ARTICLE

Check for updates

Defining and Measuring the Influences of GIS-Based Instruction on Students' STEM-Relevant Reasoning

Erin A. Jant, David H. Uttal, Bob Kolvoord, Katherine James, and Camille Msall

Abstract

Geospatial technologies, such as geographic information systems (GIS), remote sensing, and GPS have been used in a variety of educational settings to help improve student learning. A sample of 53 high school seniors was recruited from the Geospatial Semester (GSS), a course that emphasizes the use of GIS for problem-solving and students in AP Physics and AP History served as a comparison. GSS students' spatial thinking and problem solving improved across the school year in contrast to Comparison Group. Results suggest that GIS-based instruction can be used to enhance students' use of spatial reasoning when solving STEM-relevant problems.

Key Words: *education, geographic information systems, problem solving, spatial thinking, STEM*

Erin Jant is a visiting assistant professor in the Department of Psychology, Binghamton University, Binghamton, New York, USA.

David Uttal is a professor in the Department of Psychology and School of Education and Social Policy, Northwestern University, Evanston, Illinois, USA.

Robert Kolvoord is the Dean of the College of Integrated Science and Engineering, James Madison University, Harrisonburg, Virginia, USA.

Katherine James is a postdoctoral fellow in the School of Education and Social Policy, Northwestern University, Evanston, Illinois USA.

Camille Msall is a lab manager in the School of Education and Social Policy, Northwestern University, Evanston, Illinois, USA.

Introduction

This paper explores the role of spatial thinking in STEM education. Many studies have demonstrated a strong relation between spatial ability (as measured through psychometric tests) and STEM achievement and attainment, even after controlling for other related variables, such as verbal and mathematical skills (Uttal, Miller, and Newcombe 2013; Wai, Lubinski, and Benbow 2009). Some evidence suggests training spatial ability also supports improvements in STEM attainment or achievement (e.g. Duffy et al. 2016; Cheng and Mix 2014; Power et al. 2017; Uttal, Miller, and Newcombe 2013).

Here we argue for a more expansive approach to integrating spatial thinking into STEM. We call our approach to scientific thinking and learning *STEM-relevant spatial thinking*. It is characterized by what has been termed a "spatial habit of mind" (Liben, Kastens, and Stevenson 2002; Liben and Downs 2003), a tendency to think about problems and data in spatial terms and to know how to apply spatial approaches to solve complex, everyday science and engineering problems. Thus, we want to extend the importance of spatial thinking in STEM education beyond what is typically measured by spatial ability tests and bring it in line with approaches that emphasize the *practices* of STEM thinking. In this case, our focus is on spatial practices that are relevant to inquiry-based spatial thinking. Given the critical role of spatial thinking in STEM achievement, STEM education that enhances students' spatial reasoning and problem-solving skills may contribute to success in future STEM education and careers.

Background

The Role of Geospatial Technologies in Supporting and Promoting STEM-Relevant Spatial Thinking

We suggest that geospatial technologies, particularly Geographic Information Systems, (*GIS*) play a very important role in promoting the development and application of STEM-relevant spatial thinking. Geographic information systems (*GIS*) have transformed the practice of geography and geospatial education. The ability to understand and use geospatial technologies has become fundamental in the search for solutions to complex problems, such as climate change, the epidemiology of new diseases, and economic disparities (Bednarz et al. 2008; Bednarz and Van Der Schee 2006). In fact, education reforms have called for the use of GIS as a tool for improving K-12 students' spatial thinking (NGSS 2013, 2016; NRC 2006). GIS software products have been suggested as a potential powerful support system for spatial thinking because they facilitate data visualization and allow users to think flexibly about spatial patterns and relations among different kinds of data (Bednarz et al. 2008; Bednarz and Van Der Schee 2006; Bodzin 2011; Bodzin et al. 2014).

Use of GIS in Education

Curricula that incorporate GIS have been used at a variety of educational levels, from elementary school through college (see, Alibrandi 2003; Bednarz and Van Der Schee 2006; Bloom and Palmer-Moloney 2004; Bodzin 2011; Hagevik 2003; Johansson 2003; Kulo and Bodzin 2011; Milson and Alibrandi 2008; Pang 2006; Sinton 2009). Many researchers and practitioners have shown interest in integrating GIS-based curricula because they believe it can improve students' analytical thinking and learning (see, Alibrandi 2003; Bednarz and Van Der Schee 2006; Bloom and Palmer-Moloney 2004; Bodzin 2011; Hagevik 2003; Johansson 2003; Milson and Alibrandi 2008; Pang 2006).

Prior research has demonstrated the effectiveness of GIS-based education in a variety of contexts (e.g., Bodzin, Peffer, and Kulo 2012; Bodzin, et al. 2014; Edelson, Smith, and Brown 2008; Goldstein and Alibrandi 2013; Lee and Bednarz 2009). For instance, students' understanding of energy, climate change, and social studies improved through the use of a curriculum that incorporated geospatial technologies compared to students who had regularly planned instruction in these areas (e.g., Bodzin and Fu 2014; Edelson, Smith, and Brown 2008; Kulo and Bodzin 2013).

In the present research, we focus on the development of STEM-relevant spatial thinking in the context of a GIS-based, inquiry-driven STEM curriculum. Our goals were (a) to develop methods to assess STEM-relevant spatial thinking, and (b) to apply those methods to examine the effects of one approach to teaching STEM-relevant spatial thinking with GIS, the Geospatial Semester (GSS).

Present Study

To study spatial education in the classroom, we selected a test bed of schools participating in a program called The Geospatial Semester (*GSS*) (Kolvoord, Keranen, and Rittenhouse 2019). The GSS is a year-long, inquiry-based course for high school seniors that focuses on developing geospatial problem-solving skills, using state-of-the-art software, and the application of those skills to an extended local problem chosen by the students. The course is taught in the students' high school, with the assistance and sponsorship of a large public university. GSS students can engage with sophisticated STEM content that cuts across multiple disciplines. Importantly, the GSS was deliberately designed around the application of STEM-relevant spatial thinking and reasoning strategies. Students collect, analyze, model, and interpret data, and communicate their results.

Initially, students learn the mechanics of the software and a variety of analysis techniques as they solve somewhat constrained real-world problems. As their skills grow, students begin to work independently through less constrained problems. Learning GIS skills is an important part of the course however, the primary focus of the class is the extensive real-world final project that students independently develop and define, and which includes authentic science and engineering. Each student selects a question to pursue, finds relevant data (e.g., via the Internet, public databases, libraries, private businesses, etc.), tests hypotheses, constructs layered visual representations, conducts analysis, and presents that analysis orally and in a poster.

We focused on the GSS as a venue in which to investigate the learning and practice of STEM-relevant spatial thinking and problem solving because the course used an intensive, inquiry-based approach to the application of spatial information and spatial representations in STEM-problem solving. However, the implications of our research need not be limited to the GSS; the present research has the potential to catalyze efforts to develop effective forms of spatial education using a variety of methods and in a variety of different contexts besides the GSS (see Baker et al. 2015; Hsi, Linn, and Bell 1997).

We investigated whether the GSS promoted STEM-relevant spatial thinking and problem solving, such as problem identification, finding and selecting evidence, and developing arguments using that evidence. We assessed whether and when participation in the GSS influenced students' thinking and the extent to which they adopted spatial approaches to solve novel, complex, real-world problems. In addition, we investigated whether students could extend that problem solving beyond specific course content and apply new, spatial approaches to problems of science and engineering. We predicted that students in the GSS would take a spatial approach to reasoning about and solving novel problems we presented in contrast to a comparison group of students. Finally, we anticipated that GSS students who used more spatial language and scored higher on problem-solving measures would also create comparatively advanced final projects, as indicated by the final project rubric.

Methods

Overview

This quasi-experimental research study collected data across one school year. Students were interviewed and asked to solve hypothetical scenarios (Transfer Questions) several times during the school year to examine progress in STEM-relevant spatial thinking and problem solving. All interview responses were examined for the use of spatial words, as an indicator of spatial thought, and coded using a rubric to assess problem-solving skills. The GSS students were additionally evaluated on their final projects in the course.

Participants

The 53 high school juniors and seniors who participated in this study were recruited from urban and rural high schools in Virginia. The schools that participated included rural and suburban locations and ranged from approximately 500 to more than 2000 students. Thirty-eight GSS students (9 female, 29 male) volunteered to participate. As a Comparison Group, an additional 15 students (11 female, 4 male) volunteered to participate from AP Physics and AP History courses. The Comparison Group students were never enrolled in the GSS.

Materials

Assessments. We designed and validated a set of assessments of STEM-relevant spatial thinking. We explored whether participation in the GSS changed students' understanding of, and approach to problem solving. Specifically, we developed three different measures to capture the changes in students' STEM-relevant problem solving: Interview Questions, Transfer Questions, and GSS Final Projects.

Interview Questions. The interviews covered a wide range of questions and addressed students' reasons for taking the course, goals for their individual GSS projects, issues with gathering information to answer their questions, and generally the process of completing the final projects.

Transfer Questions. We designed a set of hypothetical questions to assess scientific reasoning and problem solving. These are termed *Transfer Questions (TQ)* because they present students with the opportunity to apply the problem-solving processes to solve a novel real-world problem. For example, students were asked the following questions at Interview 1 or 4 in a counterbalanced fashion:

> "If you were running a campaign for a local political office, how would you go about running your campaign?"

> "If your city needed to add an additional landfill, and you were in charge of the process, how would you go about determining where it should be?"

> "Why do you think milk prices differ from brand to brand? And how would you predict what the price would be for each brand?"

> "Why do you think gas prices differ from station to station? And how would you predict what the price would be for each station?"

The TQs could be answered either with or without the use of STEM-relevant spatial thinking. For example, a student could answer the previous question by saying that he or she would go about running the campaign by purchasing radio advertisements to inform the public about his or her campaign. This response does not reveal STEM-relevant spatial thinking. In contrast, another student might say that he or she would use data to represent the distributions of income, voter registration, and donations to political parties to highlight the best geographical regions to target the campaign. This student's approach does reveal STEM-relevant spatial thinking. Analyzing students' solutions to the TQs can provide insight into the impact of the GSS on STEM-relevant spatial thinking.

GSS Final Projects. We assessed GSS Groups' final projects along a rubric with six quality dimensions (adapted from Charles and Kolvoord 2013).

Design and Procedures

Data Collection: Interviews. To examine progress in students' STEM-relevant thinking and problem solving, the GSS Group was interviewed 4 times during the course of the school year: once at the beginning of the course (T1), at 3 months (T2), 6 months (T3), and the end of the course (T4). The Comparison Group students were interviewed at T1 and T4, at approximately the same time as the GSS students. We did not interview the Comparison Group at T2 and T3 because these interviews focused primarily on the development of GSS students' final projects and thus would not make sense to include the Comparison Group. At the T1 and T4 interviews, we asked the Comparison Group the same questions as the GSS Group regarding uses of technology, course projects, and what they found to be interesting in the courses in which they were enrolled. Additionally, at each of the four interviews, the GSS Group was asked specific questions regarding the development of their individual GSS course final projects. Both groups were asked Transfer Questions at T1 and T4. The TQs were counterbalanced across participating schools and interviews. All the interviews were conducted by an experimenter who was familiar with GIS but had no previous contact with the students. The interviews were videotaped and transcribed.

Random assignment to courses was not possible as the GSS was an elective class; therefore, the Comparison Group helped to reduce (but could not eliminate) some possible confounds. For example, students' STEM-relevant spatial thinking might improve during their senior year of high school regardless of what classes they took. We recruited students from Advanced Placement courses for the Comparison Group, as we believed that these students would be like those who enrolled in the GSS. The baseline interviews at T1 serve as another source of comparison for any preexisting group differences that could serve in the analyses as controlled covariates.

Coding

Interviews including TQs were coded to investigate the development of STEM-relevant spatial thinking and problem solving. We also evaluated the quality of students' GSS final projects.

Spatial Language. Students' responses to the interview and Transfer Questions were analyzed for the presence of spatial language. We counted the number of distinct (no repetitions) spatial words used in the interviews (e.g., cardinal directions, relational language, spatial dimensions, etc.). This measure assumed that what students talked about, and specifically, what words they used, would reflect how they were thinking about the problems (see Pennebaker 2012). For instance, Pennebaker and Seagal (1999) and Pennebaker et al. (2007) found that a clients' mental health could be reliably assessed by counting the number of positive (e.g., happy, joyful) and negative (e.g., sad, depressed) words that clients used in their journal. This method has expanded considerably and is used to assess sentiments expressed in a variety of publication outlets, ranging from novels to Facebook and Twitter postings (Qiu et al. 2012; Settanni and Marengo 2015; Wang et al. 2016). Here we applied this technique to examine whether the words that students used in answering the interview questions could provide a valid measure and insight into their STEM-relevant spatial thinking.

A critical question in this process was which words to count; what words constitute evidence of spatial thinking? This is not an easy question, as some words could be ambiguously spatial. One salient example is that the word "right" could be used as part of a spatial description, but also to mean "okay" or "correct".

Fortunately, we are not the first to use spatial language as a proxy indicator of spatial thinking, and thus we could rely on prior dictionaries of spatial words. We used two spatial dictionaries, one developed by Pennebaker and colleagues (Pennebaker et al. 2007; Pennebaker 2012) and another developed by Cannon, Levine, and Huttenlocher (2007). For the purposes of our analysis, these two dictionaries were combined. As an additional safeguard against ambiguously spatial words, we conducted a content review on the resulting dictionary to identify words with many non-spatial uses (e.g. to, from) and words that are not spatial in the sense we are exploring (e.g. shape and pattern words). Based on this analysis, 73 words were excluded from the combined dictionary. The resulting combined dictionary was used in our analysis.

We used the program Linguistic Inquiry and Word Count (LIWC) (e.g., Pennebaker 2012) to count the spatial words in each interview transcript. The program compared each word in each transcript to the Spatial Dictionaries described above and provided an output as a percentage. We converted the percentage to a raw value in order to evaluate the exact number of distinct words in each transcript that were found in the spatial dictionaries. Our goal was to assess the use of new, distinct spatial language; therefore, we only counted each spatial word once per interview, which guards against crediting frequent repetition of the same word as evidence of increases in spatial language.

Validation of Spatial Word Counting through Coding. One potential limitation of the spatial word counting method is that it does not capture the context in which the words are used (Pennebaker 2012). We wanted to be sure that the spatial words that we were counting were a valid indicator of STEM-relevant spatial thinking. Therefore, we validated the word counting by coding interview transcripts by hand. First, words in the transcripts (interviews and Transfer Questions) were indicated as potentially spatial by the LIWC software. Coders then examined the transcripts to provide a secondary check of the LIWC analyses and determined if the word was used in a spatial context in order to remove ambiguity of words such as "right". We coded 50% of the transcripts in this way and found that 98% of the words coded as spatial by the LIWC software were indeed used in a spatial context and in reference to spatial domains. Two independent coders, blind to condition, were reliable at identifying if a word was indeed used in a spatial context or not, at 96.2 percent agreement (kappa = 0.92).

STEM-Relevant Problem Solving. Counting the number of spatial words that students used in their answers to the interview and Transfer Questions can indicate whether the GSS students shifted toward spatial approaches to STEM-relevant problem solving. However, it does not address the question of whether the use of these words was in the service of providing good, scientifically-sound answers to our questions. To address this question, we used an adapted version of the McNeill and Krajcik (2011) *Constructing Explanations in Science* rubric to judge students' scientific explanation and problem solving as demonstrated by their answers to the Transfer Questions. This framework captures the depth of analysis and appropriateness of the approaches students used to construct their answers to the Transfer Questions and to compare the GSS and Comparison-group students.

Following McNeill and Krajcik (2011), we assessed three aspects of students' answers to the Transfer Questions: claim, evidence, and reasoning (Table 1). *Claims* were basic statements of hypotheses that are testable and that provided an initial answer to the Transfer Question posed. *Evidence* involved providing sources of data that were appropriate to support the stated claim. For example, to answer the Transfer Question, students would need to provide information about how and where they would collect data, what type of data that would be, and how much information gathering is needed to fully solve the issue and so forth. *Reasoning* referred to students' justifications that explicitly connected the sources of data to the claim explaining why that data would count as evidence and why that data was adequate to support the hypothesis. Each component was then evaluated on a scale of 0-2 assessing if each claim, evidence and justification given was sufficient and accurate (Table 1).

The coders were blind to condition; they did not know if the participant was enrolled in the GSS or in the Comparison Group. The two coders independently scored 25% of the Transfer Questions; one coder coded the remainder of the data with checks by the second coder. Inter-rater agreement in scoring students' answers to the Transfer Questions overall was 86.3% (kappa = 0.73) for Claims, 84.3% (kappa = 0.75) for Evidence, and 88.4% (kappa = 0.82) for Reasoning.

GSS Final Project Assessment. Finally, we evaluated the quality of GSS projects based on a scoring rubric that has been developed and validated by Charles and Kolvoord (2013). The rubric assesses six aspects of the final project: (a) Is the question specifically spatial in nature; (b) Is the problem authentic, an issue that matters to a group of individuals and that can be approached through a scientific or engineering approach; (c) Sources of data used, effectiveness of data collection or mining of existing date; (d) Sophistication of analysis; (e) Map design and analysis (f) Quality of final presentation. Each of these outcomes was assessed on a five-point scale: Advanced, Very Capable, Proficient, Basic, and Novice (Table 2). Three coders independently scored the final projects and presentations of the GSS students and reached agreement of 83.8% (kappa = 0.71).

Results

Preliminary Analyses

Preliminary analyses assessed whether there were any differences on the dependent measures that were related to the variables of interviewer, school, or sex of the student. We conducted 2 (Condition: GSS Group, Comparison Group) × 2 (Sex) ANOVAs for each dependent variable that we included in the primary analyses. There were no main effects or interactions involving sex, $Fs < 1.33$, $ps > 0.26$. Likewise, 2 (Interviewer: A, B) × 2 (Condition: GSS, Comparison Group) ANOVAs also showed no main effects or interactions, $Fs < 2.05$, $ps > 0.16$. In addition, students from the three high schools performed similarly on all dependent measures; $Fs < 1.57$, $ps > 0.22$.

Main Analyses

Spatial Problem Solving. Our first question was whether participating in the GSS led to an overall increase in spatial problem solving. We expected students to use more spatial language and provide longer answers in latter interviews as the students progressed in the school year and in their final projects.

Controlling for Amount of Overall Speech. One potential concern with the analysis of frequency of spatial words is that the use of more spatial words might simply represent more overall talk. To examine this possibility,

Table 1. Scientific problem solving rubric and corresponding examples.

Code	Score and examples		
Claim	0	1	2
Basic statements of hypotheses that were testable and that provided an initial answer to the Transfer Question posed	Ex: I don't know how I would find a landfill location	Ex: I might find some available land and make sure it is large enough	Ex: I would identify a region with little to no population and look for the correct amount of land needed for waste based on the city size
Evidence			
Involved providing sources of data that were appropriate to support the stated claim	Ex: I would ask the mayor for some site recommendations	Ex: I would need to look for people selling large pieces of land by finding information online or in the paper	Ex: I would need to collect data on population density, then I could plot the data points on a map to determine a location that is far away from houses and schools
Reasoning			
Referred to students' justifications and reasoning that explicitly connected the sources of data to the claim	Ex: The mayor or governor would probably have to approve the landfill site	Ex: Online would have information that list how big the land is that people are selling and that would help determine where to put the landfill	Ex: It is important to examine population density in certain regions on the map in order to find a site that is large enough to accommodate the city size and in a location that will not impact the water supply, schools, or houses

Table 2. Geospatial semester final project rubric.

Spatial question	Novice	Basic	Proficient	Advanced
The student poses a question with a spatial component that is based on a sound understanding of geospatial technology. -Using 21st century skills to understand and address global issues	The project does not have a spatial component.	The project has a spatial component that is based on a limited understanding of geospatial technology.	The project has a spatial component that is based on a sound understanding of geospatial technology.	The project has a spatial component that is based on an advanced understanding of geospatial technology.
Authentic problems				
The student uses geospatial technology to identify and solve complex problems in real world contexts	Student lacks expertise, and/or interest in using geospatial technology to solve complex, authentic problems.	Student applies geospatial technology to complex, authentic problems only with significant support and direction from the teacher. Products created are fairly traditional.	Student applies geospatial technology to complex, authentic problems with minimal support and direction from the teacher, occasionally creating products that have real value to audiences outside of the classroom.	Student applies technology to complex, authentic problems independently, often creating products that have real value to audiences outside of the classroom.
Data generated or mined				
The student acquires geospatial data from professional sources OR -Creates geospatial data (e.g. field data, digitized data, georeferencing historical maps) -uses information accurately and creatively for the issue or problem at hand	Student acquires incomplete geospatial data from sources OR incorrectly creates geospatial data	Student acquires limited geospatial data from sources OR Creates geospatial data with some limitations	Student acquires geospatial data from professional sources OR Creates geospatial data (e.g. field data, digitized data, georeferencing historical maps)	Student acquires particularly rich geospatial data from professional sources OR Creates particularly insightful geospatial data (e.g. field data, digitized data, georeferencing historical maps)
Map				
The student uses directly related, clear, and correct maps to describe their analysis	The student's maps are not clear or do not use good cartographic design. The maps and other deliverables do not related to the spatial problem.	Some of the student maps are clear and cartographically correct, but many are not adequate or directly related to the spatial problem.	Most of the student maps are clear and cartographically correct, but some have substantial errors or are not adequate or directly related to the spatial problem.	The student develops clear and cartographically correct maps that describes their spatial analysis. All maps and deliverables directly related to the spatial problem at hand.
Analysis				
The student is able to use geospatial technology to analyze information *Classification* *Selection* *Geoprocessing tools*	Student lacks analytical skills with geospatial tools.	Student can conduct simple analysis using geospatial tools with assistance.	Student can provide appropriate analysis using geospatial tools of the problem.	Student provides thoughtful and insightful analysis using geospatial tools, often identifying relationships and elements within a problem that are beyond expectations.

(*Continued*)

Table 2. (*Continued*).

Spatial question	Novice	Basic	Proficient	Advanced
Presenting				
The student presents information clearly and accurately	Student is unable to present information accurately, concisely, or clearly.	Student is able to present information somewhat accurately, but the presentation is neither clear nor compelling.	Student is able to present information accurately and efficiently, but the presentation is not entirely compelling.	Student is able to present information accurately, efficiently, and in a compelling manner.

we examined the total number of words that participants used during each interview; we call this measure *total talk*. We did in fact find that students who talked more used more spatial words; total talk at each interview time was correlated with the number of spatial words that students used at T1: $r(51) = 0.78$, $p < 0.001$; T2: $r(36) = 0.86$, $p < 0.001$; T3: $r(36) = 0.83$, $p < 0.001$; and T4: $r(51) = 0.93$, $p < 0.001$. Therefore, we used the total number of words that participants used as a covariate in the remaining spatial language analyses to control for the overall talkativeness of each student and better examine differences in students' use of spatial words.

Spatial Language during Transfer Questions. We examined differences between the GSS and Comparison Group in use of spatial language specifically during answers to the Transfer Questions. We evaluated the language used by the two groups at T1 and T4. Our between group comparisons focused on answers to the Transfer Questions because both groups were asked these questions at the same times (T1 and T4). As there were four different Transfer Questions, we first tested for differences regarding the type of question posed. We did not find any significant group by Transfer Question interactions at T1 or T4, $Fs < 2.95$, $ps > 0.10$. For these analyses, we combined answers to the Transfer Questions at each respective time point into a single spatial language measure for T1 and T4.

To examine the change in students' use of spatial language while answering the Transfer Questions, we conducted a repeated measures ANOVA, controlling for Total Talk, which revealed an interaction of Interview (1, 4) × Condition (GSS, Comparison Group) on amount of spatial words used by students, $F(1, 48) = 8.13.50$, $p < 0.05$, $^2 = 0.14$. There was no difference between GSS and Comparison Group at T1; both groups used about the same number of spatial words in answering the Transfer Questions at the first interview. However, at T4 students in the GSS ($M = 17.31$, $SD = 8.13$) used significantly more spatial words than the Comparison Group ($M = 10.00$, $SD = 5.44$), $F(1, 50) = 9.65$, $p < 0.05$, $^2 = 0.16$. This result indicates that at T4 the GSS students answered the questions in spatially richer ways than the Comparison Group, which suggests that they were now engaging in more spatially based reasoning. The Comparison Group used slightly *fewer* spatial words at T4 compared to T1; however, the decrease was not significant.

STEM-Relevant Scientific Reasoning and Problem Solving. The counting of spatial words suggests that the GSS led to a greater use of STEM-relevant spatial thinking. But these analyses do not reveal *how* the GSS affected scientific reasoning, and whether students were actually giving richer answers that reflected a more sophisticated understanding of the scientific process to problem solving. To address this issue, we evaluated the three dimensions (Claim, Evidence, and Reasoning) of the STEM-relevant problem-solving rubric with non-parametric inferential statistics because of the limited scale (0 to 2) of each rubric dimension.

We used a Mann-Whitney test to analyze students' scores at T1 on the scientific problem-solving rubric. GSS and Comparison Group students performed comparably on all three measures of Claim, Evidence, and Reasoning, $Us = 214.00$ - 230.00, $ps > 0.36$, with no significant differences. Thus, at the start of the school year students in both courses demonstrated similar levels of scientific reasoning and problem solving.

However, at T4 GSS students (M = 1.29, SD = 0.65) used significantly more appropriate and sufficient evidence than the Comparison Group (M = 0.60, SD = 0.73), $U = 144.50$, $p < 0.01$. The GSS' students (M = 1.45, SD = 0.55) Claims were also more complete in contrast to the Comparison Group (M = 1.07, SD = 0.65), $U = 183.00$, $p < 0.05$. The GSS students (M = 1.18, SD = 0.69) provided significantly better justifications or Reasoning that linked Evidence to the Claims, $U = 152.00$, $p < 0.01$ than the Comparison Group (M = 0.53, SD = 0.74). These results highlight that the two groups did not differ at T1 but diverged substantially by T4 (see S4 for contextualized examples). For instance, at T4, GSS Group's Reasoning and Evidence scores were nearly twice has high as those of the Comparison Group.

Scientific Problem-Solving and Spatial Language during Transfer Questions. To evaluate evidence of student's spatially relevant scientific problem solving we analyzed the relation between our two measures of problem solving: The detailed coding of scientific reasoning and problem solving when answering the Transfer Questions and the spatial language. Assessing this relation helps to validate the word counting methods and provides additional insights into students' strategies in applying a spatial approach to scientific reasoning. We used a partial Spearman correlation that controlled for students' total talk (total word count at T1). At T1, spatial language was not correlated with the rubric measures of Claims r_s (48) = 0.17, $p = 0.24$ or Evidence $r_s(48) = 0.10$, $p = 0.48$. Spatial language was moderately correlated with the measure of Reasoning $r_s(48) = 0.27$, $p = .06$. Similarly, at T4, spatial language was not correlated with the problem-solving measure of Claims, $r_s(50) = 0.15$, $p = 0.28$. However, spatial language *was* correlated with students' Evidence, $r_s(50) = 0.27$, $p < 0.05$ and Reasoning scores, $r_s(50) = 0.28$, $p < 0.05$. Therefore, the students that tended to use more spatial language also appeared to provide evidence that is more complete and demonstrated higher levels of reasoning to justify that evidence, regardless of group membership. Spatial language did not appear to be correlated with the measure of Claims on this rubric.

Final Project Rubric. The final projects of each student in the GSS were scored along a rubric that focused on the quality of the project on six dimensions (Charles and Kolvoord 2013). There was a range of proficiencies in the GSS classroom as displayed in the final projects. The following analysis applies only to the GSS students, as the Comparison Group did not complete a GSS final project.

GSS Final Projects and STEM-Relevant Spatial Thinking. Analysis of correlations between Final Project Rubric scores and other measures provides further evidence for the validity of our STEM-relevant problem-solving assessments. We again controlled for total talk by partialing it out of the Spearman correlations. The GSS students' use of spatial language at T4 was significantly correlated with 4 of the 6 Final Project Rubric dimensions $r_s = 0.27$- 0.41, $ps < 0.05$. Thus, at the end of the semester, students who used more spatial language also had more highly rated final projects in the areas of Spatial Question, Maps, Analysis, and Data. This result again reflects the development of greater STEM-relevant spatial thinking.

We also examined how students' scores on the Scientific Reasoning and Problem-Solving Rubric correlated with scores on the GSS Final Project Rubric. Spearman correlational tests showed that at T4 the problem-solving categories of Claim, Evidence, and Reasoning were correlated with Final Project Rubric measures and the Rubric categories were highly correlated with one another, $r_s = 0.27 - 0.95$, $ps < 0.05$. The skills that students demonstrated when solving a novel problem were related to their skills in the GSS course at the end of the term.

Discussion

We found that the students in the GSS and Comparison Groups began the academic year with similar levels of STEM-reasoning and problem solving. However, the two groups diverged over the course of the term. The results indicate that participating in the GSS improved students' STEM reasoning. The GSS Group gained substantially, whereas the Comparison Group did not improve, or even declined slightly. Further, participation in the GSS affected how students interpreted and solved new problems. The results indicated that the GSS students were much more likely to approach the problems from a perspective that emphasized spatial solutions. The analysis of the spatial language results revealed that at the end of the course the GSS Group used more spatial strategies to solve the novel problems in contrast to the Comparison Group. The results suggest that the GSS helped to engender a spatial approach to STEM problem solving. This adds to growing evidence indicating that spatial approaches to problem solving in science and education can be taught, and that this teaching can successfully influence how students approach new problems.

However, one potential concern regarding the word-counting method is that the words are decontextualized; they are counted without regard to how they are used in the answer. The words are simply treated as a "bag" of words, without regard to how the words are used (Lin and Demner-Fushman 2004). Thus, it was hypothetically possible that students could have used more spatial words as they progressed through GSS for a variety of reasons that would not necessarily reflect growth in STEM-relevant spatial thinking. However, the content review of the dictionary and validation of spatial word counting ensured that only relevant instances of spatial language were counted. Further, most of the spatial language, problem solving, and final project scores were significantly correlated with each other. Taken together, the convergence of these sources of evidence suggests more than just a change in use of spatial language but rather a shift in STEM-relevant spatial reasoning and problem solving.

Scientific Reasoning

The results indicated substantial improvement in the GSS Group in 2 of the 3 measures (evidence and reasoning). The experiences in the GSS influenced students' approach to problem solving such that, when given the opportunity during the Transfer Questions, the GSS students used more elaborate and complete scientific reasoning. The GSS students made connections when problem solving, bridging claims to evidence and the reason for doing so, more so than students who did not experiences the GSS curriculum. Both groups improved in the category of "claim", perhaps because making a claim was the first part of giving an answer, and all groups had learned the need to make an assertion or claim. The GSS educational experience influenced students' approach to problem solving and use of spatial language in the process.

Correlations between Spatial Language and Scientific Reasoning

We found significant, although modest, correlations between our two measures of STEM-relevant spatial thinking. Moreover, spatial language was only correlated with measures of Evidence and Reasoning at T4. Perhaps the measure of Claim represents a more general component of problem solving that is not specifically spatial. A student could provide a valid Claim without necessarily taking a spatial approach to the problem. Or perhaps in general, as students progressed in the academic year their answers became more spatial regarding data gathering and analysis, thus they used more spatial terms to describe their solutions to the Transfer Questions.

Limitations

A limitation in this work is that the correlation between spatial language to the categories of Evidence and Reasoning, while significant, is small. This could be due to the limited scale (0-2) used to measure problem solving. Perhaps a more nuanced version of the problem-solving rubric could better capture more detailed elements of scientific reasoning that provide the variability to reflect links to spatial language and thought. Importantly, the problem-solving rubric was not designed to explicitly evaluate *spatial* problem solving, but general *scientific* problem solving. Thus, finding a correlation between students who use more spatial terms and problem solve in more sophisticated ways provides one indication of a shift in students' approach to solving novel problems with a spatial approach. This work provides one insight to a link between spatial language and problem solving, however future work could address the limitations.

We compared the learning and performance of students' in the GSS Group to students enrolled in AP History or Physics (or both). We selected this Comparison Group because they would have had the same or similar prior coursework as the GSS Group had, and because all these classes were senior-level electives. Of course, it is always possible that other, unmeasured differences could account for the pattern of observed results. For example, the GSS Group may have been more interested in spatially related topics than the Comparison Group and hence learned more quickly or readily. The ideal design, of course, would involve random assignment to the GSS or Comparison Groups. However, this Comparison Group did allow for assessment of baseline similarities and differences on the dependent measures as well as growth over the school year that could be attributed to the influence of the GSS course beyond reasonable change due to maturation alone. The pattern of results obtained demonstrate the influence of the GSS experience on students' approach to problem solving and building a *spatial habit of mind*.

Another concern is that our sample was largely male, and we had unequal sized samples, with a larger number of GSS than comparison students. For both samples, volunteers were recruited with the same methods; however, more students in the GSS course volunteered to participate than those in the AP Physics and History courses. Both concerns are partially mitigated by the pre- and posttest research design and by no statistically significant differences in sex. There were no preexisting group differences on our measures of spatial language or scientific problem solving, but large differences at post-test. This pattern of results is difficult to explain solely based on differences in the sample sizes. Nevertheless, it would be helpful in future research to have more diverse and comparable samples.

Conclusions

The current study makes a novel contribution to ongoing discussions regarding inquiry-based STEM instruction by highlighting the critical role of STEM-relevant spatial thinking in this work. Taken together, our results suggest that spatial thinking, a critical component of many disciplines of science and engineering, can also be considered a central practice in science and engineering education and the standards that guide that work (NGSS 2013, 2016). A specific recommendation is that spatial thinking could, and should, be discussed as central to scientific reasoning, just as modeling and evidence-based argumentation are. Revisions to the NGSS or other problem-based approaches to science and engineering education should include spatial thinking as part of the repertoire and practices of science.

In summary, we suggest an answer to the very important question of how GIS-based instruction influences students STEM-relevant spatial thinking (see also Kerski 2003; Kim and Bednarz 2013; Kim, Bednarz, and Lee 2011; Lee and Bednarz 2009). We have demonstrated that GIS-based STEM instruction can influence not only the content of students' problem solving but also their approach to problem solving. GIS training helps student to consider, understand, and implement spatial solutions. In addition, we have provided methods that can be used in future studies to analyze changes in STEM-relevant spatial thinking and reasoning. Our findings and methods offer new support for the integration of geospatial technologies in K-12 education and may prove useful to new efforts to understand the role of spatial thinking in science and engineering education.

Note: Examples of student projects and example survey questions can be found at: https://doi.org/10.21985/n2-wb1h-bc59

Funding

This work was supported by National Science Foundation.

ORCID

Bob Kolvoord http://orcid.org/0000-0001-9678-8178

REFERENCES

Alibrandi, M. 2003. *GIS in the classroom: Using geographic information systems in social studies and environmental science*. Portsmouth, NH: Heinemann.

Baker, T. R., S. Battersby, S. W. Bednarz, A. M. Bodzin, B. Kolvoord, S. Moore, D. Sinton, and D. Uttal. 2015. A research agenda for geospatial technologies and learning. *Journal of Geography* 114 (3):118–30. doi: 10.1080/00221341.2014.950684.

Bednarz, S. W., B. Chalkley, S. Fletcher, I. Hay, E. L. Heron, A. Mohan, and J. Trafford. 2008. Community engagement for student learning in geography. *Journal of Geography in Higher Education* 32 (1):87–100. doi: 10.1080/03098260701731553.

Bednarz, S. W., and J. Van Der Schee. 2006. Europe and the United States: The implementation of geographic information systems in secondary education in two contexts. *Technology, Pedagogy and Education* 15 (2): 191–205. doi: 10.1080/14759390600769573.

Bloom, E., and L. J. Palmer-Moloney. 2004. Meeting the challenge: Integrating geographic technology into today's social studies classroom. *Meridian: A Middle School Computer Technologies Journal* 7 (1):1–5. https://projects.ncsu.edu/meridian/win2004/gis/.

Bodzin, A. M. 2010. Integrating web-based activities and site-based experiences to investigate environmental issues. In *The inclusion of environmental education in science teacher education*, 323–36. Netherlands: Springer. Q7

Bodzin, A. M. 2011. The implementation of a geospatial information technology (GIT)-supported land use change curriculum with urban middle school learners to promote spatial thinking. *Journal of Research in Science Teaching* 48 (3):281–300. doi: 10.1002/tea.20409.

Bodzin, A. M., and Q. Fu. 2014. The effectiveness of the geospatial curriculum approach on urban middle-level students' climate change understandings. *Journal of Science Education and Technology* 23 (4): 575–90. doi: 10.1007/s10956-013-9478-0.

Bodzin, A. M., Q. Fu, V. Kulo, and T. Peffer. 2014. Examining the effect of enactment of a geospatial curriculum on students' geospatial thinking and reasoning. *Journal of Science Education and Technology* 23 (4):562–74. doi: 10.1007/s10956-014-9488-6.

Bodzin, A., T. Peffer, and V. Kulo. 2012. The efficacy of educative curriculum materials to support geospatial science pedagogical content knowledge. *Journal of Technology and Teacher Education* 20 (4):361–86.

Bransford, J. D., and D. L. Schwartz. 1999. Chapter 3: Rethinking transfer: A simple proposal with multiple implications. *Review of Research in Education* 24 (1): 61–100. doi: 10.3102/0091732X024001061.

Cannon, J., S. Levine, and J. Huttenlocher. 2007. A system for analyzing children and caregivers' language about space in structured and unstructured contexts. Spatial Intelligence and Learning Center (SILC) Technical Report. http://www.silccenter.org/aigaion2/index.php/publications/show/1154.

Charles, M. T., and R. A. Kolvoord. 2013. Using visualization tools for inquiry-based science: A longitudinal study of teacher's stages of development. In *Proceedings of the 2013 ESRI Education Users Conference*, San Diego, CA.

Cheng, Y. L., and K. S. Mix. 2014. Spatial training improves children's mathematics ability. *Journal of Cognition and Development* 15 (1):2–11. doi: 10.1080/15248372.2012.725186.

Cook, T. D., W. R. Shadish, and V. C. Wong. 2008. Three conditions under which experiments and observational studies produce comparable causal estimates: New findings from within-study comparisons. *Journal of Policy Analysis and Management* 27 (4): 724–50. doi: 10.1002/pam.20375.

Driver, R., P. Newton, and J. Osborne. 2000. Establishing the norms of scientific argumentation in classrooms. *Science Education* 84 (3):287–312. doi: 10.1002/(SICI)1098-237X(200005)84:3<287::AID-SCE1>3.0.CO;2-A.

Duffy, G., S. Sorby, S. Nozaki, and B. Bowe. 2016. Exploring the role of spatial cognition in problem-solving. In *Frontiers in Education Conference (FIE), 2016 IEEE*, 1–4. Q8

Edelson, D. C., D. A. Smith, and M. Brown. 2008. Beyond interactive mapping: Bringing data analysis with GIS into the social studies classroom. In *Digital geography: Geospatial technologies in the social studies classroom*, 77–98. Charlotte, NC: IAP.

Gick, M. L., and K. J. Holyoak. 1983. Schema induction and analogical transfer. *Cognitive Psychology* 15 (1): 1–38. doi: 10.1016/0010-0285(83)90002-6.

Goldstein, D., and M. Alibrandi. 2013. Integrating GIS in the middle school curriculum: Impacts on diverse students' standardized test scores. *Journal of Geography* 112 (2):68–74. doi: 10.1080/00221341.2012.692703.

Goldstone, R. L., and S. B. Day. 2012. Introduction to "new conceptualizations of transfer of learning.

Educational Psychologist 47 (3):149–52. doi: 10.1080/00461520.2012.695710.

Hagevik, R. A. 2003. The effects of online science instruction using geographic information systems to foster inquiry learning of teachers and middle school science students. PhD diss. http://www.lib.ncsu.edu/resolver/1840.16/4389.

Hsi, S., M. C. Linn, and J. E. Bell. 1997. The role of spatial reasoning in engineering and the design of spatial instruction. *Journal of Engineering Education* 86 (2):151–8. doi: 10.1002/j.2168-9830.1997.tb00278.x.

Johansson, T. 2003. GIS in teacher education-facilitating GIS applications in secondary school geography. In ScanGIS, 285–93.

Kerski, J. 2000. Fits, starts, and headway: The implementation of geographic information systems technology and methods in beginning and advanced high school geography courses. In *The 11th Annual U.S. Geological Survey, Central Region, 1999 Poster Review*. https://pubs.usgs.gov/of/1999/0321/report.pdf

Kerski, J. 2001. A national assessment of GIS in American high schools. *International Research in Geographical and Environmental Education* 10 (1):72–84. doi: 10.1080/10382040108667425.

Kerski, J. 2003. The implementation and effectiveness of geographical information systems technology and methods in secondary education. *Journal of Geography* 102 (3):128–37. doi: 10.1080/00221340308978534.

Kim, M., and R. Bednarz. 2013. Effects of a GIS course on self-assessment of spatial habits of mind (SHOM). *Journal of Geography* 112 (4):165–77. doi: 10.1080/00221341.2012.684356.

Kim, M., R. Bednarz, and S.-I. Lee. 2011. GIS education for teachers in South Korea: Who participates and why? *Journal of the Korean Geographical Society* 46 (3):382–95.

Kolvoord, R., M. Charles, and S. Purcell. 2012. What happens after the professional development: Case studies on implementing GIS in the classroom. Chapter submitted for inclusion in *Teaching with geospatial technology: Designing effective professional development for teachers*.

Kolvoord, B., K. Keranen, and S. Rittenhouse. 2019. The geospatial semester: Concurrent enrollment in geospatial technologies. *Journal of Geography* 118 (1):3–10. doi: 10.1080/00221341.2018.1483961.

Kulo, V. A., and A. M. Bodzin. 2011. Integrating geospatial technologies in an energy unit. *Journal of Geography* 110 (6):239–51.

Kulo, V., and A. Bodzin. 2013. The impact of a geospatial technology-supported energy curriculum on middle school students' science achievement. *Journal of Science Education and Technology* 22 (1):25–36. doi: 10.1007/s10956-012-9373-0.

Lee, J., and R. Bednarz. 2009. Effect of GIS learning on spatial thinking. *Journal of Geography in Higher Education* 33 (2):183–98. doi: 10.1080/03098260802276714.

Liben, L. S., and R. M. Downs. 2003. Investigating and facilitating children's graphic, geographic, and spatial development: An illustration of Rodney R. Cocking's legacy. *Journal of Applied Developmental Psychology* 24 (6):663–79. doi: 10.1016/j.appdev.2003.09.008.

Liben, L. S., K. A. Kastens, and L. M. Stevenson. 2002. Real-world knowledge through real-world maps: A developmental guide for navigating the educational terrain. *Developmental Review* 22 (2):267–322. doi: 10.1006/drev.2002.0545.

Lin, J., and D. Demner-Fushman. 2004. "Bag of words" is not enough for strength of evidence classification. In *AMIA… Annual Symposium Proceedings. AMIA Symposium*, 1031–1031. American Medical Informatics Association.

McNeill, K. L. 2011. Elementary students' views of explanation, argumentation and evidence and abilities to construct arguments over the school year. *Journal of Research in Science Teaching* 48 (7):793–823. doi: 10.1002/tea.20430.

McNeill, K. L., and J. S. Krajcik. 2011. *Supporting grade 5-8 students in constructing explanations in science: The claim, evidence, and reasoning framework for talk and writing*. Saddle River, NJ: Pearson.

Milson, A. J., and M. Alibrandi. 2008. *Digital geography: Geospatial technologies in the social studies classroom*. Charlotte, NC: IAP.

National Research Council. 2006. *Learning to think spatially: GIS as a support system in the K-12 curriculum*. Washington, DC: National Academies Press.

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

Newcombe, N. S. 2017. Harnessing spatial thinking to support STEM learning. Organisation for Economic Co-operation and Development (OECD) Reports.

NGSS Lead States. 2013. *Next generation science standards: For states, by states*. Washington, DC: The National Academies Press.

NGSS Lead States. 2016. *Next Generation Science Standards: For States, by States*. Washington, DC: The National Academies Press.

Pang, A. 2006. Geographical information systems in education. In *Ministry of Education, Singapore IT Literature Review Educational Technology Division*.

Pennebaker, J. W. 2012. *The psychology of physical symptoms*. Verlag, NY: Springer.

Pennebaker, J. W., and J. D. Seagal. 1999. Forming a story: The health benefits of narrative. *Journal of Clinical Psychology* 55 (10):1243–54. doi: 10.1002/(SICI)1097-4679(199910)55:10<1243::AID-JCLP6>3.0.CO;2-N.

Pennebaker, J. W., C. K. Chung, M. Ireland, A. Gonzales, and R. J. Booth. 2007. The development and psychometric properties of LIWC2007. http://www.liwc.net/LIWC2007LanguageManual.pdf.

Power, J. R., S. A. Sorby, N. L. Veurink, K. Atit, and M. Carr. 2017. Preparing students for engineering success through improving 3-D spatial skills. In *2017 ASEE Annual Conference and Exposition*.

Qiu, L., H. Lin, J. Ramsay, and F. Yang. 2012. You are what you tweet: Personality expression and perception on Twitter. *Journal of Research in Personality* 46 (6):710–8. doi: 10.1016/j.jrp.2012.08.008.

Settanni, M., and D. Marengo. 2015. Sharing feelings online: Studying emotional well-being via automated text analysis of Facebook posts. *Frontiers in Psychology* 6:1045. doi: 10.3389/fpsyg.2015.01045.

Sinton, D. S. 2009. Roles for GIS within higher education. *Journal of Geography in Higher Education* 33 (sup1): S7–S16. doi: 10.1080/03098260903034046.

Steiner, P. M., T. D. Cook, W. R. Shadish, and M. H. Clark. 2010. The importance of covariate selection in controlling for selection bias in observational studies. *Psychological Methods* 15 (3):250. doi: 10.1037/a0018719.

Uttal, D. H., D. I. Miller, and N. S. Newcombe. 2013. Exploring and enhancing spatial thinking: Links to achievement in science, technology, engineering, and mathematics? *Current Directions in Psychological Science* 22 (5):367–73. doi: 10.1177/0963721413484756.

Wai, J., D. Lubinski, and C. 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): 817. doi: 10.1037/a0016127.

Wang, W., I. Hernandez, D. A. Newman, J. He, and J. Bian. 2016. Twitter analysis: Studying US weekly trends in work stress and emotion. *Applied Psychology* 65 (2):355–78. doi: 10.1111/apps.12065.

Wright, R., W. L. Thompson, G. Ganis, N. S. Newcombe, and S. M. Kosslyn. 2008. Training generalized spatial skills. *Psychonomic Bulletin and Review* 15 (4):763–71. doi: 10.3758/PBR.15.4.763.