

## **Abstract Title Page**

### **Title:**

A Method for the Microanalysis of Pre-algebra Transfer

### **Author(s):**

Philip I Pavlik Jr.

Human Computer Interaction Institute, Carnegie Mellon University

Contact email: [ppavlik@andrew.cmu.edu](mailto:ppavlik@andrew.cmu.edu)

Michael Yudelson

Human Computer Interaction Institute, Carnegie Mellon University

[yudelson@cs.cmu.edu](mailto:yudelson@cs.cmu.edu)

and

Kenneth R. Koedinger

Human Computer Interaction Institute, Carnegie Mellon University

[koedinger@cmu.edu](mailto:koedinger@cmu.edu)

**Background / Context:** As technologies for mathematics education consume larger amounts of student classroom and homework time, methods to analyze the data stream coming from this software become more and more important to maximizing the benefits of educational technologies for students. To address this growing need, the new field of educational data mining has been developing methods to detect and summarize the meaning of educational data to maximize its value to the educational research community (Romero & Ventura, 2007). While educational data mining has many methods, this paper focuses on model-based discovery, a technique that uses mathematical models to create the summary understandings that can then feed back into improvements in educational technology, and hopefully education more generally. Model-based discovery is a new area of educational data mining, and publications showing the importance of these methods are on the rise (Baker & Yacef, 2009).

**Purpose / Objective / Research Question / Focus of Study:** The objective of this research was to better understand the transfer of learning between different variations of pre-algebra problems. While we could have addressed a specific variation that might address transfer, we were interested in developing a general model of transfer, so we gathered data from multiple problem types and their variants over the course of learning (see Setting, Intervention and Design sections). We gathered our data from the classroom but used randomization of item selection and sequence for each student because we were concerned about existing data containing various sources of bias (Shadish & Cook, 2009). Our method, which has been called “*in vivo* experimentation”, blends attention to experimental method with attention to the real life issues of classroom learning (Koedinger, Alevan, Roll, & Baker, 2009; Koedinger & Corbett, 2010), such as motivation issues, attendance and classroom distractions compared to the lab.

This approach is similar to the microgenetic approach. Microgenetic methodology involves using multiple measurements of the data to understand small changes of a person’s behavior (Siegler & Crowley, 1991). To accomplish this, microgenetic experiments to understand learning behaviors have been configured with multiple pre- or post-tests so as to gather the data necessary in a controlled fashion (Siegler & Stern, 1998). This approach has been a rich source of results (e.g. Rittle-Johnson, 2006), and microgenetic methods are often advocated by researchers in the developmental psychology community (Miller & Coyle, 1999). Our use of an educational technology greatly simplified the collection of student action level data for these sorts of microgenetic analyses.

**Setting:** Data on transfer was collected from a Miami based charter school both from classroom work on a computerized educational tutoring software program and from homework on the same system. These natural settings varied widely between individuals, but because the study used full random assignment of students to condition and items to student, the data can be used for post facto analysis of causal effects. While our 10 sets of intervention items were placed as part of the Bridge to Algebra product from Carnegie Learning Inc., we are not examining the Carnegie Learning system, but rather merely using it as a piggyback vehicle to deliver our intervention. However, each of our intervention units did fit in the curriculum sequence in the Carnegie Learning system, so our interventions were appropriate for each student’s current progress in the Carnegie Learning system.

**Population / Participants / Subjects:** Approximately 250 6<sup>th</sup> and 7<sup>th</sup> graders from a charter middle school in the southeast US, participating classes included all level at the school that used the Bridge to Algebra product from Carnegie Learning Inc.

**Intervention / Program / Practice:** Each of the 10 sets of intervention items was configured with some related math content items in a simple format (see example items in Table 1). These lessons were composed of 16 single step problems selected randomly from a set of 24 possible items in each lesson. While our initial intent was to use the Bridge to Algebra software as a post-test to examine the effects of these items, we did not see effects on the Bridge to Algebra content. While we would have liked to see this long-term farther transfer of learning, we have been able to mine the data from within 6 of the lessons to find it reveals important results for a more general understanding of transfer that support specific educational recommendations.

**Significance / Novelty of study:** The design of the study is novel because it provides data to analyze learning at the level individual problem transitions, but does so in an experimentally controlled fashion. This novel data collection feeds into a model of practice that is novel in the way it separately distinguishes categorically different practice events, e.g. successes with story problems, and determines their effect on subsequent categories of problems (Pavlik Jr., Yudelson, & Koedinger, 2011, accepted). Our method of using categories of events as predictors is quite intuitive, but provides a distinct advantage for capturing asymmetrical transfer (e.g. Bassok & Holyoak, 1989) in a trial by trial learning curve model compared to popular methods that assume abstract skills such as rule-space methods (Barnes, 2005; Tatsuoka, 1983). Because rule-space methods assume a shared abstract latent skill they do not model asymmetric transfer well, since gain in the latent skill cannot be asymmetric.

**Statistical, Measurement, or Econometric Model:** Currently we call this method Contextual Factors Analysis (CFA) to capture the notion that the interaction of the contexts of learning and the context of performance determine the performance that is observed for any student (Pavlik Jr., et al., 2011, accepted). While CFA theory therefore refers to this notion of the differing importance of different prior learning contexts, the underlying formal method applies logistic regression to compute the analysis. Simply put, prior events in the students learning are each categorized and counted to predict the next practice result given the category of the problem being responded to. This procedure means that we have a single coefficient capturing, for example, the effect on item-type B of the number of prior practices with item-type A that were successful. Because there are 4 categories of prior practice – success on A, success on B, failure on A and failure on B – and because there are 2 categories of future practice – A or B only since success is not known for future events – we find that the model has 8 parameters (4 conditions of prior practice which affect 2 conditions of future practice differently). This model, see Figure 1, allows us to compose Table 1 which shows the strengths of these effects as revealed by the model. (please insert figure 1 here) All of the models we fit used fixed intercepts to capture average prior knowledge in the contrasts and overall, and modeled the individual users and individual items as sources of random effects (i.e. these are mixed-effect models). In Table 1, the notation represents transfer or learning with ‘S’ or ‘F’ for success based effects and failure based effects, and indicates the direction of the learning or transfer with the  $A \rightarrow B$  notation, which, for example indicates a transfer of learning from A to B. For example,  $S_{A \rightarrow B}$  measures the count of prior successes with A as they affect B. (please insert table 1 here)

**Usefulness / Applicability of Method:** The method is primarily useful for the implications of these models for designing, comparing and sequencing problems or other learning objects. For this usefulness of inductive implications to be manifest, it is crucial that data input to the model fitting is unbiased in its sequence, since the statistical method assumes this unbiased problem ordering. Given this unbiased problem ordering the method is applicable in situations where there are 2 or more different kinds of problems given in a sequence. In this paper we only describe differences between pairs of item types, but given enough data, the method is applicable to multiple (>2) item types in the same random sequence. Further, the method assumes that each problem provides some sort of correctness feedback (right or wrong for each response) and example-based instruction (provision of a correct response when the student is wrong). The model uses correctness to categorize each prior practice. While the requirement for randomly ordered data may be cumbersome in the classroom, we have found that short sets of related problems (which while different, are clearly in the same concept area to the teacher) have worked well to see the fine grained effect of individual item-types as students learn related ideas. These problems are very much like worksheet or test items students already work on and we received no reports of their being disruptive as an integrated activity.

**Research Design:** The experimental side of this project is best described as an experimental design in a naturalistic context, but this paper focuses on post-hoc educational data mining methodology to analyze the implications of the student results. The design used 10 sets of 24 individual pre-algebra single step questions on a variety of content (see Table 1 for example items). The 10 interventions we gave were split into “item-types” according to systematic analysis of their features (these difference were a design feature of the sets of 24 items). For example, in our first problem set, one of the 2 comparisons was between  $\frac{1}{2}$  of the problems, which were story problems with people’s names and units of measurement (e.g., “Sally visits her grandfather every 4 days and Molly visits him every 5 days. If they are visiting him together today, in how many days will they visit together again?”) and the other half of the items, which were written as word problems (e.g., “What is the least common multiple 4 and 5?”). For each of the 10 interventions students were each quizzed on 16 randomly selected items from these sets of 24 (see research design).

**Data Collection and Analysis:** Data collection was performed by the software. Analysis was performed using the mixed model logistic regression as implemented in the R software package with the `lmer` function. Mixed model logistic regression provides methods to find “random-effects” models that capture both the effect of fixed factors (e.g., the effect of prior categories of problems or the effect of success or failures) and the effect of random factors (e.g., the prior student aptitude) that are merely sampled from a population (Pavlik Jr., et al., 2011, accepted). It is important to note that the model we have settled on was validated with extensive cross-validation, a procedure that holds out a portion of the data to test predictive causal accuracy in a post hoc way. This method provides us assurance that our models were not just finding patterns in the data; they are finding patterns that generalize to unseen data.

**Findings / Results:** Table 1 reports on 6 of the 10 sets where we collected data without technical problems (e.g. set 5 could not be analyzed because it was multiple choice and proved to be too noisy for a reliable seeming analysis, sets 8 and 9 had typos and set 10 had no clear factor

contrast). We see in Table 1 that with a few exceptions in the case of learning from failures (discussed below), mostly learning is a stronger effect than transfer. For example, inspecting the first row, the four learning slope estimates (.197, .416, .179, and .079) are mostly larger than the four transfer slope estimates (.009, .087, .009, and .088).

Results in Table 1 are useful to analyze individually to see the strength of the method in capturing multiple patterns of transfer relationships. In set 1, our first contrast looked at items where the least common multiple (LCM) was simply the product and items where the LCM was less than the product. In this case the lack of transfer from LCM<product to LCM=product items ( $S_{A \rightarrow B} = .009$ ,  $F_{A \rightarrow B} = .009$ ) was interesting because it appears to contradict the conventional wisdom of starting with the easier items (LCM=product), and strengthening subskills, before moving to harder items (LCM<product). The model shows this assumption is incorrect and indicates that only LCM<product items appeared to cause transfer ( $S_{B \rightarrow A} = .087$ ,  $p < .10$ ;  $F_{B \rightarrow A} = .088$ ,  $p < .05$ ). An error analysis supported the idea that LCM=product items may prevent transfer because LCM=product was a common error for LCM<product items (e.g., entering 24 for the LCM of 6 and 4). This error of inappropriately providing a product represented 5.7% of the total commission errors for LCM<product problems. Set 6 had a similar effect where unlike denominator addition (harder) transferred while like denominator addition (easier) did not.

The second contrast in Set 1 revealed that the more abstract word problems had superior transfer ( $S_{A \rightarrow B} = .072$ ,  $p < .10$ , and  $F_{A \rightarrow B} = 0.187$ ,  $p < .001$  for success and failure). This result is supported by recent research on transfer advantages of simple symbols (in our case the more abstract word problems) compared to more concrete representations (in our case the concrete story problems) (e.g. Sloutsky, Kaminski, & Heckler, 2005). Effects in Sets 2 and 7 are perhaps similar since they both reveal some asymmetry that might be explained by appealing to the transferability of the representations.

These examples provide a useful way to consider research on mixed vs. blocked problems (e.g. Rohrer, 2009) by providing a model of micro-level transfer during a mixed practice block. Another example of the model's explanatory breadth is shown by the negative transfer seen in sets 3 and 6, which seem best explained as interference from a mental set that blocks proper attention to critical features that need to be re-encoded for each problem. For example, the first contrast in Set 3 shows negative success transfer ( $S_{A \rightarrow B} = -0.157$ ,  $p < .001$ ; and  $S_{B \rightarrow A} = -0.261$ ,  $p < .001$ ) perhaps because successful practice in subtraction (take away from 1) and addition (add from 0) item-types interfere with each other. Set 4 Cost vs. Wealth suggests a similar confusing effect of following cost problems with wealth problems.

**Conclusions:** Acknowledging the limitations for randomized data and individualized events with performance measures discussed above, we are confident in recommending these procedures more broadly to understand the problems of transfer between different mathematical exercises. The diagnostic affordances of this method make it an important addition to our tools for understanding transfer in an objective fashion. Our data shows many asymmetries, and while some prior research has focused on similar issues of asymmetric transfer (Bassok & Holyoak, 1989), we not aware of any methods to generally approach this problem with the goal of improving our ability to summarize and understand the huge quantities of data being accumulated in educational settings. Future development of this model-based method for item analysis, sequencing, and design is focused on improving the integration of a student model of learning within the transfer model so as to address individual differences in transfer more effectively with the formal model and theory we have created.

## Appendices

### Appendix A. References

- Baker, R. S. J. d., & Yacef, K. (2009). The State of Educational Data Mining in 2009: A Review and Future Visions. *Journal of Educational Data Mining*, 1(1), 3-17.
- Barnes, T. (2005). *The Q-matrix Method: Mining Student Response Data for Knowledge*. Paper presented at the American Association for Artificial Intelligence 2005 Educational Data Mining Workshop.
- Bassok, M., & Holyoak, K. J. (1989). Interdomain transfer between isomorphic topics in algebra and physics. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(1), 153-166. doi: 10.1037/0278-7393.15.1.153
- Koedinger, K. R., Alevan, V., Roll, I., & Baker, R. S. J. d. (2009). In vivo experiments on whether supporting metacognition in intelligent tutoring systems yields robust learning. In D. J. Hacker, J. Dunlosky & A. C. Graesser (Eds.), *Handbook of Metacognition in Education*. New York: Routledge.
- Koedinger, K. R., & Corbett, A. T. (2010). The Knowledge-Learning-Instruction (KLI) Framework: Toward Bridging the Science-Practice Chasm to Enhance Robust Student Learning (pp. 42). Pittsburgh: Carnegie Mellon University.
- Miller, P. H., & Coyle, T. R. (1999). Developmental change: Lessons from microgenesis. In E. K. Scholnick, K. Nelson, S. A. Gelman & P. H. Miller (Eds.), *Conceptual development: Piaget's legacy*. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- Pavlik Jr., P. I., Yudelso, M., & Koedinger, K. (2011, accepted). Using Contextual Factors Analysis to Explain Transfer of Least Common Multiple Skills *15th International Conference on Artificial Intelligence in Education*.
- Rittle-Johnson, B. (2006). Promoting Transfer: Effects of Self-Explanation and Direct Instruction. *Child Development*, 77(1), 1-15.
- Rohrer, D. (2009). The effects of spacing and mixing practice problems. *Journal for Research in Mathematics Education*, 40(1), 4-17.
- Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications*, 33(1), 135-146. doi: DOI: 10.1016/j.eswa.2006.04.005
- Shadish, W. R., & Cook, T. D. (2009). The Renaissance of Field Experimentation in Evaluating Interventions. *Annual Review of Psychology*, 60(1), 607-629. doi: doi:10.1146/annurev.psych.60.110707.163544
- Siegler, R. S., & Crowley, K. (1991). The microgenetic method: A direct means for studying cognitive development. *American Psychologist*, 46(6), 606-620.
- Siegler, R. S., & Stern, E. (1998). Conscious and unconscious strategy discoveries: A microgenetic analysis. *Journal of Experimental Psychology: General*, 127(4), 377-397.
- Sloutsky, V. M., Kaminski, J. A., & Heckler, A. F. (2005). The advantage of simple symbols for learning and transfer. *Psychonomic Bulletin & Review*, 12(3), 508-513.
- Tatsuoka, K. K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement*, 20(4), 345-354.

## Appendix B. Tables and Figures

Figure 1. Logistic regression equation model.

$$y_i = \beta_0 + \beta_1 \times \text{contrast} + \\ \beta_2 \times S_{A \rightarrow A} + \beta_3 \times S_{B \rightarrow B} + \beta_4 \times F_{A \rightarrow A} + \beta_5 \times F_{B \rightarrow B} + \\ \beta_6 \times S_{A \rightarrow B} + \beta_7 \times S_{B \rightarrow A} + \beta_8 \times F_{A \rightarrow B} + \beta_9 \times F_{B \rightarrow A} + \\ \alpha_{0i} + \alpha_{1j}$$

where:

$\beta_0$  is an overall fixed intercept for the set

$\beta_1$  is an intercept for each level of the set contrast

$\beta_{2...9}$  capture the effects of counts of prior practice categories relevant to the item predicted

$\alpha_0$  is the random effect intercept for each student  $i$

$\alpha_1$  is the random effect intercept for each item in the set  $j$

Table 1. Results of logistic regression coefficients for learning and transfer across 6 datasets with 9 comparisons.

Set	Contrast	Item-type A Example	Item-type B Example	Success Learning		Failure Learning		Success Transfer		Failure Transfer	
				$S_{A \rightarrow A}$	$S_{B \rightarrow B}$	$F_{A \rightarrow A}$	$F_{B \rightarrow B}$	$S_{A \rightarrow B}$	$S_{B \rightarrow A}$	$F_{A \rightarrow B}$	$F_{B \rightarrow A}$
1	LCM=product strategies vs. LCM<product	What is the least common multiple 3 and 5?	What is the least common multiple 8 and 12?	0.197 ***	0.416 ***	0.179 *	0.079 *	0.009	0.087 +	0.009	0.088 *
1	Word problem vs. Story Problems	What is the least common multiple 3 and 5?	Sally visits her grandfather every 3 days and Molly visits him every 5 days. If they are visiting him together today, in how many days will they visit together again?	0.380 ***	0.339 ***	0.176 **	0.016	0.072 +	-0.028	0.187 ***	-0.010
2	Groups vs. Factors	How many groups of 7 items can you make in 14?	If 8 is a factor of 32, what is the matching factor in the factor pair with 8?	0.253 **	0.436 ***	0.249	0.195 +	0.095 +	-0.031	0.340 *	0.275 *
3	Add to 0 vs. Take away from 1	On a number line, the interval between "0" and "1" is partitioned into 5 equal parts. What is the fraction number that corresponds to the point that is 2 parts away from "0"?	If we have a string and cut it into 5 equal pieces, what is the fraction we have if we take away 2 pieces? (Remember, 1 whole is the length of the original string).	0.279 ***	0.470 ***	0.280 ***	0.055	-0.157 ***	-0.261 ***	0.086	-0.051
3	String vs. Number line	If we have a string and cut it into 5 equal pieces, what is the fraction we have if we take away 2 pieces? (Remember, 1 whole is the length of the original string).	On a number line, the interval between "0" and "1" is partitioned into 7 equal parts. What is the fraction number that corresponds to the point that is 2 parts away from "1"?	0.183 ***	0.234 ***	0.044	0.169 ***	-0.051	-0.075	0.034	0.012
4	Cost vs. Wealth	Mike has \$20. He wants to buy a t-shirt and t-shirts cost 1/2 of his money. How many dollars is a t-shirt?	Jane wants to buy a DVD which is \$5. If the DVD's price is 1/3 of her money, how many dollars does she have?	0.424 ***	0.307 ***	0.253 ***	0.428 ***	-0.049	-0.158 ***	0.079	-0.019
6	Same denominators vs. Different denominators	What is the denominator in the solution to $4/5 + 2/5$ ? (do not reduce or convert to a mixed number before answering)	What is the denominator in the solution to $1/5 + 2/3$ ? (do not reduce or convert to a mixed number before answering).	0.410 ***	0.417 ***	0.084	0.048	-0.018	0.143 *	0.009	-0.142 **
6	Answer is denominator vs. Answer is numerator	What is the denominator in the solution to $4/5 + 2/5$ ? (do not reduce or convert to a mixed number before answering)	What is the numerator in the solution to $4/5 + 2/5$ ? (do not reduce or convert to a mixed number before answering)	0.480 ***	0.491 ***	0.150 **	0.051	-0.069	0.034	-0.102 +	-0.096 +
7	Gas volume vs. Number line	A car's tank has 11/5 gallons of gas how many whole gallons of gas does the tank have?	On a number line if 11/5 is located, what is the biggest whole number that comes on or before that point?	0.382 ***	0.288 ***	0.047	-0.010	0.175 **	0.140 *	0.213 **	0.056

\*\*\*p < .001. \*\*p < .01. p < .05. +p < .10.