Does Representational Understanding Enhance Fluency – Or Vice Versa? Searching for Mediation Models

Martina A. Rau
Human-Computer Interaction Institute
Carnegie Mellon University
5000 Forbes Ave
Pittsburgh, PA 15213, USA
marau@cs.cmu.edu

Richard Scheines
Department of Philosophy
Carnegie Mellon University
5000 Forbes Ave
Pittsburgh, PA 15213, USA
scheines@cmu.edu

Vincent Aleven
Human-Computer Interaction Institute
Carnegie Mellon University
5000 Forbes Ave
Pittsburgh, PA 15213, USA
aleven@cs.cmu.edu

Nikol Rummel
Institute of Educational Research
Ruhr-Universität Bochum
Universitätsstraße 150
44801 Bochum, Germany
nikol.rummel@rub.de

ABSTRACT
Conceptual understanding of representations and fluency in using representations are important aspects of expertise. However, little is known about how these competencies interact: does representational understanding facilitate learning of fluency (understanding-first hypothesis), or does fluency enhance learning of representational understanding (fluency-first hypothesis)? We analyze log data obtained from an experiment that investigates the effects of intelligent tutoring systems (ITS) support for understanding and fluency in connection-making between fractions representations. The experiment shows that instructional support for both representational understanding and fluency are needed for students to benefit from the ITS. In analyzing the ITS log data, we contrast the understanding-first hypothesis and the fluency-first hypothesis, testing whether errors made during the learning phase mediate the effect of experimental condition. Finding that a simple statistical model does not the fit data, we searched over all plausible causal path analysis models. Our results support the understanding-first hypothesis but not the fluency-first hypothesis.

Keywords

Causal path analysis modeling, multiple representations, intelligent tutoring systems.

1. INTRODUCTION
Representational understanding and *representational fluency* are important aspects of learning in any domain [1]. When working with representations (e.g., formulae, line graphs, path diagrams), students need conceptual understanding of these representations (representational understanding). Students also need to use the representations to solve problems fast and effortlessly (representational fluency). Science and mathematics instruction typically employs multiple graphical representations to help students learn about complex domains [2]. For instance, instructional materials for fractions use circle and rectangle diagrams to illustrate fractions as parts of a whole, and number lines to depict fractions in the context of measurement [3-5]. Multiple representations have been shown to lead to better learning than a single representation, provided that students make connections between them [6-7]: to benefit from the multiplicity of representations, students need to conceptually understand how different representations relate to one another, and they need to translate between them [8-11]. Yet, students find it difficult to make these connections [8], and tend not to make them spontaneously [12]. Therefore, they need to be supported in doing so [7]. Based on [1], we distinguish between *representational understanding* as conceptual understanding of connections between different graphical representations, and *representational fluency* as the ability to fast and effortlessly make these connections. To benefit from multiple graphical representations, students need to acquire both representational understanding [8], and they need to develop representational fluency [13].

In the present paper, we use log data obtained from a classroom experiment that uses a successful type of intelligent tutoring system (ITS) to help students learn about fractions while comparing different ways to support *representational understanding* and *representational fluency*. The experiment demonstrates that both instructional support for representational understanding and representational fluency are necessary for students to benefit from multiple graphical representations of fractions [14]. The goal of the present paper is to augment the findings from the traditional analysis of pretest and posttest data by using causal path analysis modeling to analyze mediation effects that can explain the nature of *how* representational understanding and representational fluency interact. Does representational understanding facilitate students' acquisition of representational fluency? Or does representational fluency enhance students' ability to acquire representational understanding? We contrast two competing hypotheses. According to the *understanding-first hypothesis*, representational understanding equips students with the necessary knowledge about structural correspondences between graphical representations and about what differences between the representations are incidental, allowing students to attend to relevant aspects of the graphical representations while developing representational fluency. Therefore, students who receive support for representational understanding should make fewer errors on fluency-building problems compared to students who do not receive support for representational understanding. By contrast, the *fluency-first hypothesis* predicts that representational fluency frees up the cognitive resources that students need to acquire understanding of these connections. Therefore, students who receive fluency-building support should make fewer errors on problems supporting representational understanding compared to students who do not receive fluency-building support. The answer to the question of how acquisition of representational understanding and representational fluency interact has important implications for the instructional design of ITSs and other educational technologies. If representational understanding enhances the acquisition of representational fluency (understanding-first hypothesis), instructional materials should support representational understanding before representational fluency. If, on the other hand, representational fluency facilitates students' acquisition of representational understanding(fluency-first hypothesis), instructional materials should support representational fluency before supporting representational understanding.

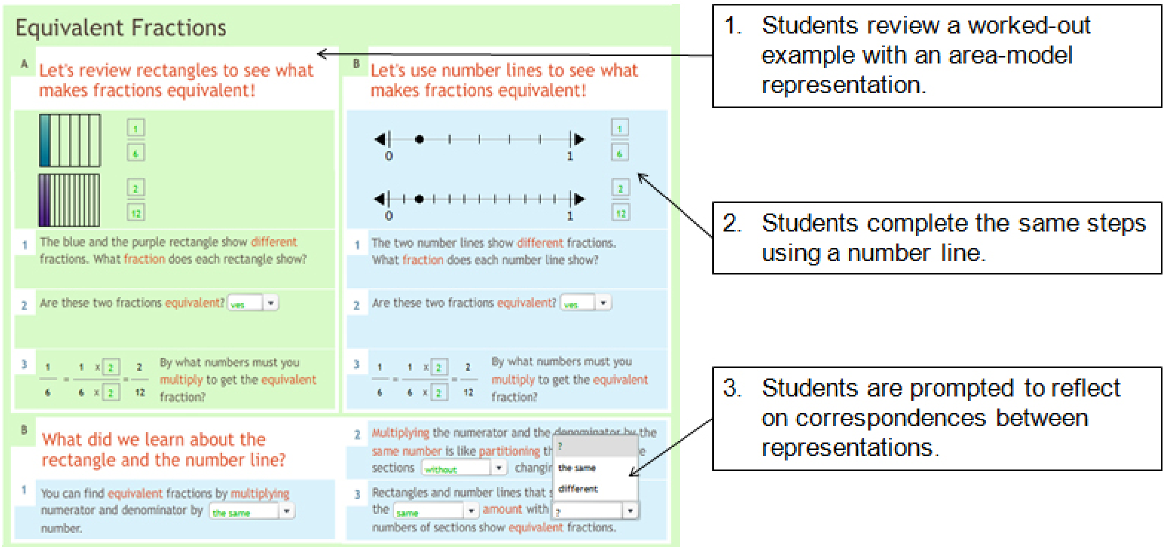

Figure 1. Worked example support for representational understanding: students use a worked example with a rectangle (part A, upper left) to guide their work on a fractions problem with a number line (part B, upper right). At the end (part C, bottom), students are prompted to integrate both representations by responding to drop-down menu questions.

To gain further insights into *how* support for representational understanding and representational fluency affect students' interactions with an ITS for fractions, we employ causal path analysis. In doing so, we contrast mediation models that correspond to the understanding-first hypothesis, and to the fluency-first hypothesis. Specifically, we investigate whether errors that students make during the learning phase mediate the interaction effect between support for representational understanding and representational fluency on students' learning. Our results are in line with the understanding-first hypothesis, but not with the fluency-first hypothesis.

The remainder of this paper is structured as follows. We first describe the ITS that we used to carry out the experimental study. We then provide a brief overview of the experimental design and the results obtained from the analysis of pretests and posttests. The main focus of this paper is on describing the causal path analysis we conducted to investigate the interaction of instructional support for representational understanding and representational fluency on students' learning behaviors as identified by the tutor log data. We end by discussing the implications of our analysis for the instructional design of learning materials, and by outlining open questions that future research should be address.

2. THE FRACTIONS TUTORING SYSTEM

The Fractions Tutor used in the experiment is a type of Cognitive Tutor. Cognitive Tutors are grounded in cognitive theory and artificial intelligence. Cognitive Tutors have been shown to lead to substantial learning gains in a number of studies [15]. We created the Fractions Tutor with Cognitive Tutor Authoring Tools [16]. The design of the tutor interfaces and of the interactions students engage in during problem solving are based on a number of small-scale user studies, a knowledge component model developed based on Cognitive Task Analysis of the learning domain [17], and a series of in vivo experiments [6, 12, 18].

The Fractions Tutor uses multiple interactive graphical representations (circles, rectangles, and number lines) that are typically used in instructional materials for fractions learning [2-3, 5]. The Fractions Tutor covers a comprehensive set of topics ranging from identifying fractions from graphical representations, to equivalent fractions and fraction addition. Taken together, the Fractions Tutor comprises about ten hours of supplemental instructional material. Students solve tutor problems by interacting both with fractions symbols and with the graphical representations. As is common with Cognitive Tutors, students receive error feedback and hints on all steps. In addition, each tutor problem includes conceptually oriented prompts to help students relate the graphical representations to the symbolic notation of fractions.

3. EXPERIMENT

The goal of the experimental study (cf. [14] for a detailed description) was to investigate the hypothesis that students learn more robustly when receiving instructional support for both representational understanding and support for representational fluency. We conducted a classroom experiment with 599 4^{th}- and 5^{th}-grade students from five elementary schools in the United States. Students worked with the Fractions Tutor for about ten hours during their regular mathematics class.

We contrasted two experimental factors. One factor, *support for representational understanding* in making connections had three levels: no support, auto-linked support in which the Fractions Tutor automatically made changes in one representation as students manipulated another, and worked examples. Figure 1 provides an example of the Fractions Tutor problem that uses worked examples (WEs) to support representational understanding. Students used a worked example with a familiar representation as a guide to make sense of an isomorphic problem with a less familiar representation. This factor was crossed with a second experimental factor, namely, whether or not students received *support for representational fluency* in making connections: students had to visually estimate whether different types of graphical representations showed the same fraction. Figure 2 shows an example of a fluency-building problem (FL). Students in all conditions worked on 80 tutor problems: eight problems per topic (e.g., equivalent fractions, addition, subtraction, etc.). In each topic, the first four tutor problems were single-representation problems (i.e., they included only a circle, only a rectangle, or only a number line, and no connection-making support). The last four tutor problems were multiple-representation problems and differed between the experi-

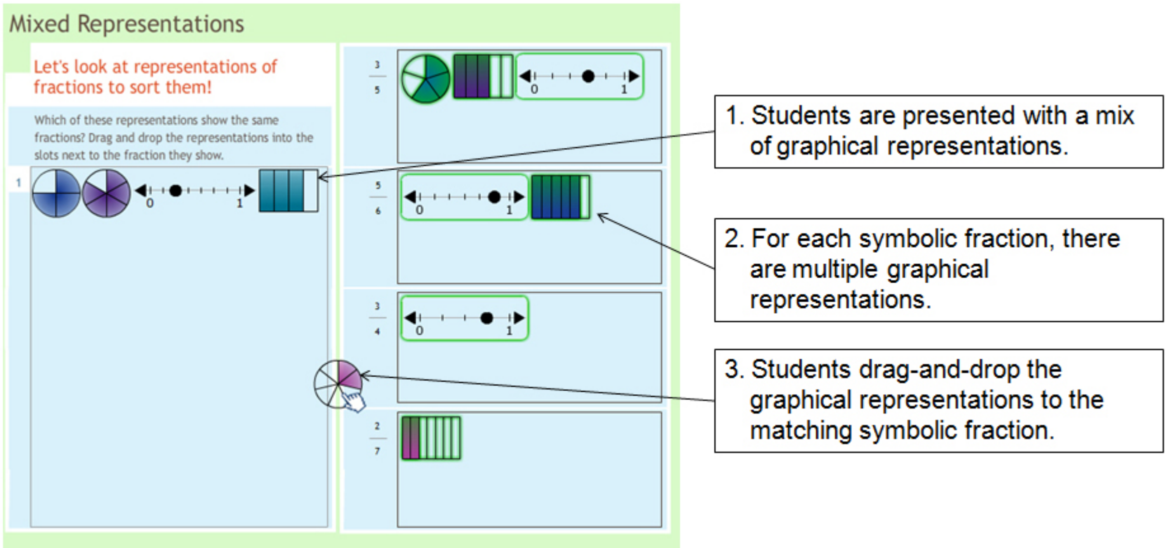

Figure 2. Fluency-building support: students sort graphical representations by dragging-and-dropping them into slots that show equivalent fractions.

mental conditions. For instance, students in the worked examples only condition (WE) received four worked examples problems. Students in the fluency-only condition (FL) received four fluency-building problems. Students in the worked examples plus fluency condition (WE-FL) received two worked examples problems, followed by two fluency-building problems. Table 1 illustrates this procedure for two consecutive topics for each of these three conditions. The same sequence of eight problems was repeated for each of the ten topics the Fractions Tutor covered.

Table 1. Problem sequence per condition: for each topic, problems 1-4 (P1-P4) are single-representation problems (S); problems 5-8 are multiple-representation problems: worked examples (WE, blue-underlined) or fluency-building problems (FL, green-italicized).

Cond.	Topic	P1	P2	P3	P4	P5	P6	P7	P8
WE	1	S	S	S	S	WE	WE	WE	WE
	2	S	S	S	S	WE	WE	WE	WE
	…	…							
FL	1	S	S	S	S	FL	FL	*FL*	*FL*
	2	S	S	S	S	FL	FL	*FL*	*FL*
	…	…							
WE-FL	1	S	S	S	S	WE	WE	*FL*	*FL*
	2	S	S	S	S	WE	WE	*FL*	*FL*
	…	…							

Results based on the analysis of pretest, immediate posttests, and delayed posttest (administered one week after the immediate posttest) from 428 students confirmed the hypothesis that a combination of instructional support for representational understanding and representational fluency is most effective: the interaction between support for understanding and fluency was significant, $F(2, 351) = 3.97$, $p < .05$, $\eta_p^2 = .03$, such that students who received both types of support performed best. Worked examples are the more effective type of support for representational understanding, when paired support for representational fluency: within the conditions with support for representational fluency, there was a significant effect of support for representational understanding, $F(2, 343) = 4.34$, $p < .05$, $\eta_p^2 = .07$. However, within the conditions without support for representational fluency, there was no significant effect of support for representational understanding ($F < 1$). Finally, our results show an advantage of the WE-FL condition over the number-line control, $t(115) = 2.41$, $p < .05$, $d = .27$.

The results from the experimental study raise interesting new questions about the relation between representational understanding and representational fluency. It is surprising that there were no significant main effects for support for representational understanding or representational fluency alone; only the combination of both enhanced students' learning from multiple graphical representations. Did support for understanding enable students to benefit from fluency-building support, or vice versa? We address this question in the remainder of this paper.

4. DATA SET

The analyses in this paper are based on the data obtained from the experimental study just described. Students in the experiment received a pretest on the day before they started to work with the Fractions Tutor. The day after students finished working with the Fractions Tutor, they received an immediate posttest. One week after the immediate posttest, students were given a delayed posttest. All three tests were equivalent (i.e., they contained the same items with different numbers). Students worked with the Fractions Tutor for about ten hours and had to complete each tutor problem. All interactions with the Fractions Tutor were logged.

4.1 Selecting Conditions to Include into Causal Path Analysis Modeling

In the light of the interaction effect between support for representational fluency and support for representational understanding through worked examples, the experimental conditions of interest for further analyses are worked example (WE), fluency (FL), and worked examples paired with support fluency (WE-FL). We thus selected these three conditions to include into the causal path analysis model. A total of 190 students were included in the analysis (n = 59 in the WE condition, n = 73 in the FL condition, and n = 58 in the WE-FL condition). Table 2 shows the means and standard deviation of students' performance on pretest, immediate and delayed posttest by condition.

4.2 Defining Mediation Variables

As the goal was to investigate whether support for representational understanding helps students benefit from support for representational fluency or vice versa, we compared students' *performance on worked-example problems* (i.e., support for representational understanding) between the WE and the WE-FL condition, and students' *performance on fluency-building problems* between the FL and the WE-FL condition. Specifically, we compared performance on those tutor problems that were *the same* across these pairs of conditions. To compare the WE and WE-FL conditions, we used errors students made on problems P5 and P6 (see the blue-underlined problems in Table1). To compare the FL and WE-FL conditions, we used errors students made on problems P7 and P8 (see the green-italic problems in Table1). We expect that, if representational understanding facilitates the acquisition of representational fluency, students in the WE-FL condition will make fewer errors on fluency-building problems than students in the FL condition. If representational fluency facilitates the acquisition of representational understanding, we expect the WE-FL condition to make fewer errors on worked-examples problems than students in the WE condition.

Table 2. Means and standard deviation (in parentheses) on pretests and posttests per condition.

Condition	Pretest	Immediate posttest	Delayed posttest
WE	.36 (.22)	.43 (.20)	.49 (.26)
FL	.31 (.21)	.37 (.22)	.44 (.24)
WE-FL	.39 (.21)	.52 (.24)	.58 (.26)

A first step in this analysis was to use the tutor log data to identify measures of errors that students made on these problems. Rather than using the overall error rate, we applied the knowledge component model [17] that underlies the problem structure of the Fractions Tutor to categorize the errors students made while working on the tutor problems. Doing so allows for a much more fine-grained analysis of students' errors than the overall error rate does. The knowledge component model describes a meaningful set of steps within a tutor problem which provide practice opportunities for practicing a "unit" of knowledge. For example, every time a student is asked to enter the numerator of a fraction, he/she has the opportunity to practice knowledge about what the numerator of a fraction is. Worked-example problems and fluency-building problems cover a different set of knowledge components, but the same knowledge components occur repeatedly across different worked example problems and fluency-building problems, respectively. Altogether, the knowledge component model led to 12 types of errors that students could make on worked-example problems, and 11 types of errors that students could make on fluency-building problems.

Next, we had to narrow the number of error categories to include in the causal path analysis model. We included only those error types which (1) were significant predictors of students' posttest performance, while controlling for pretest performance, and (2) significantly differed between conditions. To determine whether an error type was a significant predictor of students' immediate posttest performance, we conducted linear regression analyses with posttest performance as the dependent variable, and pretest performance and number of error type as predictors.

To determine whether error types differed significantly between conditions, we conducted Chi-square tests with number of error type as dependent variable and condition as independent variable (i.e., WE vs. WE-FL for error types that students could make on worked-example problems, and FL vs. WE-FL for error types that students could make on fluency-building problems). For both analyses, we adjusted for multiple comparisons using the Bonferroni correction. On worked-example problems, six error types differed significantly between conditions, but only two error types were significant predictors of posttest performance (both of them passed both the Chi-square test and the regression test). On fluency-building problems, eight error types differed significantly between conditions, and four were significant predictors of posttest performance (three of them passed both the Chi-square test and the regression test). Table 3 provides an overview of the error types we selected for further analyses.

Table 3. Selected error types and number of error-types per condition.

Error type	Description	# in WE	# in FL	# in WE-FL
place1Error	Locating 1 on the number line given a dot on the number line and the fraction it shows	150	n/a	222
SE-Error	Self-explanation error, response to reflection questions in drop-down menu format	132 0	n/a	1629
equivalen-ceError	Finding equivalent fraction representa-tions	n/a	289 9	2157
improper-MixedError	Finding representa-tions of improper frac-tions	n/a	138 0	1608
Name-Circle-MixedError	Finding circle repre-sentations that show the same fraction as a number line or a rec-tangle	n/a	355	126

5. PATH ANALYSIS MODELING

In order to investigate whether and how error types mediate the effect of condition, we first specified, estimated, and tested two path analytical structural equation models [19-20] – one which compared the WE and WE-FL conditions using error types made on the worked-example problems as mediators, and one which compared the FL and WE-FL conditions using error types made on the fluency-building problems as mediators. Structural equation models provide a unified framework within which to test mediation hypotheses, to estimate total effects, and also to separate direct from indirect effects. The models that represented our hypotheses in both experiments were decisively rejected by the data, and in such a case it is not appropriate to use the model to test mediation hypotheses or estimate effects. Our strategy was to use the Tetrad IV program[1] to search for alternative models that

[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 [23] Spirtes, P., Glymour, C. and Scheines, R. *Causation, Prediction, and Search.* MIT Press, 2000.

are both theoretically plausible and consistent with the data. In this section, we describe the path analytic models that represent our hypotheses, describe the search algorithms we use to find for alternative models, and briefly summarize the results of our search.

5.1 Modeling our Hypotheses

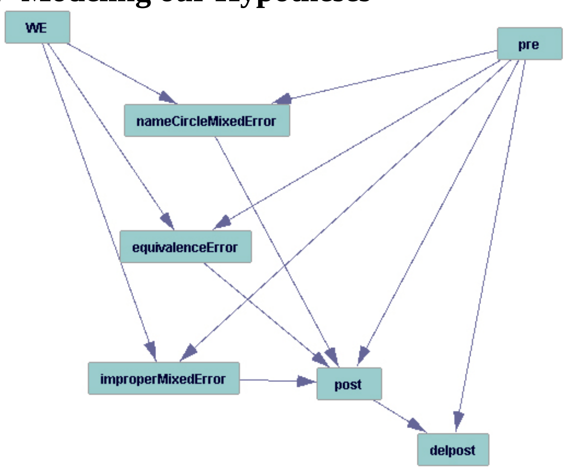

Figure 2. Path model for understanding-first hypothesis.

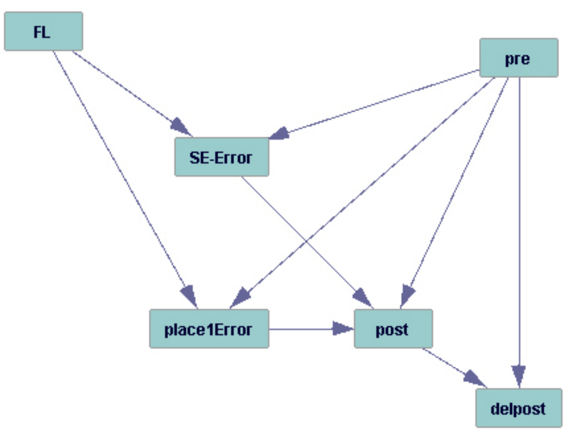

Figure 3. Path model for fluency-first hypothesis.

Our model hypotheses correspond to the understanding-first hypothesis and the fluency-first hypothesis described above. The understanding-first hypothesis predicts that support for representational understanding enhances students' ability to benefit from fluency-building problems by equipping students with the knowledge they need to attend to relevant features of the graphical representations while developing representational fluency. Therefore, students who receive support for representational understanding should make fewer errors on fluency-building problems compared to students who do not receive support for representational understanding. Therefore, the understanding-first hypothesis predicts that support for representational understanding increases learning by reducing the number of errors made on fluency-building problems. Figure 3[2] depicts the model we specified to test the understanding-first hypothesis. Each node in the path model refers to a variable in the data set: *WE* = whether or not students receive worked-example support for representational understanding (i.e., whether they are in the FL vs. in the FL-WE condition), *nameCircleMixedError*, *equivalenceError*, and *improperMixedError* being the errors students could make on fluency-building problems (see Table 3), *pre* = performance on the pretest, *post* = performance on the immediate posttest, *delpost* = performance on the delayed posttest.

The fluency-first hypothesis predicts that support for representational fluency enhances students' ability to benefit from support for representational understanding because representational fluency frees up the cognitive resources that students can invest in sense-making processes that lead to representational understanding. Therefore, students who receive support for representational fluency should make fewer errors on worked-example problems, compared to students who do not receive support for representational fluency. Therefore, the fluency-first hypothesis predicts that support for representational fluency increases learning by reducing the number or errors made on worked-example problems. Figure 4 depicts the model that we specified to test the fluency-first hypothesis. Each node in the path model refers to a variable in the data set: *FL* = whether or not students receive support for representational fluency (i.e., whether they are in the WE vs. in the FL-WE condition), *SE-Error* and *place1Error* being the errors students could make on worked-example problems (see Table 3), *pre* = performance on the pretest, *post* = performance on the immediate posttest, *delpost* = performance on the delayed posttest.

Using normal theory maximum likelihood to estimate the parameters of these models, we find that in each case the deviation between the estimated and the observed covariance matrix is too large to be explained by chance (for the model for the understanding-first hypothesis in Figure 3: $\chi^2 = 30.88$, df = 9, $p < .0001$[3], and for the model for the fluency-first hypothesis in Figure 4: $\chi^2 = 49.14$, df = 6, $p < .0001$), thus the models do not fit the data and the parameter estimates cannot be trusted[4].

5.2 Model Search

To search for alternatives, we used the GES algorithm in Tetrad IV along with background knowledge constraining the space of models searched [19] to those that are theoretically tenable and compatible with our experimental design. In particular, we assume that our intervention variables are exogenous, that our intervention variables are causally independent, that the pretest is exogen-

[2] In path models of this type, also called "causal graphs" [22] Ibid., each arrow, or directed edge, represents a direct causal relationship relative to the other variables in the model. For example, in Figure 3 the condition is a direct cause of the mediator variables, but only affect the posttest indirectly through these mediators.

[3] The usual logic of hypothesis testing is inverted in path analysis. The p-value reflects the probability of seeing as much or more deviation between the covariance matrix implied by the estimated model and the observed covariance matrix, conditional on the null hypothesis that the model that we estimated was the true model. Thus, a low p-value means the *model* can be rejected, and a high p-value means it cannot. The conventional threshold is .05, but like other alpha values, this is somewhat arbitrary. The p-value should be higher at low sample sizes and lowered as the sample size increases, but the rate is a function of several factors, and generally unknown.

[4] We also tested variations of these models in which we added direct paths from the condition variables to the post-test and delayed post-test. These variants are also clearly rejected by our data.

ous and causally independent of intervention, that the mediators are prior to the immediate posttest and to the delayed posttest, and that the immediate posttest is prior to the delayed posttest. Even under these constraints, there are at least 2^{25} (over 33 million) distinct path models for the understanding-first hypothesis, and 2^{25} (over 33 million) for the fluency-first hypothesis.

The qualitative causal structure of each linear structural equation model can be represented by a Directed Acyclic Graph (DAG). If two DAGs entail the same set of constraints on the observed covariance matrix,[5] then we say that 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* [20-21]. Instead of searching in DAG space, the GES algorithm achieves efficiency by searching in pattern space. The algorithm is asymptotically reliable,[6] and outputs the *pattern* with the best Bayesian Information Criterion (BIC) score.[7] 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 are 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.

5.3 Results

Figure 5 shows a model found by GES for the understanding-first hypothesis, with coefficient estimates included. The model fits the data reasonably well[8] (χ^2 = 16.10, df = 6, p = .013). Students with higher pretest scores make fewer nameCircleMixedErrors, and they perform better on the immediate and the delayed posttest. Receiving worked-example support for representational understanding (i.e., being in the WE-FL condition and not in the FL condition) increases nameCircleMixedErrors, which in turn decreases performance on the immediate posttest. In other words, nameCircleMixedErrors mediate a negative effect of worked examples on students' learning. Receiving worked-example support for representational understanding also reduces equivalenceErrors and improperMixedErrors. Since making more improperMixedErrors leads to worse performance on the immediate and the delayed posttests, equivalenceErrors and improperMixedErrors mediate the positive effect of the worked-example support on students' learning. Support for representational understanding through worked examples does not have a direct impact on students' posttest performance. The overall positive effect of worked examples on students' learning through equivalenceErrors and improperMixedErrors is larger than the negative effect through nameCircleMixedErrors. (See Table 3 for a description of the errors.)

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

[6] 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.

[7] 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.

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

Figure 6 shows a model found by GES for the fluency-first hypothesis. The model fits the data well (χ^2 = 8.32, df = 5, p = .14). Students with higher pretest scores make fewer SE-Errors and perform better on both posttests. Having fluency-building support (i.e., being in the WE-FL condition as opposed to being in the WE condition) increases SE-Errors, which reduces performance on the immediate and the delayed posttest. In other words, SE-Errors mediate a negative effect of fluency-building support. There are no further mediations of having fluency-building support, but there is a direct positive effect of fluency-building support on students' performance on the immediate posttest. (See Table 3 for a description of the errors.)

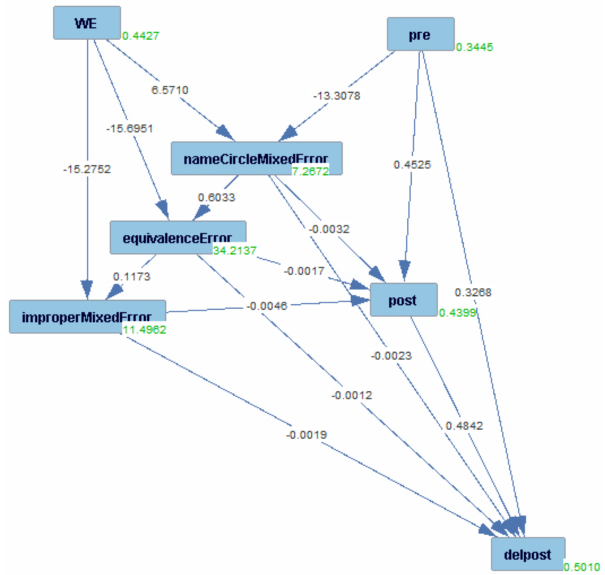

Figure 4. The model found by GES for the understanding-first hypothesis, with parameter estimates included.

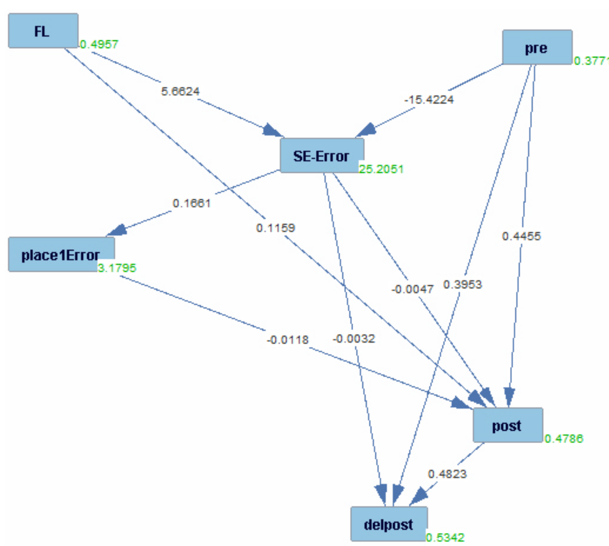

Figure 5. The model found by GES for the fluency-first hypothesis, with parameter estimates included.

6. DISCUSSION

Taken together, results from the causal path analysis models support the understanding-first hypothesis but not the fluency-first hypothesis: receiving worked-example support for representational understanding helps students learn from fluency-building problems. The model in Figure 5 demonstrates that, although students who receive worked-example support make more nameCircle-MixedErrors, they make fewer equivalenceErrors and improper-MixedErrors. NameCircleMixedErrors are possible early in the Fractions Tutor curriculum, whereas equivalenceErrors and improperMixedErrors occur later in the Fractions Tutor curriculum. The analysis therefore suggests that support for representational understanding reduces errors later during the learning phase, which leads to better overall learning. This finding is particularly interesting when we recall that we only compare the errors students make on fluency-building problems P7 and P8 (see Table 1). For the FL condition, problems P5 and P6 are also fluency-building problems, whereas for the WE-FL condition, problems P5 and P6 are worked-examples problems. That is, students in the FL condition receive more practice on fluency-building problems, which should *increase* their performance on fluency-building problems. Based on practice effects, we would thus expect that students in the FL condition would *outperform* students in the WE-FL condition on problems P7 and P8 (e.g., P7 is the first time the WE-FL condition encounters a fluency-building problem, but the third time the FL condition encounters a fluency-building problem, for the given topic). However, we find the opposite for errors that occur later in the curriculum: worked-example support for representational understanding leads to better performance on fluency-building problems, even compared to students who received more practice on the same types of fluency-building problems. Since higher performance on these problems (i.e., fewer equivalenceErrors and fewer improperMixedErrors) leads to better performance on the immediate posttest, while controlling for pretest, it seems that support for representational understanding prepares students to learn better from subsequent fluency-building problems.

The model in Figure 6 does not provide support for the fluency-first hypothesis. We do not find evidence that fluency-building support helps students benefit from support for representational understanding. Although we find a direct positive effect of fluency-building support on students' learning, the mediation effect shown in Figure 6 is evidence that receiving fluency-building support comes at the cost of lower performance on worked-examples problems: students tend to make more SE-Errors and more place1Errors. This finding is somewhat expected. Students in the WE condition work on twice as many worked-examples problems than the WE-FL condition, so they receive more practice on the worked-examples problems compared to the WE-FL condition (see Table 1). As students in the WE-FL condition have less practice on worked-examples problems, they are expected to perform somewhat worse on those problems – and that is what the model in Figure 6 confirms. Yet, since we do not find evidence that receiving fluency-building support also benefits students' learning from worked-example support for representational understanding, our results do not support the fluency-first hypothesis.

Our findings from path analysis modeling demonstrate the importance of model search. None of our initial hypothesis models fit the data, but there are thousands of plausible alternatives. Further, estimating path parameters with a model that does not fit the data is scientifically unreliable. Parameter estimates, and the statistical inferences we make about them with standard errors etc., are all conditional on the model specified being true everywhere except the particular parameter under test.

Even if our initial hypotheses had fit the data well, it would have been important to know whether there were alternatives that explained the same data. The GES algorithm implemented in Tetrad IV enabled us to find plausible models that fit the data well. The models we found in Figures 5 and 6 allow us to estimate and test path parameters free from the worry that the model within which the parameters are estimated is almost surely mis-specified, as is the case for the models in Figures 3 and 4.

Several caveats need to be emphasized, lest we give the false impression that we think we have "proved" the causal relationships that appear in the path diagrams shown in Figures 5 and 6. First, the GES algorithm assumes that there are no unmeasured confounders (hidden common causes), an assumption that is almost certainly false in this and in almost any social scientific case, but one that is routinely employed in most observational studies.[9] In future work, we will apply algorithms (e.g., FCI) that do not make this assumption, and see whether our conclusions are robust against this assumption. Second, although we did include intervention interaction in our model search and did test for interactions between pretest and mediators, by no means were our tests exhaustive, and by no means can we rely on the assumption that the true relations between the variables we modeled are linear, as the search algorithms assume. The assumption of linear relationships is reasonable but not infallible. Third, we have a sample of 190 students, and although that is sizable compared to many Cognitive Tutor studies, model search reliability goes up with sample size, but down with model complexity and number of variables, and is impossible to put confidence bounds on finite samples [23].

7. CONCLUSIONS AND FUTURE WORK

Our findings provide important insights into the nature of the interaction between students' acquisition representational understanding and representational fluency. Our analysis supports the notion that the acquisition of representational understanding enhances students' ability to benefit from instructional support for representational fluency, more so than the other way around. Therefore, our findings suggest that instruction should provide support for representational understanding *before* providing support for representational fluency.

Although our analyses provide support for the understanding-first hypothesis, but not for the fluency-first hypothesis, both remain valid hypotheses. One important caveat of the analyses presented here is that, within each curricular topic of the Fractions Tutor, all students in the WE-FL condition received support for representational understanding before support for representational fluency (i.e., there was no FL-WE condition). We therefore cannot draw definite conclusions about the relative effectiveness of providing support for representational understanding before support for representational fluency (WE-FL) and providing support for representational fluency before support for representational understanding (FL-WE). Our findings based on causal path analysis modeling merely suggest that the WE-FL condition would lead to better learning than a FL-WE condition. This notion remains to be tested empirically.

Future research should also investigate whether our findings are specific to the domain of fractions learning, and to the acquisition of representational understanding and representational fluency in making connections between multiple graphical representations. Graphical representations are universally used as instructional tools to emphasize and illustrate conceptually relevant aspects of the domain content. Furthermore, in any given domain, students

[9] Although our data are from a study in which we intervened on intervention, we did not directly intervene on our mediator or outcome variables. Thus these parts of our model are subject to the same assumptions as a non-experimental study.

need to develop representational fluency in using graphical representations to solve problems, and they need to effortlessly translate between different kinds of representations. But representational understanding and representational fluency are not limited to learning with graphical representations: representational understanding and representational fluency also play a role in using symbolic and textual representations. For example, should students acquire representational fluency in applying a formula to solve physics problems before understanding the conceptual aspects the formula describes, or should they first conceptually understand the phenomenon of interest and then learn to apply a formula to solve problems related to that phenomenon? This is a crucial question for instructional design and one that remains open. While the analysis presented in this paper takes an important step towards answering this question by providing novel insights into how representational understanding and representational fluency interact, more research is needed to investigate implications and applications related to the question of how best to support students to develop expertise with representational understanding and representational fluency.

The use of search algorithms over plausible causal path analysis models is a promising method to analyze the effects of instructional interventions on, because we can get insights into *how* an intervention affects problem-solving behaviors, and *how* these effects account for the advantage of one intervention over the other. Basing our analysis on cognitive task analysis and knowledge component modeling, we make use of common techniques in the analysis of tutor log data [23]. The results from our causal path analysis not only provide insights into the nature of the interaction of the experimental study, but also raise new hypotheses that can be empirically tested in future research. Thereby, our findings illustrate that causal path analysis modeling is a useful technique to augment regular tutor log data analysis.

8. ACKNOWLEDGEMENTS

We thank the NSF REESE-21851-1-1121307, the PSLC, funded by NSF award number SBE-0354420, Ken Koedinger, the Datashop team, and the students and teachers.

9. REFERENCES

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

[2] NMAP: Foundations for Success: Report of the National Mathematics Advisory Board Panel. U.S. Government Printing Office (2008).

[3] Cramer, K. Using models to build an understanding of functions. *Mathematics teaching in the middle school*, 6, 310-318 (2001).

[4] Charalambous, C.Y., Pitta-Pantazi, D.: Drawing on a Theoretical Model to Study Students' Understandings of Fractions. Educational Studies in Mathematics, 64, 293-316 (2007).

[5] Siegler, R. S., Thompson, C. A. and Schneider, M. An integrated theory of whole number and fractions development. *Cognitive Psychology*, 62, 273-296 (2011).

[6] Rau, M.A., Aleven, V. Rummel, N.: Intelligent tutoring systems with multiple representations and self-explanation prompts support learning of fractions. In: Dimitrova, V. et al. (eds) *Proceedings of the 2009 conference on Artificial Intelligence in Education*. IOS Press, Amsterdam, The Netherlands. pp. 441-448 (2009).

[7] Bodemer, D., Ploetzner, R., Feuerlein, I., Spada, H.: The Active Integration of Information during Learning with Dynamic and Interactive Visualisations. Learning and Instruction, 14, 325-341 (2004).

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

[9] de Jong, T., Ainsworth, S. E., Dobson, M., Van der Meij, J., Levonen, J., Reimann, P., Simerly, C. R., Van Someren, M. W., Spada, H. and Swaak, J. *Acquiring knowledge in science and mathematics: The use of multiple representations in technology-based learning environments*. Oxford (1998).

[10] Bodemer, D. and Faust, U. External and mental referencing of multiple representations. *Computers in Human Behavior*, 22, 27-42 (2006).

[11] Schwonke, R., Berthold, K. and Renkl, A. How multiple external representations are used and how they can be made more useful. *Applied Cognitive Psychology*, 23, 1227-1243 (2009).

[12] Rau, M.A., Rummel, N., Aleven, V., Pacilio, L., Tunc-Pekkan, Z.: How to schedule multiple graphical representations? A classroom experiment with an intelligent tutoring system for fractions. In: Aalst, J.v., et al. (eds) The future of learning: *Proceedings of the 10th ICLS*, ISLS, Sydney, Australia. 64-71 (2012).

[13] Kellman, P. J., Massey, C. M. and Son, J. Y. Perceptual Learning Modules in Mathematics: Enhancing Students' Pattern Recognition, Structure Extraction, and Fluency. *Topics in Cognitive Science*, 22009), 285-305.

[14] Rau, M. A., Aleven, V., Rummel, N. and Rohrbach, S. *Sense Making Alone Doesn't Do It: Fluency Matters Too! ITS Support for Robust Learning with Multiple Representations*. Springer Berlin / Heidelberg (2012).

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

[16] Aleven, V., McLaren, B. M., Sewall, J. and Koedinger, K. R. Example-tracing tutors: A new paradigm for intelligent tutoring systems. *International Journal of Artificial Intelligence in Education*, 19, 105-154 (2008).

[17] Baker, R. S. J. d., Corbett, A. T. and Koedinger, K. R. The difficulty factors approach to the design of lessons in intelligent tutor curricula. *International Journal of Artificial Intelligence in Education*, 17, 341-369 (2007).

[18] Rau, M. A., Aleven, V. and Rummel, N. Blocked versus Interleaved Practice with Multiple Representations in an Intelligent Tutoring System for Fractions, In: Yacef, K. et al. (eds), *Proceedings of the 5th Intl Conference of ITS*, Springer Berlin / Heidelberg (2010).

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

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

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

[22] Robins, J., Scheines, R., Spirtes, P. and Wasserman, L.: Uniform Consistency in Causal Inference. *Biometrika*, 90, 491 – 515 (2003).

[23] Koedinger, K. R. et al.: A data repository for the EDM community: The PSLC Data-Shop, In: C. Romero (ed.), *Handbook of educational data mining* (pp. 10-12). Boca Raton, FL: CRC Press (2010).