvement in spatial skills (not as large as the experimental group), and the liberal arts students showed very small, mostly insignificant gains in spatial skills. The authors comment on the placebo and control groups, saying that their gains "suggest that taking introductory physics may itself influence visual-spatial abilities of students" [24].

Therefore, we can see that physics learning can improve spatial skills, and we know from other work that having good spatial skills helps in learning physics: Kozhevnikov *et al.* observed that spatial skills contribute to solving kinematics problems in physics [15, 16], and Mac Raighne *et al.* observed a correlation between spatial visualisation and a Newtonian mechanics concept inventory [20]. Based on the models presented by Margulieux and Parkinson and Cutts, which apply to most STEM education, it is possible that the bi-directionality observed in physics education also exists in computing education.

2.4 The Role of Prior Experience

Since non-verbal encoding skills are associated with learning, Margulieux states in SpES that spatial skills are not likely to be useful beyond initial learning in STEM since students develop their own domain-specific strategies and patterns for solving problems. Verifying this, Parkinson & Cutts have repeatedly found evidence that students with more experience in computing appear to rely less on spatial skills to solve computing problems. In previously mentioned studies, while they observed a moderate correlation between spatial skills and module exams in CS0, the relationship was much weaker and did not reach significance for the CS1 cohort [29]; similarly, in their expression evaluation experiment, the correlation between their expression evaluation instrument and spatial skills was not significant for students with more experience [31].

However, Parkinson & Cutts have also found that the correlation between spatial skills and CS GPAs actually grows over time, indicating that those with better spatial skills increasingly do better in later years of study [30]. Additionally, the correlation between spatial skills and grades in computationally complex modules is still fairly high in the third and fourth years of study. Recently, Liu *et al.* also observed relationships between spatial skills and success in an introductory computer graphics module, which was studied by students in their third or fourth year of study: the students had at least two years of computing or software engineering learning experience but the domain was new [18].

Parkinson & Cutts explain their findings by considering the nature of learning and the role of prior experience. They hypothesise that *initial* learning in computing requires abstract encoding skills, but domain-specific strategies do develop, just as Margulieux theorises [30]. However, they do not assume that learning in CS is only an initial event: as a student progresses through a degree, they encounter new languages, new paradigms, and new models and processes of computation. Every time they encounter something new, they will revert back to their abstract skills to assimilate the information and develop their new strategies. With this theory, spatial skills are important for any individual intending on learning new computational content, whether it is introductory or advanced.

2.5 Replications and MIMN Studies

Recent work at ICER has highlighted a lack of replication studies in our field [5]. Spatial skills research in computing is still fairly novel and there are currently few studies that can be considered replication studies, leaving a gap in the field. It is important to replicate results across multiple studies to demonstrate their efficacy.

Additionally, there is value in conducting multi-institutional, multi-national research. This permits many perspectives to be included from multiple different contexts which helps us to draw broader conclusions based on more diverse data [12]. They also demonstrate the efficacy of replications not with new data from the same or similar sources, but with multiple different contexts where differences in culture or attitudes may impact the broader conclusions being drawn.

2.6 Research Questions

Although theories have been proposed regarding the relationship between computing and spatial reasoning [22, 28], only limited work has been done to attempt to expand upon or provide empirical evidence for these contributions.

The goal of the current study is partially to perform replications of existing correlations between spatial skills and module grades at institutions where this has not been explored before, contributing to the existing but varied results already observed. Additionally, we wish to answer the call to action posed by Malmi *et al.* and build on theory work developed by previous researchers [21]. We present work similar to Pallrand and Seeber [24] discussed above to test the model proposed by Parkinson and Cutts [28]: does learning introductory computing lead to improvements in spatial skills? We also wish to consider how prior programming fluency affects the relationship between spatial skills and CS success, which has been theorised about by Margulieux [22] but evidence supporting it is limited.

Therefore, we have three research questions:

RQ1 Do spatial skills correlate with introductory CS module grades? (replication)

RQ2 Do spatial skills change over a period of introductory CS learning? (new, replicating physics)

RQ3 How does prior programming fluency affect the interplay between spatial skills and introductory CS module grades? (new)

We will also explore these relationships in more detail in the analysis of our results.

3 METHOD

3.1 Study Design

This study is the result of a multi-institutional project designed to investigate spatial skills in introductory CS across a year. Data is collected from three institutions located in the UK, the USA and Vietnam.

The study was designed such that tests and surveys were delivered to students at various points over a year of learning. A spatial skills test was delivered at the start of the first semester, at the end of the first semester or the beginning of the second, and at the end of the academic year prior to exams. This allows us to examine changes in spatial skills across a semester, as Pallrand & Seeber did [24], but also across a whole year to see if additional instruction yields additional changes.

Module grades from participating students were also collected at the end of the academic year after assessments had been conducted. Additionally, a self-reported level of prior programming fluency was collected at the start of the period of study.

3.2 Participants

All participants were taking part in an introductory CS module in their respective universities. The institutions involved were based in [redacted for anonymity, but will be included in the camera-ready manuscript]. Participation in the research was voluntary and was not compensated. The number of participants was not the same for every point of data collection due to some participants explicitly indicating that they wanted only certain data points to be used or due to students not taking part in the study. Ethical approval was acquired for the data collection at each institution and student participation was collected alongside a consent form explicitly permitting their data to be used for publication in aggregate and without identifying details.

Table 1: Number of participants taking part in each instrument/data collection point from each institution (Where MG1: module grade 1, MG2: module grade 2 and PPFS: prior programming fluency survey)

Inst.	PSVT:R 1	PSVT:R 2	PSVT:R 3	MG1	MG2	PPFS
A	122	87	27	136	126	146
B	38	37	4	17	*n/a*	18
C	50	35	*n/a*	50	*n/a*	44
Total	210	159	31	203	127	209

The number of participants for each data point is shown in table 1. Note, however, that some numbers are lower in our analysis in section 4: this is because most tests conducted required an intersection of at least two data points, and some students may have only contributed to one or the other. Also note that there is a decline in participation in later test points. We believe that this is because the tests were voluntary and students were probably experiencing test fatigue and assessment crunch by the end of the academic year.

3.3 Instruments and Data Collected

3.3.1 Spatial Skills. To measure spatial skills, the Revised Purdue Spatial Visualisation Test of Rotations (PSVT:R) [46] was used. The test requires participants to identify a sequence of rotations applied to a 3-D object to give the same object in a different, given orientation. They must then apply the same series of rotations to another object provided and choose the correct resulting orientation from a range of five options. The test consists of 30 items of increasing complexity. In each instance, the test was delivered with a 20-minute time limit.

As previously discussed in the Background section, the PSVT:R is specifically a test of spatial visualisation. Spatial visualisation is the factor of spatial skills most explored in spatial skills and CS research, and we continue this trend in this work. Also note that the revised PSVT:R specifically has been used in multiple works already discussed, particularly by Parkinson & Cutts and Bockmon, so we maintain this continuity here.

3.3.2 Prior Programming Fluency. We allowed students to self-report their prior programming fluency regarding the language they are most fluent in on the following scale:

(1) Professional fluency (contributed to open-source code, sold an application or worked as a professional programmer)
(2) Expert fluency (very comfortable using the language, created useful programs)

(3) Moderate fluency (write programs using basic control structures in the language)
(4) Novice fluency (basics of input/output and simple control flow only)
(5) None (very little to no programming experience)

We attempted to alleviate subjectivity in responses by including specific tasks and applications in the parentheses for each response, reducing ambiguity about what terms like "expert" and "novice" may entail. While this does not entirely eliminate subjectivity, and results may still be skewed by confidence, we believe that it is adequate to distinguish fluency for this study.

3.3.3 Module Grades. Module grades were collected using assessments from each of the institutions involved. In all instances, these are recorded as a single value of composite assessment scores according to the institution's own weighting scheme.

For institution A, module grades were a percentage which comprised of 20% weekly quizzes and 80% weekly programming exercises in the first semester and 20% weekly quizzes, 30% debugging assignments, and 50% on long-form programming assignments in the second semester.

For institution B, module grades were a letter grade in an ordinal range with 11 divisions which comprised of equal 25% contributions from homework assignments (7 in total), smaller practice problems (21 in total), quizzes (10 in total), and a final project.

For institution C, module grades were a letter grade in an ordinal range with 7 divisions which comprised of a wide range of formative and summative components:

- 20% weekly participation activities (including interactive pre-reading participation exercises, asynchronous in-lecture quizzes, post-lecture quizzes, and weekly team lab work activities)
- 25% auto-graded weekly short programming exercises
- 20% manually graded (human marked) weekly short programming exercises
- 20% across four module tests, one every three weeks
- 15% final exam

Different statistical tests are required for different forms of data, which we address in section 4.1. Institutions B and C collected grade data for the first semester only while institution A was able to collect grade data for both semesters individually.

3.4 Delivery

At each institution, time was taken during contact time to conduct the tests and surveys. They were all completed online, with students either using devices in the lecture room or completing the surveys alone as a class activity, using their institution's own preferred survey and testing platforms.

Institution C was not able to conduct the initial round of testing due to delays in ethical approval from the internal review board, so the data provided for institution C is the spring delivery of the introductory CS module. This places the students in similar circumstances to those starting their introductory CS in the autumn at other institutions, though the timing did not permit institution C to provide data across the following semester. Figure 2 shows the semester-based timeline and the data collected at each point by each institution.

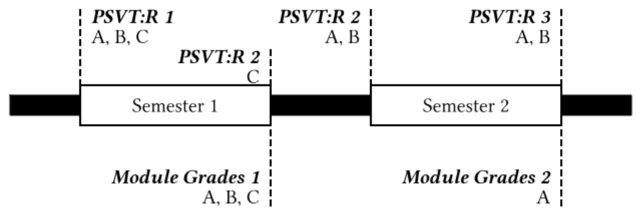

Figure 2: Timeline of data collection for each institution across both semesters

Three of the authors were instructors on one course each at their respective institutions which provided data. However, the invitation and subsequent handling of student participation was ruled as non-coercive by each institution's IRB/Ethics Review Committees.

3.5 On Statistical Significance

In the author guidelines, ICER promotes the position held by the American Statistical Association regarding p-values and significance [45]. In short, the guidance is not to draw conclusions about statistical significance based on an arbitrary value of $p<.05$ and rather to present p-values as they are and comment more broadly and thoughtfully on their meaning in the given context. For this reason, we present the p-values as they are calculated without any assumptions about statistical significance and comment more generally on the possibility of false positives occurring.

4 RESULTS

4.1 Pre-analysis: Normality and Distribution

The number of datapoints collected are shown in table 1. In order to determine the suitability of statistical tests to conduct, the data had to be examined for assumptions of normality and distribution. When analysed with a Shapiro-Wilk test, spatial skills and module grades were mostly found to be non-normally distributed with a positive skew. This also applied for most of the breakdowns by time period, institution and prior programming fluency sub-categories. Therefore, non-parametric analyses are used: these were Spearman's Rank tests for correlation, Wilcoxon Signed-Rank tests for comparisons of repeated measures at two intervals and the Kruskal-Wallis test for comparisons between groups. Non-parametric analyses are also the more appropriate choice for ordinal data, which applies to the letter-grade module results from institutions B and C. Additionally, since the data was skewed, medians are provided as the preferred measure of central tendency where applicable, and spread is indicated by the interquartile range as indicated by Q_1—Q_3. The results of the Shapiro-Wilk tests and the histograms of the data distribution can be found in the appendix.

4.2 RQ1: Spatial Skills and CS Module Results

In order to determine whether introductory programming module grades were associated with spatial skills, we conducted a series of correlations between spatial skills scores and module results. It was not appropriate to combine the module results between institutions

Table 2: Spearman's correlations between modules grades and spatial skills test scores for each institution

Institution	Semester	r	p	n
A	1	.44	.000	81
B	1	.58	.015	17
C	1	.50	.002	35
A	2	.53	.005	26

or between testing periods, since they were all measuring different things. Therefore, individual correlations are conducted for each institution.

In each instance, the spatial skills score used was the test result *chronologically closest* to the delivery of the major assessments. Although each institution had some summative components spread across the whole semester (particularly institution B) assessment was weighted towards the end of the first semester, so the chronologically closest spatial skills test was PSVT:R 2. Institution A collected a second semester of data, which is chronologically closest to PSVT:R 3. The results of the correlation tests can be found in table 2.

Each of these results indicates a moderate correlation of $r \approx .5$ with a low chance of a false positive result. Therefore, it appears that spatial skills are related to assessment scores at each institution involved, which is broadly in line with previous discoveries.

4.3 RQ2: Changes in Spatial Skills while Studying CS

Our second research question has a core component: do spatial skills change over a period of introductory CS learning across all institutions? Following this, we would also like to determine at which *institutions* spatial skills change by the most and which *students*, characterised according to their spatial skills scores, saw the greatest spatial skills changes.

4.3.1 Changes over Time. A series of Wilcoxon Signed-Rank tests were conducted between spatial skills tests at different times. Since the same version of the PSVT:R was used across every institution, unlike the module assessments scores, they can be combined and analysed together. The results of these three tests can be found in table 3. These results indicate that there is an increase in spatial skills across semester 1 and the full year which is unlikely to be a false positive, but the minimal gains across semester 2 are much more likely to be due to a false positive based on the high p-value.

4.3.2 Changes by Institution. We then conducted Wilcoxon Signed-Rank tests between spatial skills tests over the first semester for each institution. We only consider the first semester due to data restrictions: institution C has no data and institution B has only 4 data-points for PSVT:R 3, so meaningful analysis could only be conducted for institution A, allowing no comparisons. The results of the Wilcoxon Signed-Rank tests are found in table 4. The results indicate that there were gains in about 2 points in the PSVT:R across a semester for institutions A and C, and about 4 points for institution B. Note that according to traditional statistical significance expectations, we would not consider the changes in institution C to be significant, however in the spirit of moving away from arbitrary p-values as dictators of significance, we instead highlight that there is a greater chance of this result being a false positive than at other institutions, but this chance is still relatively low.

4.3.3 Changes by Initial Spatial Skills. In engineering, Sorby has previously split spatial skills scorers into three categories according to their PSVT:R scores: 18 and below indicates low scorers, who would require training; 19-21 inclusive indicates marginally high scorers, who would benefit from training but would not be required to participate in it; 22 and above indicates high scorers, who would probably not benefit from spatial skills instruction [36]. When categorising by these divisions, Ly *et al.* discovered that the students with the lowest initial spatial scores improved in spatial skills by the most points in the PSVT:R on average after spatial skills training [19]. To determine whether this applied in computing instruction, gains were calculated for the first semester for all students with enough data by subtracting their PSVT:R 1 score from their PSVT:R 2 score. Based on the PSVT:R 1, students from all institutions were grouped Sorby's three categories: Low (scoring 0-18 in the PSVT:R), Mid (19-21) and High (22-30). A Kruskal-Wallis test was then conducted with the students' gain data grouped by their categories. The results can be found in table 5.

These results indicate that there is a difference between the gains in spatial skills of students depending on their initial spatial skills. Students with initially low spatial skills improve by about 3 points on average while marginal students improved by about 1 point and students with high initial spatial skills did not improve on average.

Another way to examine this relationship is to conduct correlations between initial spatial skills and differences in spatial skills over the first semester. If negative correlations are observed, this would mean that the lower the initial spatial skills, the higher the gain. Spearman's correlations were conducted between PSVT:R 1 and PSVT:R 2 - PSVT:R 1 across all institutions and for each institution. The results are shown in table 6.

These results indicate that there is a negative correlation between spatial skills and gains across all institutions. This means that students with lower initial spatial skills tended to show the largest gains across the first semester, with institution B showing the highest correlation.

4.4 RQ3: Prior Programming Fluency

4.4.1 Fluency and Initial Spatial Skills. We predicted that those with greater programming fluency would have better spatial skills because they developed these skills while they first learned to program. To test this, we conducted a Kruskal-Wallis test with PSVT:R 1 scores grouped by fluency. Note that not enough participants indicated *Professional* programming fluency to include them in the Kruskal-Wallis test. The results are shown in table 7. These results indicate that there is a difference between groups which is not likely to be a false positive result. Students with no prior programming fluency have the lowest initial spatial skills and spatial skills rise with fluency levels.

4.4.2 Fluency and Correlation between Spatial Skills and Module Grades. Parkinson & Cutts theorised that students with high prior programming fluency would rely less on spatial skills, and therefore

Table 3: Results of Wilcoxon Signed-Rank tests between each spatial skills test across all institutions

	Semester 1		Semester 2		Whole Year	
PSVT:R	1	2	2	3	1	3
Median	20	22	24	24	20	25
Q_1—Q_3	16—24	17—25	18—26	18—27	15—23	18—27
n		133		30		26
Z		-3.34		-1.03		-2.54
p (2-tailed)		.001		.303		.011

Table 4: Results of Wilcoxon Signed-Rank tests at the start and end of the first semester for each institution

	Inst. A		Inst. B		Inst. C	
PSVT:R	1	2	1	2	1	2
Median	21	23	17	21	18	20
Q_1—Q_3	19—25	20—28	14—21	16—25	13—25	13—25
n		61		37		35
Z		-2.61		-3.14		-1.71
p (2-tailed)		.009		.002		.087

Table 5: Results of a Kruskal-Wallis test where students are grouped by Sorby's categories (Low, Mid, High) and then the median gains in spatial skills over a semester are compared

	Low	Mid	High
Median	3	1	0
Q_1—Q_3	0—7	0—3	0—2
n	55	22	56
H-stat	13.425		
p	.001		

Table 6: Spearman's correlations between initial spatial skills and the difference in spatial skills between the start and end of the first semester (spatial skills gains). Negative correlations indicate that students with the lowest initial spatial skills gain by the largest margins.

Institution	r	p	n
All	-.36	.000	133
A	-.25	.050	61
B	-.59	.000	37
C	-.36	.032	35

would exhibit a lower correlation between their spatial skills and module grades [29], in line with Margulieux's Spatial Encoding Strategy theory regarding domain specific skills [22]. In order to test this, we examined the correlation between grades and spatial skills as the end of the first semester split by prior programming fluency. Institution B did not have enough data points to conduct the correlations. Institution A did not have enough *Professional* respondents and institution C did not have enough *Professional* or *Expert* respondents who also completed the requisite PSVT:R and consented to have their grade data used. The results are shown in table 8.

The results show that there is a high correlation between spatial skills and module grades for students with no prior programming fluency, which decreases with each level of fluency. Note the high p-values of the more experienced students which indicate higher chances of the observed relationship being random and due to false positive result. These correlations indicate a trend for spatial skills having a stronger relationship with grades for those with less prior programming experience.

5 DISCUSSION

It is encouraging to see that the correlation between spatial skills and CS module scores has been observed here, as it has been in other studies. This work provides three additional institutions where the connection has not been observed before, which contributes to our more general understanding of spatial skills in relation to CS.

A key contribution of this study was to determine whether spatial skills change over a period of introductory CS learning, as they have been observed to in physics [24] and is proposed by the model presented by Parkinson & Cutts [28]. We have determined that spatial skills do appear to change over the first semester – and the first year – of study when looking at the dataset as a whole. As would be expected, there is a difference in how much spatial skills improve at each institution. Institution B observed the greatest gains, with institution A showing some gains and institution C observing similar gains when accounting for a higher chance of error.

We believe the reasons for this are due to institutional differences. Development of spatial skills during a spatial intervention happens

Table 7: Results of a Kruskal-Wallis test where the initial spatial skills were grouped and compared based on self-reported prior programming fluency. Note that not enough participants indicated *Professional* to reasonably contribute to the statistical analysis, so the values for this group are shown in italics for completeness but were not included in the calculation of the *H*-stat or the *p*-value.

	No Experience	Novice	Moderate	Expert	*Professional*
Median	18	22	24	24	*25*
Q_1—Q_3	15—24	15—-25	19—26	18—27	*18—30*
n	51	39	68	20	*3*
H-stat	6.539				
p	.001				

Table 8: Spearman's correlations between modules grades and spatial skills test scores for institutions A and C across semester 1, categorised by self-reported prior programming fluency

	Institution A			Institution C		
Fluency	r	p	n	r	p	n
No Experience	.51	.022	20	.61	.019	14
Novice	.47	.084	14	.33	.667	6
Moderate	.33	.069	31	.14	.764	7
Expert	.13	.641	15		*n/a*	
Professional		*n/a*			*n/a*	

through gradually increasing the challenge of spatial exercises, with activities such as sketching 3-D objects after rotations in multiple axes being some of the most effective training exercises [36]. Parkinson & Cutts proposed that the underlying cognitive skills, which Margulieux identified as non-verbal encoding skills [22], could be improved through the correct kind of CS exercises [28]. While we cannot be certain exactly what these exercises may look like, we would expect them to involve complex internalisation of operations, procedures or concepts, just like the complex rotation exercises. In order to determine whether these kinds of complex exercises are being used, we need to look closely at an institution's instruction. In this case, it might be possible to see if there are some distinct differences in the way institution B delivers content compared to institutions A and C. The nature of the instruction and its possible implications is explored in more detail in section 5.1.

Our work also validates a component of Margulieux's theory about expertise and CS learning: experts are likely to have developed domain specific strategies over time and will rely less on the abstract strategies underpinned by spatial skills to solve problems [22]. In our results, the correlation between spatial skills and module grades is exceptionally high for the students with no programming experience (higher than any relationship between spatial skills and computing success we have observed in the literature to date) and is negligible for experts. This greatly strengthens the theory that spatial skills, and their associated non-verbal encoding skills, are of the most value to students learning completely new content in computing.

A still-outstanding question is whether these self-identified *Expert* students would require spatial skills when engaging with completely new CS content, if it is actually possible to define this at all. Parkinson & Cutts' later work would seem to indicate that spatial skills are valuable all the way through a student's computing degree, affecting their module selection later in their programme [30], so it is possible that we would observe a stronger relationship between spatial skills and grades for our self-identified experts if we examined them further along their computing learning pathways.

5.1 Differences in Institutions

To attempt to tease out why one institution showed greater spatial skills gains than the others, we provide a discussion of the content and teaching methods conducted at each institutions. We also provide some theories on why some of these practices may develop spatial skills and not others.

Institution A teaches functional Haskell as a first language, which is not commonly taught in schools in their context and we do not believe it is a common self-taught language for incoming CS1 students. Therefore, we expect that this novelty reduces the effectiveness of students' pre-built strategies and language models, requiring them to build new understanding using their non-verbal encoding skills, thus relying on spatial skills. This may also explain why even students with *Moderate* prior fluency show a moderate correlation between spatial skills and success in the module, since even though they have some programming experience, it has limited bearing on the new language paradigm context. Ultimately, the added use of these non-verbal encoding skills should develop them, leading to improvements in spatial skills. However, since students with lower spatial skills develop them the most, institution A may not be observing gains as drastic as institution B because their students already started with high median spatial skills. They improved by 2 points, but from an already high starting point – if the students had begun with lower spatial skills, the gains may have been more substantial.

Institution B teaches a much more common initial language, Python, which it is reasonable to assume that students with prior programming fluency may have encountered before. However, the method of instruction and assessment may be making the difference: over the semester, students have multiple small exercises to work on per week, 7 assessed programming homework assignments and a large project. The homework assignments expect pair programming to be conducted and the final project is a group project. This high number of summative tasks and the need to engage with peers may result in spatial skills improvement through the repeated need to externalise and seriously recall mental models. Peer instruction has been regarded as an effective method of developing a deeper understanding for students [33], and while the practice of pair programming and group work is not peer instruction itself, it must involve the discussion and interrogation of students' conceptual models in a way that they must be constantly developed, altered and re-evaluated. This process of repeated mental model building is very probably the reason why spatial skills develop: building conceptual models rely on non-verbal encoding skills, and using these skills develops them, thus improving spatial ability. As already discussed regarding institution A, institution B students had the lowest incoming spatial skills, which also probably set them up to improve the most.

Institution C teaches a "late objects" approach in Java, which involves mainly procedural Java with some light-touch object oriented topics (basic classes and inheritance) at the end of the semester. There is a strong focus on Java comprehension, with explicit scaffolding to support the development of problem solving strategies. It is possible that the explicit strategy building leads to the development of domain-specific strategies rather than generic non-verbal encoding skills, leading to the higher chance of error observed in institution C's spatial skills gains. However, the correlation between spatial skills and assessment is still fairly high for a naturalistic study, suggesting that spatial skills still contribute meaningfully to CS outcomes. Additionally, in the delivery of content, institution C still incorporates multiple pair- and team-based discussions and activities, which we already suggested could have led to improvements in spatial skills in institution B (though perhaps the difference here is that students working together are practising their domain specific strategies rather than generic non-verbal strategies).

These theories do not explain why there was such limited development in semester 2. Both institutions A and B delivered a second-semester module teaching Java programming and data structures. Both teach a new language, however, there is a higher likelihood that Java will have been experienced by those with some prior programming fluency, giving them some domain-specific strategies to handle it. Institution B uses the same teaching structure as in semester 1, except that the group project is supplanted by an exam, so we would expect there to be some spatial skills gains. Despite this, we must also consider the semester 2 results with some caution: the Wilcoxon Signed-Rank test involves only 30 students, with only 4 from institution B, so they may not necessarily be indicative of the overall cohort. No data was collected from semester 2 for institution C, so their second semester delivery is not reported.

In short, although we have examined the nature of the delivery and content at each institution studied and provided some theories, there are no obvious, repeated characteristics which we believe specifically would lead to spatial skills improvement. We discuss this further as possible future work in section 5.4.

5.2 Limitations and Threats to Validity

Regarding gains in spatial skills over time, we did not have a control group as Pallrand & Seeber did in their study in physics [24]. We do not see this as a major threat to validity for several reasons. First, if there were a consistent retest gain, we would expect this to be consistent over time and by institution, where we see several variations. Additionally, neither the answers nor the test results were shared with the students and they had no means to revisit the questions once the test was completed, so the participants had no means to revise or practice the test itself.

The means of observing CS success was module grades, which can be seen in section 3.3.3 to be very different per institution and spread out across the semester. This means that spatial skills may be being correlated with quite different things in each instance. A more precise and homogeneous measure of CS success (such as the SCS1 [25]) would probably give more consistent results and could be compared across institutions. However, examining the relationship between spatial skills and module-specific measures of success is still valuable: in the context of each module, we expect that there have been contextual considerations put into the measures being used to determine success, and even if these are different between institutions, it is still valuable to know how spatial skills contribute to them. By comparing these measures, we examine the role of spatial skills with success in different contexts, *as measured* in different contexts.

Using a self-reported measure of programming fluency may be conflated with confidence. Over-confident students may report higher than their actual level of fluency and vice-versa with students with low confidence levels. We attempted to alleviate this by including more concrete measures of fluency for each category, but even then there is a risk that students would self-report differently from their actual fluency levels. However, the results broadly align with expectations and the need for extremely precise categorisation is not necessary for this research: the results speak for themselves. Even if there are some datapoints which we would more formally categorise in a different level of fluency, there is a clear difference in how the self-reported *No Experience* students differ from the *Moderate* and *Expert* students.

The analysis of the nature of each institution's CS learning experiences and content was conducted in an ad-hoc, discursive manner. We did not apply any kind of framework or rubric. This means that we may be missing some important components which would factor into the gains of spatial skills that may have been captured if we had followed a more formulaic and structured process to tease them out. We are not currently aware of such a framework or rubric which has been used to characterise a module in this manner, but if one exists it would have been valuable. Additionally, since this is a naturalistic study, it is not possible to capture every student's extra-curricular experiences. Such activities as playing video games [41] and technical drawing [39] are known to develop spatial skills, and we cannot guarantee that students did not engage with these activities throughout their modules to the degree that they may have improved spatial skills.

Bockmon & Cooper [1] recently published a viewpoint on participation bias in CS Education studies, noting that student who participate in voluntary activities may not be representative of the whole cohort (and also suggest that they are more likely to be skewed towards being higher academic performers). Given that all participation was entirely voluntary in this study, we may have encountered similar pitfalls. As stated by Bockmon & Cooper, these are challenges faced by our whole community and can be challenging to address; they suggest making testing instruments mandatory in classes, which in exploratory contexts or emerging fields – such as spatial skills – is a hard sell for institutions. However, this should be considered in any future work conducted.

5.3 Implications

Our first finding, that spatial skills are correlated with module grades, has implications for the breadth and far-reaching nature of the relationship. We have shown the relationship being replicated in institutions where it has not yet been observed and in the case of one institution, on a continent where spatial skills and CS specifically have never been explored before. This, along with all the other correlations observed in prior work, indicates that the relationship is not a fluke: there is a relationship between spatial skills and multiple measures of CS success which appear in many different deliveries of introductory computing content. Since spatial skills are easy to train (with gains being observed in multiple different kinds of spatial intervention deliveries in CS [27] and beyond [43]) we can help students with initially low skills to succeed in their programmes.

That spatial skills can be developed through CS learning is an encouraging finding. Not only does it take a step towards validating Parkinson & Cutts' underlying cognition model [28], it also means that it is possible that some deliveries of introductory CS will improve the spatial skills of students who – for whatever reason – begin their programme with low spatial skills. This means that some deliveries of CS might help to close gaps in spatial skills observed in incoming undergraduates naturally without a need for a spatial skills intervention.

However, there is still much more work to do. Most deliveries of Sorby's spatial skills intervention [36, 39], when run in the context of an introductory CS module, raise PSVT:R scores by about 6 points [27]; the highest observed gain in this study was only 4 points, and only at one institution. We don't yet know what it is about CS instruction which can improve spatial ability, and until we do we cannot simply assume that any CS delivery will make sure that students have opportunities to develop these skills. So while the implications of this study are positive, and indicate that we can improve these skills without the need for dedicated spatial interventions, we have some way to go before we can figure out just what we should be including in our programmes to help develop these skills.

Finally, our final finding about prior experience validates a component of Margulieux's Spatial Encoding Strategy theory [22] and has broader implications about who may need spatial skills training. It appears that the students with the least experience would benefit from these skills more, so under tight conditions spatial skills training would be of more benefit to them. However, as there is still a connection between spatial skills and later study [18, 30], we must be careful not to dismiss the possible need for students to have good spatial skills in order to succeed in later years.

5.4 Future Work

In CS we should conduct additional testing with students further along in their academic careers. Parkinson & Cutts demonstrate that spatial skills correlate with success in later study [30], so it would be valuable to determine whether spatial skills are still being developed at these stages or if the learning experiences of early study essentially dictate the students' later course selections.

As with traditional spatial skills training, there appears to be an upper threshold at which students are not likely to improve spatial skills through their CS learning. However, we have also discussed background work showing that spatial skills are still related to later years of study. Additionally, Parkinson & Cutts observed that spatial skills are higher among those further in their academic journeys: on average, CS faculty have substantially higher spatial skills than first year students [28]. This could be due to spatial skills continuing to develop over time as students encounter more complex concepts, or it could be because only those with initially higher spatial skills end up progressing and keep the average scores high. This should be investigated in more detail: do students develop spatial skills in their second year of study? What about even later: do the spatial skills of final year students or practitioners in industry change over time?

It would also be valuable to more deeply explore just what kind of learning experiences in computing improve spatial skills. We have discussed these in broad terms here, but are unable to draw any conclusions about which learning experiences may influence spatial skills development. More robust and rigorous exploration of these activities would help us to identify the best ways to gradually improve the spatial skills of our students and avoid the need for supplemental spatial skills interventions. Initial steps could involve pre- and post-testing spatial skills of students around different groups participating in different kinds of learning activities (reading lecture notes alone, pair programming, peer instruction activities, etc.) or after short bursts of being introduced to different content (early language constructs, objects and classes, recursion, etc.).

It is also worth noting that the entirety of this study took up only about an hour and a half of student time in each institution, most of it spent on the spatial skills tests (which require about 30 minutes to administer). Anyone wishing to replicate the work could do so fairly easily if they were able to garner the correct institutional support or student motivation to examine how spatial skills may be involved in their own context.

6 CONCLUSION

We are still working to understand the relationship between spatial skills and CS. The discovery that spatial skills can be developed through CS learning is encouraging because it means that we can structure our teaching so that all students get an opportunity to develop these important skills while working through their normal degree. However, these findings do not constitute permission to dismiss spatial skills as skills that will naturally develop in our students and are not worth further attention: not all computing instruction appears to develop spatial skills at the same rate, so not all computing students will naturally pick them up in class. Also, even the highest gains we have observed are not as effective as dedicated interventions: evidence shows that spatial skills can be improved by up to 6 points on the PSVT:R over a short and lightweight spatial skills intervention taken alongside CS learning [27]. The next steps – to understand how students use these skills and in what ways we can effectively and naturally develop them in our programmes – are important.

In order to do this, we need to understand the theory. We need to know how these connections manifest and how they can be leveraged into helping our students and improving our learning experiences. This paper also contributes to this process. We have discovered (more) evidence that spatial skills are connected with success in computing, and we also now know that this is of high importance for students who are just beginning their journey and have not programmed before. This work builds towards a broader understanding of how students are expected to learn. Continuing to tease out this relationship and generate a good understanding of how these cognitive skills can be developed, measured, and applied will help researchers and instructors to innovate education practices that are more equitable and beneficial to all students.

REFERENCES

[1] Ryan Bockmon and Stephen Cooper. 2022. What's your placebo? *Commun. ACM* 65, 10 (Oct. 2022), 31–33. https://doi.org/10.1145/3528085

[2] 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*. ACM, Minneapolis MN USA, 552–557. https://doi.org/10.1145/3287324.3287372

[3] 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*. ACM, Trondheim Norway, 446–451. https://doi.org/10.1145/3341525.3387380

[4] 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*. ACM, Portland OR USA, 766–772. https://doi.org/10.1145/3328778.3366829

[5] Neil C. C. Brown, Eva Marinus, and Aleata Hubbard Cheuoua. 2022. Launching Registered Report Replications in Computer Science Education Research. In *Proceedings of the 2022 ACM Conference on International Computing Education Research V.1*. ACM, Lugano and Virtual Event Switzerland, 309–322. https://doi.org/10.1145/3501385.3543971

[6] Jeffrey Buckley, Niall Seery, and Donal Canty. 2018. A Heuristic Framework of Spatial Ability: a Review and Synthesis of Spatial Factor Literature to Support its Translation into STEM Education. *Educational Psychology Review* 30, 3 (Sept. 2018), 947–972. https://doi.org/10.1007/s10648-018-9432-z

[7] John B. Carroll. 1993. *Human Cognitive Abilities: A Survey of Factor-Analytic Studies* (1 ed.). Cambridge University Press, Cambridge, UK. https://doi.org/10.1017/CBO9780511571312

[8] 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, Omaha Nebraska USA, 13–20. https://doi.org/10.1145/2787622.2787728

[9] Anthony Cox, Maryanne Fisher, and Philip O'Brien. 2005. Theoretical Considerations on Navigating Codespace with Spatial Cognition. In *Proceedings of the 17th Annual Workshop of the Psychology of Programming Interest Group*. Psychology of Programming Interest Group, Brighton, UK, 9. https://ppig.org/papers/2005-ppig-17th-cox/

[10] Brian Dorn and Allison Elliott Tew. 2015. Empirical validation and application of the computing attitudes survey. *Computer Science Education* 25, 1 (Jan. 2015), 1–36. https://doi.org/10.1080/08993408.2015.1014142

[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*. Technical Report. University of Kent.

[12] Sally Fincher and Marian Petre (Eds.). 2004. *Computer science education research*. Taylor & Francis, London ; New York.

[13] Sue Jones and Gary Burnett. 2007. Spatial skills and navigation of source code. In *Proceedings of the 12th annual SIGCSE conference on Innovation and technology in computer science education - ITiCSE '07*. ACM Press, Dundee, Scotland, 231. https://doi.org/10.1145/1268784.1268852

[14] Sue Jones and Gary Burnett. 2008. Spatial Ability And Learning To Program. *Human Technology: An Interdisciplinary Journal on Humans in ICT Environments* 4, 1 (2008), 47–61. https://doi.org/10.17011/ht/urn.200804151352

[15] Maria Kozhevnikov, Mary Hegarty, and Richard Mayer. 2002. Spatial Abilities in Problem Solving in Kinematics. In *Diagrammatic Representation and Reasoning*, Michael Anderson, Bernd Meyer, and Patrick Olivier (Eds.). Springer London, London, 155–171. https://doi.org/10.1007/978-1-4471-0109-3_9

[16] Maria Kozhevnikov, Michael A. Motes, and Mary Hegarty. 2007. Spatial Visualization in Physics Problem Solving. *Cognitive Science* 31, 4 (July 2007), 549–579. https://doi.org/10.1080/15326900701399897

[17] Patrick C. Kyllonen, David F. Lohman, and Richard E. Snow. 1984. Effects of aptitudes, strategy training, and task facets on spatial task performance. *Journal of Educational Psychology* 76, 1 (Feb. 1984), 130–145. https://doi.org/10.1037/0022-0663.76.1.130

[18] Ken Liu, Burkhard C. Wünsche, and Andrew Luxton-Reilly. 2022. Relationship Between Spatial Skills and Performance in Introductory Computer Graphics. In *Proceedings of the 27th ACM Conference on on Innovation and Technology in Computer Science Education Vol. 1*. ACM, Dublin Ireland, 304–310. https://doi.org/10.1145/3502718.3524756

[19] Anna Ly, Jack Parkinson, Quintin Cutts, Michael Liut, and Andrew Petersen. 2021. Spatial Skills and Demographic Factors in CS1. In *Koli Calling*. ACM, Joensuu, Finland, 1–10. https://doi.org/10.1145/3488042.3488049

[20] Aaron Mac Raighne, Avril Behan, Gavin Duffy, Stephanie Farrell, Rachel Harding, Robert Howard, Edmund Nevin, and Brian Bowe. 2015. Examining the Relationship Between Physics Students' Spatial Skills and Conceptual Understanding of Newtonian Mechanics. In *The 6th Research in Engineering Education Symposium*. TU Dublin, Dublin, Ireland, 8. https://doi.org/10.21427/8ET5-X129 Publisher: Technological University Dublin.

[21] Lauri Malmi, Judy Sheard, Päivi Kinnunen, Simon, and Jane Sinclair. 2019. Computing Education Theories: What Are They and How Are They Used?. In *Proceedings of the 2019 ACM Conference on International Computing Education Research*. ACM, Toronto ON Canada, 187–197. https://doi.org/10.1145/3291279.3339409

[22] Lauren E. Margulieux. 2019. Spatial Encoding Strategy Theory: The Relationship between Spatial Skill and STEM Achievement. In *Proceedings of the 2019 ACM Conference on International Computing Education Research*. ACM, Toronto ON Canada, 81–90. https://doi.org/10.1145/3291279.3339414

[23] 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 (July 1986), 605–610. https://doi.org/10.1145/6138.6142

[24] George J. Pallrand and Fred Seeber. 1984. Spatial ability and achievement in introductory physics. *Journal of Research in Science Teaching* 21, 5 (May 1984), 507–516. https://doi.org/10.1002/tea.3660210508

[25] 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 (ICER '16)*. Association for Computing Machinery, New York, NY, USA, 93–101. https://doi.org/10.1145/2960310.2960316 event-place: Melbourne, VIC, Australia.

[26] 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*. ACM, Espoo Finland, 97–105. https://doi.org/10.1145/3230977.3230987

[27] Jack Parkinson, Ryan Bockmon, Quintin Cutts, Michael Liut, Andrew Petersen, and Sheryl Sorby. 2021. Practice report: six studies of spatial skills training in introductory computer science. *ACM Inroads* 12, 4 (Dec. 2021), 18–29. https://doi.org/10.1145/3494574

[28] 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*. ACM, Espoo Finland, 106–114. https://doi.org/10.1145/3230977.3230990

[29] 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*. ACM, Trondheim Norway, 439–445. https://doi.org/10.1145/3341525.3387413

[30] Jack Parkinson and Quintin Cutts. 2022. Relationships between an Early-Stage Spatial Skills Test and Final CS Degree Outcomes. In *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education*. ACM, Providence RI USA, 293–299. https://doi.org/10.1145/3478431.3499332

[31] Jack Parkinson, Quintin Cutts, and Steve Draper. 2020. Relating Spatial Skills and Expression Evaluation. In *United Kingdom & Ireland Computing Education Research conference*. ACM, Glasgow United Kingdom, 17–23. https://doi.org/10.1145/3416465.3416473

[32] Herbert L. Pick and Jeffrey J. Lockman. 1983. Map Reading and Spatial Cognition: Discussion. In *Spatial Orientation*, Herbert L. Pick and Linda P. Acredolo (Eds.). Springer US, Boston, MA, 219–224. https://doi.org/10.1007/978-1-4615-9325-6_10

[33] Leo Porter, Cynthia Bailey Lee, Beth Simon, Quintin Cutts, and Daniel Zingaro. 2011. Experience report: a multi-classroom report on the value of peer instruction. In *Proceedings of the 16th annual joint conference on Innovation and technology in computer science education*. ACM, Darmstadt Germany, 138–142. https://doi.org/10.1145/1999747.1999788

[34] Bruno Poucet. 1993. Spatial cognitive maps in animals: new hypotheses on their structure and neural mechanisms. *Psychological review* 100, 2 (1993), 163. Publisher: American Psychological Association.

[35] Simon, Sally Fincher, Anthony Robins, Bob Baker, Ilona Box, Quintin Cutts, Michael De Raadt, Patricia Haden, John Hamer, Margaret Hamilton, and others. 2006. Predictors of success in a first programming course. In *Conferences in Research and Practice in Information Technology*, Vol. 52. Australian Computer Society, Hobart, Australia, 189–196.

[36] Sheryl Sorby. 1999. Developing 3-D Spatial Visualisation Skills. *The Engineering Design Graphics Journal* 63, 2 (1999), 21–32.

[37] Sheryl Sorby, Norma Veurink, and Scott Streiner. 2018. Does spatial skills instruction improve STEM outcomes? The answer is 'yes'. *Learning and Individual Differences* 67 (Oct. 2018), 209–222. https://doi.org/10.1016/j.lindif.2018.09.001

[38] Sheryl A. Sorby. 2009. Educational Research in Developing 3-D Spatial Skills for Engineering Students. *International Journal of Science Education* 31, 3 (Feb. 2009), 459–480. https://doi.org/10.1080/09500690802595839

[39] Sheryl Ann Sorby, Anne Frances Wysocki, and Beverly Gimmestad Baartmans. 2003. *Introduction to 3-D spatial visualization: an active approach.* Thomson/Delmar Learning, Clifton Park, N.Y. OCLC: ocm51301756.

[40] Sorby, Sheryl A. and Baartmans, Beverly J. 1996. A Course for the Development of 3-D Spatial Visualization Skills. *Engineering Design Graphics Journal* 60, n1 (1996), 13–20.

[41] Ian Spence and Jing Feng. 2010. Video Games and Spatial Cognition. *Review of General Psychology* 14, 2 (June 2010), 92–104. https://doi.org/10.1037/a0019491

[42] Denise Tolhurst, Bob Baker, John Hamer, Ilona Box, Raymond Lister, Quintin Cutts, Marian Petre, Michael De Raadt, Anthony Robins, Sally Fincher, and others. 2006. Do map drawing styles of novice programmers predict success in programming? A multi-national, multi-institutional study. *Research in Practice in Information Technology* 52 (2006), 213–222.

[43] 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 (March 2013), 352–402. https://doi.org/10.1037/a0028446

[44] 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 (Nov. 2009), 817–835. https://doi.org/10.1037/a0016127

[45] Ronald L. Wasserstein, Allen L. Schirm, and Nicole A. Lazar. 2019. Moving to a World Beyond " $p < 0.05$". *The American Statistician* 73, sup1 (March 2019), 1–19. https://doi.org/10.1080/00031305.2019.1583913

[46] So Yoon Yoon. 2011. *Psychometric properties of the Revised Purdue Spatial Visualization Tests: Visualization of Rotations (the Revised PSVT:R).* Ph. D. Dissertation. Purdue University.

A DATA HISTOGRAMS AND NORMALITY TEST RESULTS

A.1 Spatial Skills

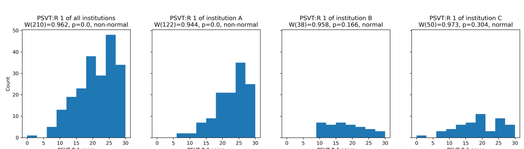

Figure 3: PSVT:R 1 distribution and Shapiro-Wilk test results for the full dataset and each institution

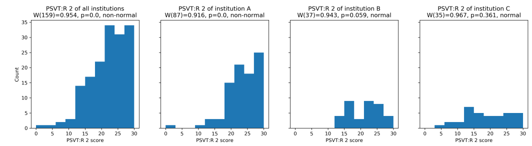

Figure 4: PSVT:R 2 distribution and Shapiro-Wilk test results for the full dataset and each institution

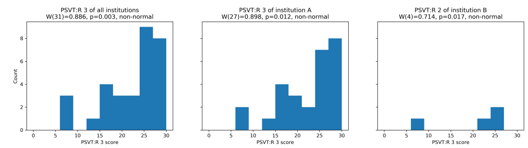

Figure 5: PSVT:R 3 distribution and Shapiro-Wilk test results for the full dataset and each institution (where data was provided)

A.2 Module Grades

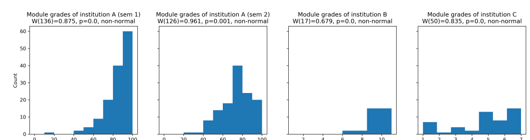

Figure 6: Grade distribution and Shapiro-Wilk test results for each institution