Why do the rich get richer? A structural equation model to test how spatial skills affect learning with representations

Martina A. Rau
Department of Educational Psychology
University of Wisconsin—Madison
1025 W. Johnson St
Madison, WI 53706
+1-608-262-0833
marau@wisc.edu

ABSTRACT

Spatial skills predict students' success in STEM domains. This paper aims to better understand the difficulties of students with low spatial skills in using interactive graphical representations. I present a mediation analysis with test and log data from 117 students who worked with an intelligent tutoring system for chemistry. The analysis is based on (1) a knowledge component model that describes knowledge students acquire as they solve problems with graphical representations, (2) a search for features that describe students' interactions with the representations and that are predictive of students' learning gains, and (3) a structural equation model that tests whether these features statistically mediate the effect of spatial skills on students' learning gains. Results show that only students' ability to plan representations before they construct them mediates the effect of spatial skills on learning gains. This finding suggests that these students may need more support *before* they construct representations.

Keywords

Spatial skills, intelligent tutoring systems, interactive representations, STEM learning.

1. INTRODUCTION

Students' spatial skills predict learning success in STEM domains [1, 2]: students with low spatial skills tend to show lower achievements in STEM domains and they are less likely to pursue careers in these domains. Spatial skills are important for STEM learning because many concepts in STEM domains are inherently visuo-spatial. For example, astronomers have to visualize the solar system, engineers have to visualize interactions among components of a machine, and chemists have to visualize movements of atoms and electrons. To make these concepts accessible to students, instructional materials in STEM domains tend to heavily rely on the use of graphical representations [5, 6]. Graphical representations are external representations that use visuo-spatial features to depict domain-relevant concepts (as opposed to text or symbols). As a consequence, students have to make sense of visuo-spatial relationships depicted by graphical representations to understand abstract concepts in STEM domains [7].

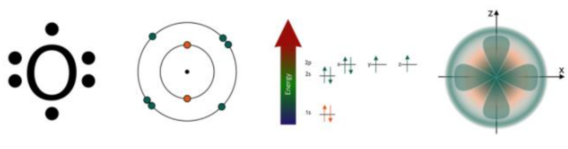

Figure 1. Graphical representations of an oxygen atom: Lewis structure, Bohr model, energy diagram, orbital diagram.

Consider, for example, a student who is learning about atomic structure. Figure 1 shows the graphical representations that instructional materials typically use to illustrate atomic structure [8]. Lewis structures (left) show paired an unpaired valence electrons, Bohr models (center-left) show all electrons in atomic shells, energy diagrams (center-right) depict electrons in orbitals with their energy level, and orbital diagrams (right) show the spatial arrangement of non-empty orbitals. To understand atomic structure, students have to integrate the information depicted in these graphical representations into a visuo-spatial mental model of how electrons are arranged relative to the atom's nucleus, and how they move according to probabilistic laws.

Integrating such information into a mental model of the domain-relevant concepts requires students to hold the relative location of the depicted objects in working memory and to mentally rotate these objects [9]. The cognitive load imposed by this task is arguably higher for students with low spatial skills than for students with high spatial skills [1]. As a consequence, students with low spatial skills may fail at this task, which might jeopardize their learning success [1, 5, 9]. On the flip side, students with high spatial skills are more successful at integrating visuo-spatial information into mental models, and—consequently—are likely to show higher learning gains. Thus, the rich (in spatial skills) get richer (in content knowledge).

Educational technologies such as intelligent tutoring systems (ITSs) hold particular promise for breaking the "the-rich-get-richer" rule and for creating an "everyone-gets-richer" rule, because they can address the needs of students with low spatial skills in several ways. First, ITSs can provide interactive tools that students can use to construct representations while receiving assistance and feedback. Such support for learning with interactive graphical representations can enhance learning outcomes [10], in particular for students with low spatial skills [11]. Second, ITSs have the capability to provide individualized support that adapts to student characteristics [12]. Adapting instructional support to the individual student's spatial skills has been shown to improve their spatial skills [13] as well as their learning of content knowledge [14].

However, before we can design ITSs that tailor support for using interactive representations to the needs of students with low spatial skills, we first have to understand what makes this learning task difficult for these students. This paper presents a first step towards this goal. Specifically, this paper investigates the following two questions: (1) Which aspects of problem solving with interactive graphical representations are more difficult for students with low spatial skills than for students with high spatial skills? (2) Which of these difficulties explain why students with

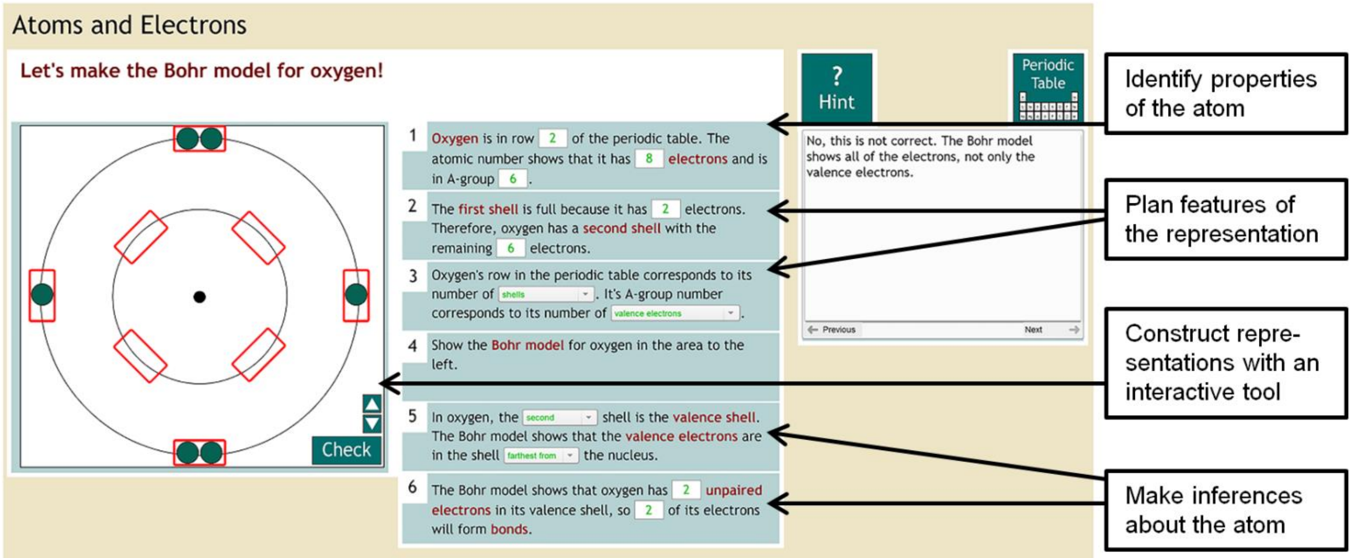

Figure 2. Example screen shot of a tutor problem: students construct a Bohr model of oxygen.

low spatial skills have lower learning outcomes in chemistry? To address these questions, I conducted a mediation analysis that tested which aspects of students' problem-solving performance account for the effect of spatial skills on learning outcomes. The mediation analysis was carried out with a data set obtained from an experiment with an ITS for chemistry learning in which students had to use interactive tools to construct graphical representations of atoms.

2. CHEM TUTOR

The data set used in this paper was obtained from an experiment with Chem Tutor: an ITS for undergraduate chemistry [15]. The goal of Chem Tutor is to enhance learning by helping students understand graphical representations of abstract concepts [16]. Chem Tutor targets foundational concepts of introductory undergraduate courses, such as atomic structure and bonding. The design of Chem Tutor is based on surveys with undergraduate chemistry students and instructors, interviews and eye-tracking studies with undergraduate and graduate students, and extensive pilot testing in the lab and the field [15]. Chem Tutor was built with Cognitive Tutor Authoring Tools [17], which facilitates rapid iterations of prototyping and pilot-testing involved in such user-centered design approaches.

In the present experiment, students worked with the atoms and electrons unit of Chem Tutor. This unit features interactive tools that students use to construct a variety of graphical representations of atoms: Lewis structures, Bohr models, energy diagrams, and orbital diagrams (see Figure 1). The tutor problems are structured as follows. First, students are prompted to think about the properties of the atom. They can use the periodic table to look up information about the atom (e.g., oxygen has eight electrons). Second, students are prompted to plan what the given representation will look like (e.g., the Bohr model of oxygen should show two shells). Third, students use an interactive tool to construct the representation of the given atom. Students receive error-specific feedback on their interactions (e.g., "The Bohr model shows all of the electrons, not only the valence electrons"). Students have to construct a correct graphical representation before they can continue. Fourth, students are prompted to make inferences from the given graphical representation about the atom (e.g., the number of valence electrons allow to approximate the number of bonds the atom forms). Figure 2 shows an example tutor problem in which students construct the Bohr model of an oxygen atom. The interface of the problems builds up step-by-step, as shown in Figure 3.

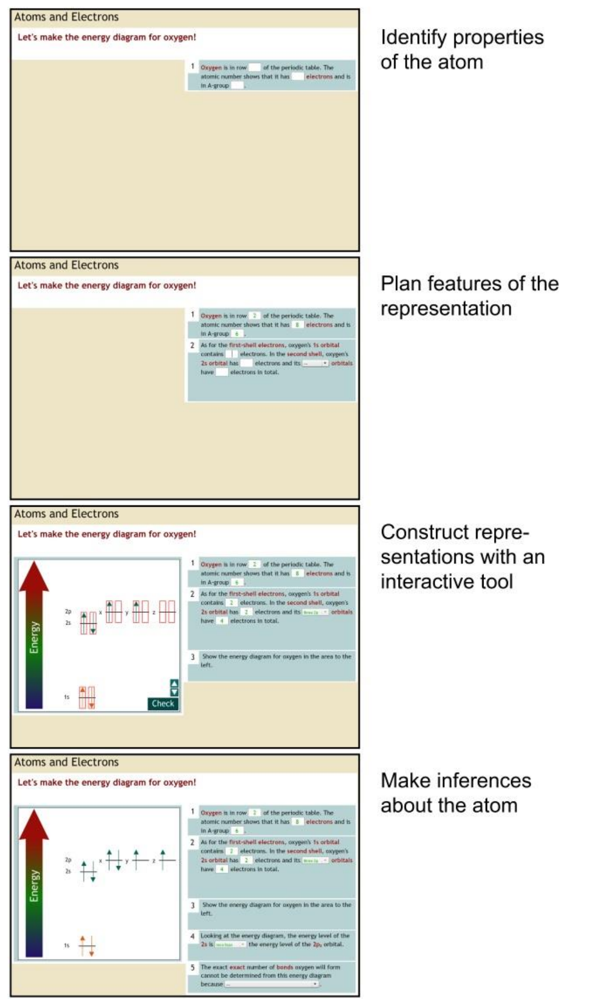

Figure 3. Sequence of screen shots showing how the interface updates step by step as students construct an Energy diagram.

3. EXPERIMENT
The experiment investigated whether Chem Tutor helps undergraduate students learn chemistry. For a detailed description of the experiment, refer to [18].

3.1 Participants
117 undergraduate students from a university in the mid-western United States participated in the experiment. 79% of the students were enrolled in general chemistry for non-science majors. According to the instructor of this course, these students had no experience with the graphical representations used in the Chem Tutor unit, with the exception of the common Lewis structure. 13.4% of the students were enrolled in general chemistry for science majors, 2.5% were enrolled in advanced general chemistry. According to the instructors of these courses, these students had experience with all graphical representations used in the Chem Tutor unit. The remaining 5% of the students were not currently enrolled in a chemistry course.

3.2 Assessments
Students' chemistry knowledge was assessed three times: before they started working with Chem Tutor (pretest), after they completed half of the tutor problems (intermediate posttest), and after they completed all tutor problems (final posttest). Three isomorphic test forms were used: they asked structurally identical questions but used different problems (e.g., with different atoms). The order in which students received the test forms was counterbalanced. The tests assessed reproduction and transfer of the chemistry content covered in Chem Tutor. Reproduction items used a format similar to the Chem Tutor problems. Transfer items asked students to apply the knowledge Chem Tutor covered in ways they had not been asked to do in the Chem Tutor problems. The tests included items with and without representations. In addition, spatial skills were assessed with the Vandenberg & Kuse mental rotation ability test [19]. This test presents students with a drawing of an object and asks them to identify which of four other drawings show the same object. This task requires spatial skills because students have to mentally rotate the given object to align it with the comparison objects. This test was chosen because it has been used in prior research on the impact of students' spatial skills on STEM learning [1, 2, 4, 5, 7].

3.3 Procedure
The experiment took place in the laboratory and involved two sessions of about 90 minutes each. Sessions were scheduled no more than three days apart. In session 1, students first completed the mental rotation test and the chemistry pretest. They then received an introduction into using Chem Tutor. Next, they worked through half of the problems in Chem Tutor's atoms and electrons unit. At the end of session 1, students took the intermediate chemistry posttest. In session 2, students worked through the remainder of the tutor problems. At the end of session 2, they took the final chemistry posttest. All students worked on the tutor problems at their own pace and were able to finish the assigned tutor problems in the available time.

3.4 Results
Results from the analysis of the test data show that there were significant learning gains on the chemistry knowledge test, $F(2,230) = 6.18$, $p < .01$. A regression of students' spatial skills on learning gains (i.e., performance on the posttest, controlling for pretest performance) showed that spatial skills were a significant predictor of learning gains ($\beta = .34$, $p < .01$), such that students with high spatial skills showed higher learning gains than students with low spatial skills.

4. OPEN QUESTIONS
The finding that students with lower spatial skills had lower learning gains as the result of an intervention that relies on graphical representations is not surprising: it aligns with prior research on the role of spatial skills in STEM learning [1, 4, 5, 9]. It is conceivable that working with interactive graphical representations requires students to make sense of how abstract properties of atoms can be translated into visuo-spatial elements of graphical representations. It is well documented that this is more difficult for students with lower spatial skills [1, 4, 5, 9].

A first question that remains thus far unanswered, however, is how these difficulties affect how students interact with tutor problems. There are several aspects of the problems in Chem Tutor that may be more difficult for students with low spatial skills. First, these students may struggle with the first part of the tutor problems: identifying properties of atoms. Students with low spatial skills may have trouble retrieving facts that describe properties of atoms because they cannot imagine what an atom looks like. They might also struggle in using resources such as the periodic table to retrieve this information. Second, students with low spatial skills may struggle with the planning part of the tutor problems, because this step requires them to think about how properties of an atom can be visualized. Third, it is possible that these students struggle more when constructing graphical representations because they have to translate text-based information into visuo-spatial elements of the graphical representations. Finally, it is possible that these students struggle more in using representations to make inferences about the atom because this requires them to imagine how the visualized properties determine dynamic behavior of electrons (e.g., electron movement) and of atoms (e.g., tendency to form bonds).

A second question that remains open is how these difficulties relate to learning gains. While it is possible that all of the aspects just described are more difficult for students with low spatial skills, some difficulties may play a larger role than others in explaining why these students show lower learning gains. Understanding which difficulties account for the fact that students with lower spatial skills show lower learning gains will enable us to provide more appropriate support for these students.

5. FEATURE SELECTION
To investigate why spatial skills predict students' learning gains as they work with interactive graphical representations, I used a structural equation model to conduct a mediation analysis. Structural equation models provide a unified framework to test mediation hypotheses, estimate total effects, and separate direct from indirect effects. The first step in constructing a structural equation model is to determine candidate mediator variables to be included in the model. To do so, I first investigated how best to represent the knowledge students acquire as they are working on the tutor problems by comparing different knowledge component models. Second, I used the knowledge component model to generate a number of features that describe student performance during problem solving. Third, I searched for features that are predictive of learning outcome, using linear regressions.

5.1 Knowledge component model
First, I constructed a knowledge component model that adequately describes knowledge students acquire when working with interactive representations to learn about atomic structure. Knowledge components are "acquired units of cognitive function or structure that can be inferred from performance on a set of related tasks" [19]. I contrasted the following knowledge component models:

1. A *single-step baseline model* that treats all problem-solving step as one skill;
2. A *step-type model* that does not distinguish between the graphical representation used in the given problem but distinguishes between step types (i.e., providing information about atoms, planning the graphical representation of the atom, constructing graphical representations, and making inferences about the atom; see Figures 2 and 3);
3. A *representation-construct model* that distinguishes between the graphical representation used in the given problem (i.e., Lewis structure, Bohr model, energy diagram, and orbital diagram; see Figure 1) for the step in which students are asked to construct the graphical representation, but that does not distinguish between graphical representations for the remaining step types;
4. A *step-type / representation model* that distinguishes between the graphical representation used in the given problem for each step types except for providing information about atoms.

Each model was evaluated as to how well it predicts student behavior during problem solving. Following standard practice in ITS research [19, 20], I considered each step in a given tutor problem as a learning opportunity for the particular knowledge component involved in the step. Student behavior was assessed based on whether a student solved the step correctly (i.e., without hints and without errors). To evaluate model fit, I used the Additive Factors Model (AFM) in the PSLC DataShop [20]. As a metric for model fit, I used 3-fold item-stratified cross validation [21]. Table 1 shows the root mean squared errors (RMSEs) for each knowledge component model. The *step-type / representation* model had the best model fit. Hence, this knowledge component model was used as a basis to generate features that describe students' learning about atomic structure with interactive graphical representations.

Table 1. RMSEs for knowledge component models.

Knowledge component model	**Knowledge components**	**Item-stratified RMSE (lower is better)**
Single-step baseline model	1	0.464794
Step-type model	4	0.375733
Representation-construct model	7	0.372553
Step-type / representation model	13	0.363908

5.2 Feature generation

Based on the step-type / representation model, I generated features that describe how students interact with the tutor problems. Students' problem-solving behaviors can be described based on the outcome (proportion of incorrect first attempts, proportion of hint requests at the first attempt, proportion of total incorrect attempts, proportion of total hint requests) and based on durations (time spent per step in total, time spent on steps with first correct attempt / steps with at least one incorrect attempt, time spent before first attempt, time spent before first attempt if it was a correct / incorrect attempt). Additionally, when students use an interactive tool (e.g., to construct representations) they can make a large variety of errors. Thus, the number of different error types when constructing representations is another measure of interest. To generate features, I computed these metrics for each knowledge component, yielding a total of 134 features (i.e., four outcome-based and six duration-based set of metrics for each of the 13 KCs, plus number of mistake types for constructing each of the four representations).

5.3 Search for predictive features

Since it is impractical to include all 134 features in a structural equation model, it was necessary to narrow down the number of features to consider. The most interesting features when investigating the role of spatial skills on learning outcomes are those features that are predictive of students' learning outcomes. To find predictive features, I conducted linear regressions on each set of features (i.e., proportion of correct steps, time spent on correct steps, etc.), computed for the given KCs. It was necessary to conduct separate regressions for each set of feature because the feature sets are not independent of one another. For example, the total incorrect attempts subsume the first incorrect attempts. Learning outcomes on the final posttest was the dependent variable in each linear regression model. Pretest performance was included as a predictor in all regression models. Regressions were conducted using 10-fold cross-validation. I used the results from the regression analyses to determine what characterizes predictive features. To do so, I compared the standardized coefficients and significance of features based on the metric they used and based on the KC they described. Table 2 shows the results for the regression analyses.

The goal of the selection procedure was to identify a set of predictive features that are independent of one another. Overall, features based on *knowledge components* related to planning, constructing, and making inferences were predictive of learning outcomes. However, features based on retrieving information about atoms were not predictive of learning outcomes. Thus, atoms steps were excluded from further analysis. Among the *outcome-based features*, those using proportion of incorrect first attempts and those using proportion of total incorrect attempts were equally predictive of learning outcomes. However, when excluding atoms steps, the features based on proportion of incorrect total attempts were slightly more predictive than those based on incorrect first attempts. Thus, features based in incorrect total attempts were selected for further analysis. Features based on proportion of hint requests at first attempt and proportion of total hint requests had low predictive value because hint use was generally low. Thus, these features were excluded. Features describing error types while constructing representations had high predictive value. Thus, these features were selected for further analysis. Among the *duration-based features*, those based on time spent on steps with at least one incorrect attempt as a metric were selected because they were more predictive than the other duration-based features.

Based on these findings, the following variables were selected for the structural equation model:

- Average duration of planning steps with at least one incorrect attempt (plan_timeError)
- Average duration of representation-construction steps with at least one incorrect attempt (repr_timeError)
- Average duration of inference steps with at least one incorrect attempt (infer_timeError)
- Proportion of total incorrect attempts on planning steps (plan_incorrect)
- Proportion of total incorrect attempts on representation-construction steps (repr_incorrect)
- Proportion of total incorrect attempts on inference steps (infer_incorrect)
- Number of error types on representation-construction steps (repr_errorTypes)

Table 2. Standardized coefficients for mediators in regression models, using color gradients to illustrate the strength of association with performance on the final posttest.

predictor	outcome-based features			duration-based features					
	total incorrects	incorrect 1st attempt	error-Types	total step duration	correct step duration	error step duration	before 1st attempt	before 1st correct	before 1st error
pretest	0.275	0.281	0.307	0.364	0.356	0.258	0.334	0.372	0.305
atom	-0.002	-0.027		0.013	0.009	-0.076	0.006	-0.007	-0.054
planning-Bohr	0.112	0.082		-0.137	0.018	-0.039	0.068	0.024	0.016
planning-Energy	-0.393	-0.112		-0.163	-0.001	0.230	0.075	0.036	0.025
planning-Lewis	-0.116	-0.114		-0.025	-0.006	-0.048	-0.118	-0.093	-0.046
planning-Orbital	0.018	0.112		-0.004	0.112	-0.118	-0.066	0.07	-0.071
construct-Bohr	-0.028	0.230	-0.201	-0.080	-0.053	-0.050	0.031	-0.103	0.062
construct-Energy	-0.030	-0.174	-0.093	0.269	-0.144	0.003	0.087	-0.086	-0.155
construct-Lewis	0.203	-0.053	-0.169	-0.109	0.025	-0.113	-0.077	0.029	-0.158
construct-Orbital	-0.028	-0.119	0.139	0.056	0.045	-0.166	-0.211	0.064	-0.202
inference-Bohr	-0.030	0.011		-0.080	-0.138	-0.114	-0.017	-0.046	0.059
inference-Energy	-0.121	-0.064		0.269	0.196	0.091	0.121	0.064	0.116
inference-Lewis	0.071	0.040		0.169	0.013	-0.093	-0.023	0.025	0.053
inference-Orbital	-0.140	-0.147		-0.107	-0.044	-0.106	0.075	0.039	0.010
Average of absolute values	0.112	0.112	0.182	0.132	0.083	0.108	0.094	0.076	0.095

6. STRUCTURAL EQUATION MODEL

The goal of the structural equation model was to investigate why students with low spatial skills show lower learning gains. The structural equation model allows testing whether students' problem-solving behaviors statistically mediate the effect of spatial skills on learning gains. To carry out this analysis, I considered the variables that I identified as predictive of students' learning outcomes as potential mediators of the effect of spatial skills on learning outcomes at the final posttest, controlling for pretest.

6.1 Model Search

Since there are many models that might describe the nature of the effect of spatial skills on learning outcomes, I conducted a model search. Because a factor analysis indicated that the chemistry content pretest and the mental rotation ability test load onto separate factors that correlate weakly, I assumed that pretest and spatial skills are independent. I assumed that pretest is prior to the mediators and to the final posttest, that spatial skills are prior to the mediators and to the final posttest, and that mediators are prior to the final posttest. For the mediators, I assumed that planning is prior to constructing representations, which is prior to making inferences. Even under these constraints, there are at least 2^{49} distinct models that are consistent with these assumptions. Figure 4 shows the fully saturated model that would be compatible with these assumptions. A fully saturated model contains all possible edges (or "effects") compatible with the assumptions. Therefore, Figure 4 illustrates the search space of models: the search was conducted among models that had all, none, or a subset of the edges in the fully saturated model.

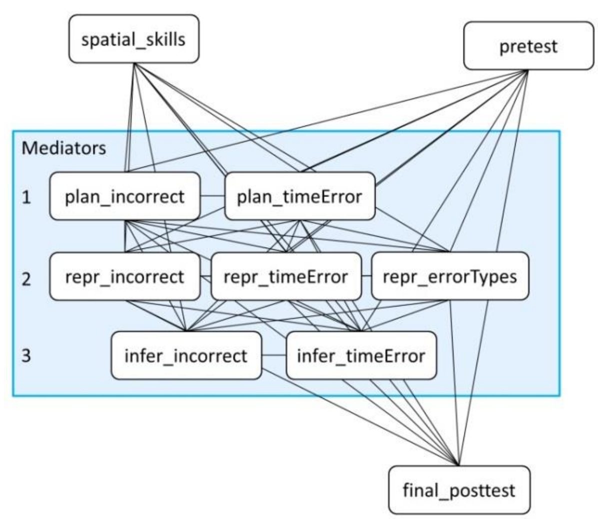

Figure 2. Fully saturated model consistent with the assumptions. Mediators are highlighted in blue and organized by tiers (1 = planning; 2 = representation-construction, 3 = inference).

To search for models that are theoretically plausible and consistent with the data, I used the Tetrad V program's[1] GES algorithm along with background knowledge constraining the space of models searched [22] to those that are theoretically tenable and compatible with my assumptions [23]. In the model search, each edge shown in Figure 4 is evaluated as to whether including it yields a better model fit than not, and whether it is a statistically reliable effect. As Figure 4 illustrates, there are many distinct models consistent with the background knowledge and that are plausible tests for the mediation hypothesis. Yet, it is important to know which of these models fits the data best, because parameter estimates and the statistical inferences we make about them are conditional on the model being true. Parameter estimates of models that do not fit the data well are scientifically unreliable. Thus, searching for the model that is most consistent with the data ensures that the parameters of the model can be trusted.

To conduct the model search at a technical level, I represented the qualitative causal structure of each model by a Directed Acyclic Graph (DAG). If two DAGs entail the same set of constraints on the observed covariance matrix,[2] then they are empirically indistinguishable. If the constraints considered are independence and conditional independence, which exhaust the constraints entailed by DAGs among multivariate normal varieties, then the equivalence class is called a *pattern* [23, 24]. The GES algorithm is asymptotically reliable,[3] and outputs the *pattern* with the best BIC score.[4] The pattern identifies features of the causal structure that are distinguishable from the data and background knowledge, as well as those that are not. The algorithm's limits lie primarily in its background assumptions involving the non-existence of unmeasured common causes and the parametric assumption that causal dependencies can be modeled with linear functions. The outcome of the model search is a structural equation model model that (1) is theoretically plausible, (2) fits the data well, and (3) contains only edges that describe statistically reliable effects.

6.2 Results

Figure 5 shows a model found by GES, with unstandardized parameter estimates. Table 2 shows standardized parameter estimates. Each edge is evaluated as to whether it is a reliable effect using *t*-tests, assuming an alpha-level of .05. A Bonferroni correction of the *p*-values is not necessary in a structural equation model because the significance tests are not independent. Table 2 shows the results from these tests. Altogether, the model fits the data well[5] (χ^2 = 32.77, df = 27, p = .21).

[1] Tetrad, freely available at www.phil.cmu.edu/projects/tetrad, contains a causal model simulator, estimator, and over 20 model search algorithms, many of which are described and proved asymptotically reliable in [24].

[2] An example of a testable constraint is a vanishing partial correlation, e.g., $\rho_{XY.Z} = 0$.

[3] Provided the generating model satisfies the parametric assumptions of the algorithm, the probability that the output equivalence class contains the generating model converges to 1 in the limit as the data grows without bound. In simulation studies, the algorithm is quite accurate on small to moderate samples.

[4] All the DAGs represented by a pattern will have the same BIC score, so a pattern's BIC score is computed by taking an arbitrary DAG in its class and computing its BIC score.

[5] The usual logic of hypothesis testing is inverted in path analysis: a *low* p-value means the model can be rejected.

Table 3. Parameter estimates (PE) for all edges and result of *t*-tests assessing whether the PE is significantly different from 0.

Edge from…	**to…**	**PE**	***t***	***p***
infer_timeError	infer_incorrect	.0124	3.2999	.0013
plan_incorrect	final_posttest	-.1116	-2.4706	.0150
plan_incorrect	infer_incorrect	.3704	6.3202	< .001
plan_incorrect	plan_timeError	4.6959	3.7759	< .001
plan_incorrect	repr_incorrect	3.0573	9.5622	< .001
plan_incorrect	repr_timeError	19.8158	2.8253	.0056
plan_timeError	infer_incorrect	-.0079	-2.2244	.0281
plan_timeError	infer_timeError	.2706	3.1104	.0024
plan_timeError	repr_timeError	1.8936	4.7591	< .001
pretest_content	final_posttest	0.2293	2.9336	.0040
pretest_content	plan_incorrect	-.4975	-3.1908	.0018
repr_incorrect	infer_incorrect	.0329	2.6633	.0088
repr_incorrect	repr_timeError	11.068	7.529	< .001
repr_timeError	infer_timeError	.0394	3.1303	.0022
repr_timeError	repr_errorTypes	.0083	8.6891	< .001
spatial_skills	final_posttest	.147	1.9078	.0589
spatial_skills	plan_incorrect	-.4102	-2.6326	.0096
spatial_skills	plan_timeError	-3.5242	-1.639	.1039

The final model shows that spatial skills have a direct positive effect on students' learning outcomes at the final posttest. Furthermore, spatial skills predict students' problem-solving behaviors while they are planning the graphical representation, which, in turn, has an effect on outcome-based and duration-based measures of problem-solving behaviors while they construct the graphical representation and while they make inferences from graphical representations about domain-relevant concepts. Only the proportion of incorrect attempts on planning steps mediates the effect of spatial skills on learning outcomes: plan_incorrect is the only variable that mediates the effect of spatial_skills on final_posttest. The edge from spatial_skills to plan_incorrect shows that a student with a perfect score on the spatial skills test makes .4102 fewer incorrect attempts per step than a student with the lowest possible score on the spatial skills test. The edge from plan_incorrect to final_posttest means that a student who makes one incorrect attempt per step scores 11.16% lower on the final posttest than a student who makes no incorrect attempts (controlling for pretest performance). In sum, the mediated effect of spatial_skills to final_posttest through plan_incorrect is .4102 * .1116 = .0458. Incorrect attempts while planning representations only partially mediate the effect of spatial skills on learning outcomes, because there is a direct effect of .147 from spatial_skills to final_posttest. Yet, making more incorrect attempts while planning graphical representations explains a considerable portion (about 25%) of the effect of spatial skills on learning outcomes.

7. CONCLUSIONS

The goal of the mediation analysis was to investigate (1) which aspects about working with interactive representations are harder for students with low than with high spatial skills and (2) which of these aspects explain why students with low spatial skills show lower learning gains than students with high spatial skills. With respect to the first question, results show that spatial skills have an effect on all aspects of students' problem-solving behaviors,

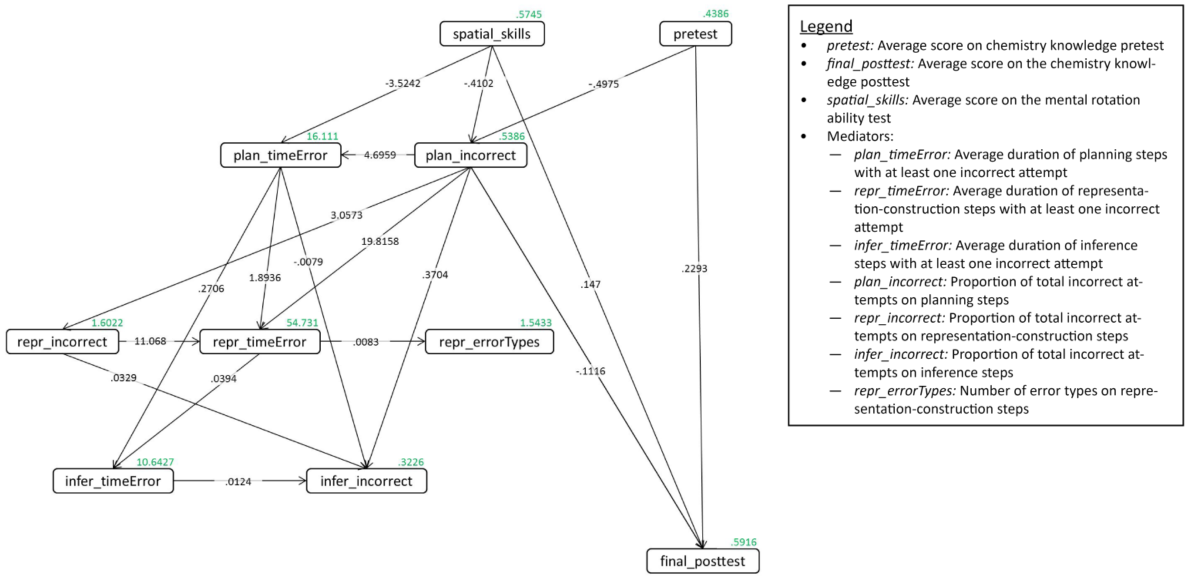

Figure 3. Final structural equation model with unstandardized parameter estimates. Green values show means.

except for looking up information about the atoms: planning, constructing, and making inferences from graphical representations. Spatial skills affect outcome-based measures of performance as well as duration-based measures of performance. Yet, the structural equation model shows that planning has a central role: students' ability to plan graphical representations has an impact on all further problem-solving behaviors as students construct graphical representations and make inferences about domain-relevant concepts based on the graphical information. With respect to the second question, results show that planning is the only aspect that mediates the effect of spatial skills on learning gains. The difficulties that students with low spatial skills have in constructing representations and in making inferences may merely be symptomatic—they do not explain why these students show lower learning gains. Only the fact that students with low spatial skills tend to struggle more in planning representations explains why they benefit less from interactive representations.

Why might students' ability to plan graphical representations be so strongly affected by their spatial skills? Planning a representation requires students to describe what the representation should look like, based on the properties of the atom. This task requires them to mentally picture visuo-spatial features based on text-based information about the atom's properties. This takes more cognitive effort for students who struggle with such visuo-spatial tasks. Hence, these students are at risk of cognitive overload during planning, which jeopardizes learning. Perhaps difficulties in planning are amplified by the fact that the interactive representation tool is not visible during the planning step (see Figure 3).

Why might the ability to plan representations determine students' learning gains? Learning with graphical representations means that students have to visualize new information externally while integrating this information with their internal mental models of the domain-relevant concepts [26]. Planning might play a central role because it helps students organize their initial mental model of the domain-relevant concepts. Having a well-organized initial mental model might facilitate integration of new information into this model: learning occurs as students expand and repair their mental models throughout the learning intervention, for instance by self-explaining how the new information relates to their initial mental models [27].

In summary, the findings from the mediation analysis shed light into the broader theoretical question of how spatial ability affects learning outcomes in STEM. Spatial skills seem to be important because students' benefit from interactive representations depends on their ability to mentally visualize abstract concepts *before* they use an external representation to visualize the concept. Mental visualization may play a key role in students' learning of abstract concepts because it allows students to integrate new information into their mental models. These findings also yield new hypotheses about the practical question of how best to support students with low spatial skills. These students might benefit from receiving additional assistance in planning graphical representations. They might benefit from seeing the interactive representation tool during the planning steps, so that they can more easily visualize the representation. They may also benefit from receiving examples of successful planning. It would be interesting to investigate whether such support increases learning gains for students with low spatial skills. In light of the interpretation that planning is so important because it helps students organize their initial mental models, it would be interesting to conduct a think-aloud study to assess whether, indeed, helping students plan representations facilitates mental model integration.

Several limitations of the present analysis need to be discussed. First, performance on planning steps only partially mediates the effect of spatial skills on learning outcomes. Thus, there might be other mediators that we did not assess. Further research is needed to investigate other aspects of problem solving that explain why students with low spatial skills tend to show lower learning gains. Second, the data is correlational: it is impossible to randomly assign students to having "low" or "high" spatial skills. As in any correlational data set, there may be other unknown factors that affect the effects of interest. Third, the structural equation model assumes linear relations between the variables in the model. This assumption is reasonable but not infallible. Finally, the analysis is based on a sample of 117 students. Even though that is sizable compared to many ITS studies, model search reliability increases with sample size, but decreases with model complexity. Hence, it is impossible to put confidence bounds on finite samples [21].

To conclude, the mediation analysis presented in this paper yields new insights into why students with lower spatial skills struggle in learning with interactive graphical representations. It seems that planning representations is a crucial aspect of learning success. This finding yields new hypotheses about what types of interventions these students may benefit from. Even though the present paper merely presents a first step towards better understanding the mechanisms that underlie the “the-rich-get-richer” rule in STEM domains, it may help us address the unfortunate fact that students with low spatial skills tend to show lower achievements in STEM domains and they are less likely to pursue careers in these domains. In other words, this paper is a first step towards creating an “everyone-gets-richer” rule for STEM learning.

8. ACKNOWLEDGMENTS

This work was supported by the UW-Madison Graduate School and WCER. We thank Teri Larson, Ned Sibert, Stephen Block, Amanda Evenstone, and Jocelyn Kuhn for their help with recruitment, and Sally Wu and the RAs in the Learning, Representations, & Technology Lab for their help in conducting the experiment.

9. REFERENCES

[1] Uttal, D.H., Meadow, N.G., Tipton, E., Hand, L.L., Alden, A.R., Warren, C., Newcombe, N.S.: The malleability of spatial skills: a meta-analysis of training studies. Psychological Bulletin 139, 352-402 (2013)

[2] Wai, J., Lubinski, D., Benbow, C.P.: Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. Journal of Educational Psychology 101, 817-835 (2009)

[3] Barnea, N., Dori, Y.J.: High-school chemistry students' performance and gender differences in a computerized molecular modeling learning environment. Journal of Science Education and Technology 8, 257-271 (1999)

[4] Stieff, M.: Sex differences in the mental rotation of chemistry representations. Journal of Chemical Education 90, 165-170 (2013)

[5] Stieff, M.: Mental rotation and diagrammatic reasoning in science. Learning and Instruction 17, 219-234 (2007)

[6] Kozma, R., Russell, J.: Students becoming chemists: Developing representationl competence. In: Gilbert, J. (ed.) Visualization in science education, pp. 121-145. Springer, Dordrecht, Netherlands (2005)

[7] Stieff, M., Hegarty, M., Deslongchamps, G.: Identifying representational competence with multi-representational displays. Cognition and Instruction 29, 123-145 (2011)

[8] Griffiths, A.K., Preston, K.R.: Grade-12 students' misconceptions relating to fundamental characteristics of atoms and molecules. Journal of Research in Science Teaching 29, 611-628 (1992)

[9] Hegarty, M., Waller, D.A.: Individual differences in spatial abilities. In: Shah, P., Miyake, A. (eds.) The Cambridge handbook of visuospatial thinking, pp. 121–169. Cambridge University Press, New York, NY (2005)

[10] Clements, D.H.: 'Concrete' Manipulatives, Concrete Ideas. Contemporary Issues in Early Childhood 1, 45-60 (1999)

[11] Ai-Lim Lee, E., Wong, K.W.: Learning with desktop virtual reality: Low spatial ability learners are more positively affected. Computers & Education ahead of print (2014)

[12] VanLehn, K.: The relative effectiveness of human tutoring, intelligent tutoring systems and other tutoring systems. Educational Psychologist 46, 197-221 (2011)

[13] Tuckey, H., Selvaratnam, M., Bradley, J.: Identification and rectification of student difficulties concerning three-dimensional structures, rotation, and reflection. Journal of Chemical Education 68, 460-464 (1991)

[14] Davidowitz, B., Chittleborough, G.: Linking the macroscopic and sub-microscopic levels: Diagrams. In: Gilbert, J.K., Treagust, D.F. (eds.) Multiple representations in chemical education, pp. 169-191. Springer, Dordrecht, Netherlands (2009)

[15] Rau, M.A., Michaelis, J.E., Fay, N.: Connection making between multiple graphical representations: A multi-methods approach for domain-specific grounding of an intelligent tutoring system for chemistry. Computers and Education 82, (2015)

[16] Ainsworth, S.: DeFT: A conceptual framework for considering learning with multiple representations. Learning and Instruction 16, 183-198 (2006)

[17] Aleven, V., McLaren, B.M., Sewall, J., Koedinger, K.R.: A new paradigm for intelligent tutoring systems: Example-tracing tutors. International Journal of Artificial Intelligence in Education 19, 105-154 (2009)

[18] Rau, M.A., Wu, S.P.W.: ITS support for conceptual and perceptual processes in learning with multiple graphical representations. submitted to AIED 2015 (under review)

[19] Peters, M., Laeng, B., Latham, K., Jackson, M., Zaiyouna, R., Richardson, C.: A Redrawn Vandenberg & Kuse Mental Rotations Test: Different Versions and Factors that affect Performance. Brain and Cognition 28, 39-58 (1995)

[20] Koedinger, K.R., Corbett, A.T., Perfetti, C.: The Knowledge-Learning-Instruction Framework: Bridging the Science-Practice Chasm to Enhance Robust Student Learning. Cognitive Science 36, 757–798 (2012)

[21] Koedinger, K.R., Baker, R., Cunningham, K., Skogsholm, A., Leber, B., Stamper, J.: A data repository for the EDM community: The PSLC Data-Shop. In: Romero, C. (ed.) Handbook of educational data mining, pp. 10-12. CRC Press, Boca Raton, FL (2010)

[22] Stamper, J.C., Koedinger, K.R., McLaughlin, E.A.: A Comparison of Model Selection Metrics in DataShop. In: D'Mello, S.K., Calvo, R.A., Olney, A. (eds.) Proceedings of the 6th International Conference on Educational Data Mining (EDM 2013), pp. 284-287. International Educational Data Mining Society (2013)

[23] Chickering, D.M.: Optimal Structure Identification with Greedy Search. Journal of Machine Learning Research 3, 507-554 (2002)

[24] Spirtes, P., Glymour, C., Scheines, R.: Causation, Prediction, and Search. MIT Press (2000)

[25] Pearl, J.: Causality: Models, Reasoning, and Interference. Cambridge University Press (2000)

[26] Bodner, G.M., Domin, D.S.: Mental models: The role of representations in problem solving in chemistry. University Chemistry Education 4, 24-30 (2000)

[27] Wylie, R. and Chi, M.T., 2014. The Self-Explanation Principle in Multimedia Learning. In *The Cambridge Handbook of Multimedia Learning*, R.E. Mayer Ed. Cambridge University Press, New York, NY, 413-43