What Were They Thinking? Diagnostic Coding of Conceptual Errors in a Mathematics Learning Software Data Archive

Christine Massey, Jennifer Kregor, Laura Cosgrove & Himchan Lee
University of Pennsylvania

Abstract

As new adaptive learning technologies become ubiquitous in education, they bring opportunities both to facilitate conceptual development in more focused ways and to gather data that may yield new insights into students' learning processes. This study analyzed data archives from an adaptive perceptual learning intervention designed to help students master key concepts related to linear measurement and fractions. Using algorithmic data coding on a database of 78,034 errors from a sample of 716 6th graders, both conceptual errors and other errors were captured and analyzed for change over time. Results indicate that improved encoding of relevant structure led to decreases in common conceptual errors and gains in accuracy.

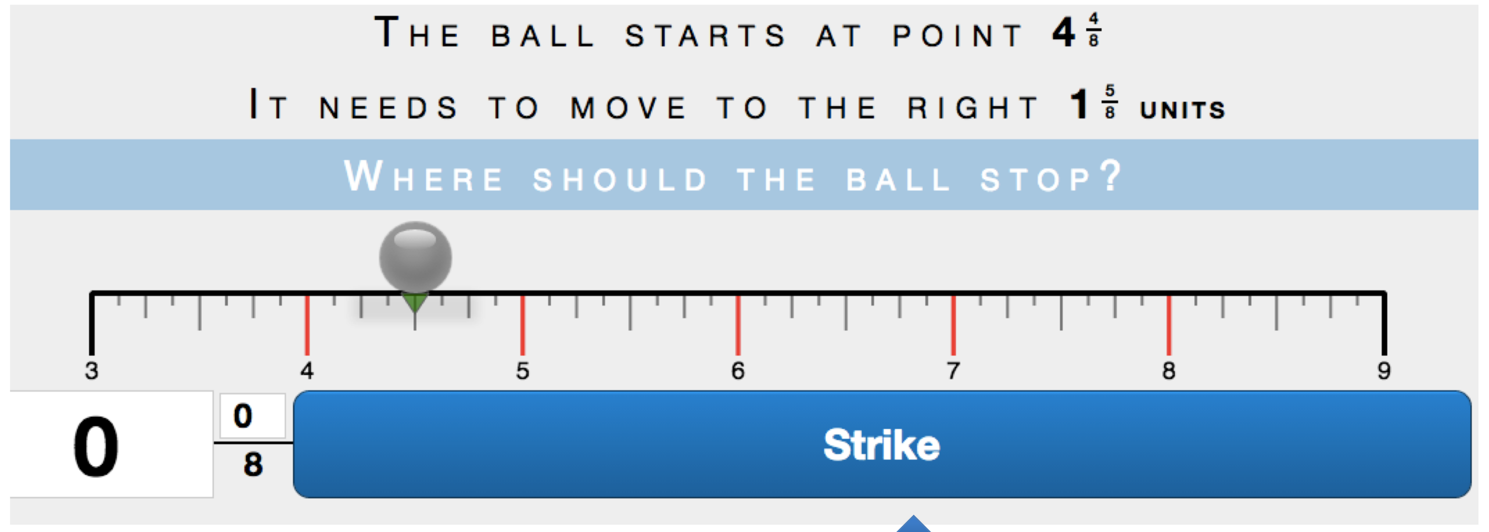
Linear Measurement Perceptual Learning Module (PLM)

The intervention consisted of a web-based perceptual learning module (PLM) (Kellman & Massey, 2013; Massey, Kellman, Roth & Burke, 2010) designed to improve students' ability to extract the structure of units of linear measurement and to accurately and fluently process points and intervals on rulers (or number lines) for both integer and fraction values. Students used onscreen tools to actively engage in a variety of problems with custom animated feedback until they reached mastery criteria for all problem categories.

The PLM is designed to counteract common conceptual problems, such as counting discrete numbered hash marks on rulers rather than measuring continuous intervals; failing to distinguish between a position and a distance on a number line; not understanding fractional subdivisions of intervals; and confusion over multiple labels for the same point (e.g., 2/4 and 4/8).

Separate studies have demonstrated significant learning gains for students using the Linear Measurement PLM compared to control groups in previously reported studies (Massey et al., 2010) and in 2 cohorts of a large RCT (Scull, 2015). Outcomes include significant long-lasting treatment effects on one-year delayed posttests (Scull et al., in preparation).

Example Item & Error



The student's task is to enter the Endpoint, given a Start Point and Distance traveled. After a response is entered and "Strike" is pressed, the software executes the action and provides animated feedback. This problem crosses an integer boundary, requiring regrouping of fraction and integer components.

Regrouping Error

$4\frac{4}{8} + 1\frac{5}{8}$

$4 + 1 \text{ and } \frac{4}{8} ? \frac{5}{8}$ — Student adds whole numbers correctly but doesn't know how to combine fractions that cross an integer boundary

$= 5\frac{1}{8} \text{ or } 5\frac{4}{8} \text{ or } 5\frac{5}{8}$ — Student subtracts fraction components instead of adding or simply repeats one fraction and ignores the other

$\text{Correct answer} = 6\frac{1}{8}$

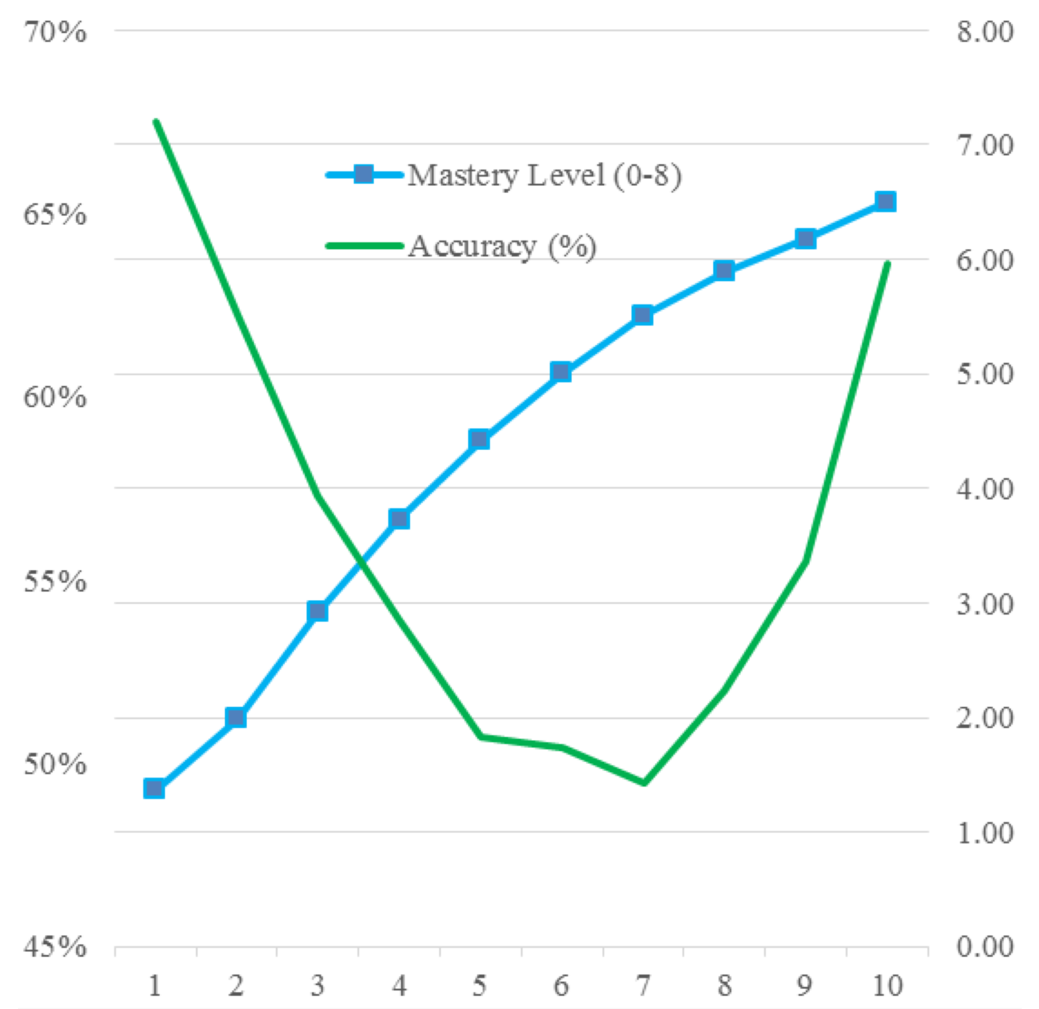
Method

The dataset consisted of 157,147 problems completed by 716 6th graders from 30 classrooms in urban schools serving predominantly low-income minority students. Students used the PLM in class as part of their normal math curriculum. Each student had to complete at least 20 problems but did not have to complete the entire PLM to be included. 52.8% of students mastered all 8 levels of the PLM and 60.5% mastered at least 6 levels.

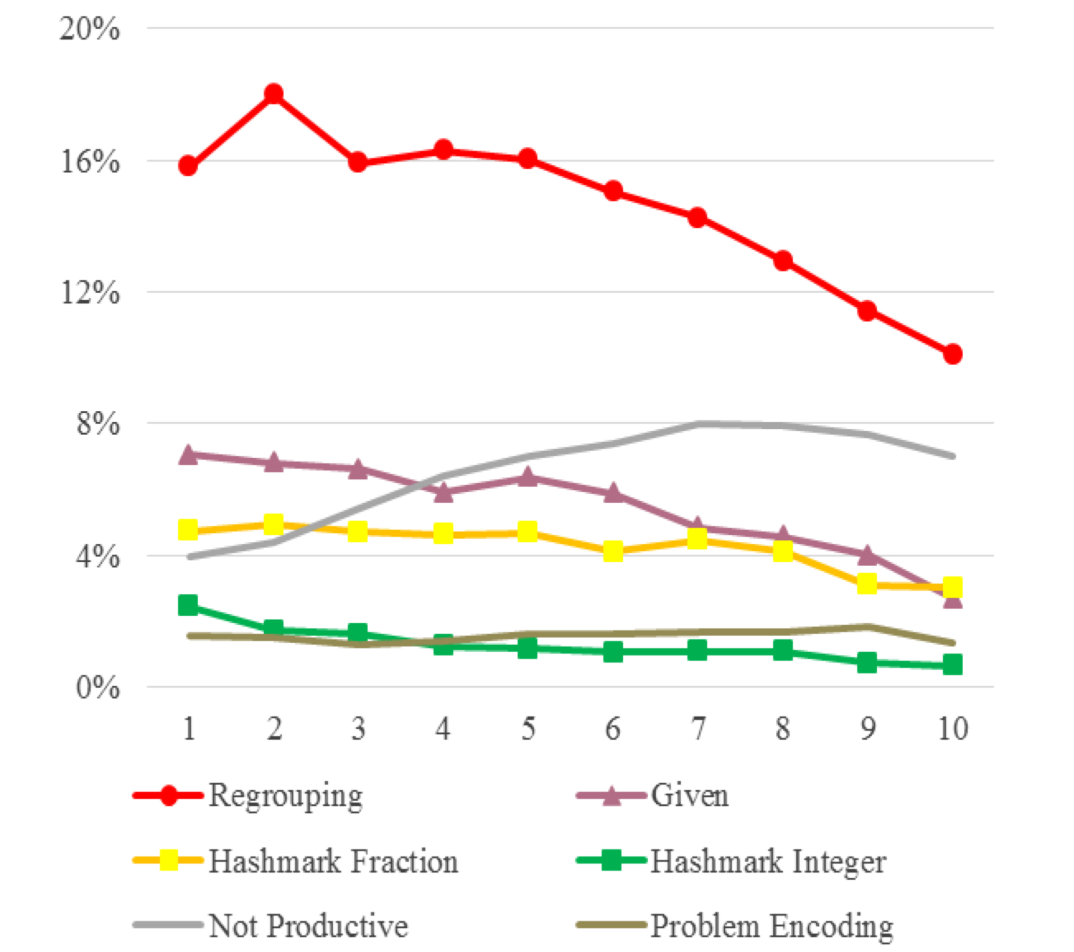
Each unique problem can be deconstructed into a set of parameters that were used to create algorithms to operationally define a set of targeted errors, with a particular focus on conceptual errors involving miscounting of hash marks and regrouping errors involving fractions that cross an integer boundary. Out of 78,034 total errors, 38,337 (49.1%) were captured by a single well-specified error category. Errors that were captured by multiple codes or by no codes were not included in the analysis.

Error	Total Errors Coded	% of Errors All Students	N Students with error 5+ times
Unproductive Responses	11,080	14.2%	479
Given Information	9,588	12.3%	463
Regrouping	8,363	10.7%	464
Hashmark Fraction	5,073	6.5%	382
Problem Encoding	2,533	3.2%	193
Hashmark Integer	1,700	2.2%	111
Total Errors/Total N	78,034		716

Results



To examine changes in the rates of various error types over time, each student's time-ordered sequence of trials was divided into 10 phases. As the figure above shows, learners typically made steady progress through the PLM, accumulating up to 8 mastery levels. It also shows a distinctive U-shaped curve for average accuracy, which started around 68% as the PLM begins with the simplest integer problems, which are adaptively retired as they are mastered. Accuracy dropped to a low near 50% during the middle of training before climbing back up, as performance improved on more difficult problem types.



The figure above compares the proportion of errors of each type made by each student at each phase of learning, averaged across all students. Regrouping errors showed the highest rate (relative to eligible problems) in all phases, and they decreased steadily in phases 5-10, mirroring increases in average accuracy and mastery levels across phases 7-10. Unproductive responses initially increased and then leveled off. Problem Encoding errors were relatively uncommon and remained steady across phases. Repeated measure ANOVAs run on each error type across phases indicated that Regrouping, Hash Mark, and Given Information errors all decreased significantly ($p<.001$). Unproductive Responses increased significantly ($p<.001$) and Problem Encoding errors did not vary significantly.

Conclusions

Prior studies have demonstrated that the *Linear Measurement* PLM leads to significant and long-lasting learning gains on external tests. The present analysis of changes in error patterns illuminates the theory of action behind these gains by indicating that the PLM was successful in mitigating several specific conceptual errors, such as helping students to perceive units of linear measurement as continuous intervals rather than discrete hash marks and to comprehend how fractions of units are represented on the number line and how they relate to integer units.

Future analyses will examine whether additional types of errors can be captured with well-defined algorithms and will also investigate learning patterns across subgroups of learners. Findings from error analyses of large data archives can also be a powerful input to the design of next generation adaptive learning software.

The research reported here was supported in part by the Institute of Education Sciences, U.S. Department of Education, through Grants #R305B090015 and #R305A120288 to the University of Pennsylvania and UCLA. The opinions expressed are those of the authors and do not represent views of the U.S. Department of Education.