Learning Curve Analysis in a Large-Scale, Drill-and-Practice Serious Math Game: Where is Learning Support Needed?

Zhongxiu Peddycord-Liu[1], Rachel Harred[1], Sarah Karamarkovich[1], Tiffany Barnes[1], Collin Lynch[1], and Teomara Rutherford[1]

North Carolina State University, Raleigh, NC, USA,
`zliu24, rlharred, skessle, tmbarnes, cflynch, taruther@ncsu.edu`

Abstract. This paper applies data-driven methods to understand learning and derives game design insights in a large-scale, drill-and-practice game: Spatial Temporal (ST) Math. In order for serious games to thrive we must develop efficient, scalable methods to evaluate games against their educational goals. Learning models have matured in recent years and have been applied across e-learning platforms but they have not been used widely in serious games. We applied empirical learning curve analyses to ST Math under different assumptions of how knowledge components are defined in the game and map to game contents. We derived actionable game design feedback and educational insights regarding fraction learning. Our results revealed cases where students failed to transfer knowledge between math skills, content, and problem representations. This work stresses the importance of designing games that support students' comprehension of math concepts, rather than the learning of content- and situation-specific skills to pass games.

Keywords: Learning Analytics; Serious Games; Educational Data Mining; Fraction Learning

1 Introduction

Serious games are in high demand with a fast-growing market [1]. As such, they are poised to significantly impact future generations of learners. As more games emerge, we face unprecedented challenges in evaluating the games against their educational goals. Traditional methods fall short. Games are growing in size and complexity, making it increasingly costly to evaluate games through controlled-trial experiments. Games evolve quickly during production and maintenance cycles, making it necessary to pinpoint design flaws and to derive actionable feedback efficiently. We need new methods to penetrate the black box between pre- and post-tests and to inform game design in an efficient, scalable fashion.

Serious game analytics [14] is an emerging field that uses data to connect gameplay with learning. In contrast to traditional game analytics [7] that primarily focuses on player enjoyment, serious game analytics grounds game design

in students' learning and performance on targeted skills. Serious game analytics is extremely valuable for large-scale, curriculum-integrated games. This approach not only helps pinpoint game design issues efficiently, but can also be used to investigate the learning process and derive educational insights across domains [13, 14, 19]. Serious game analytics shares similar and potentially transferable methods with Learning Analytics [27] and Educational Data Mining [23]. However, serious game analytics is concerned with more specific domains, where learning is interleaved with game mechanisms as well as other factors. We need more research to verify the applicability of these methods across domains, and to adapt these methods to solve contemporary challenges in serious games.

In this paper, we applied empirical learning curve analyses. We fit and combined learning curves under different assumptions of how knowledge components are defined in the game and mapped to game contents. Our analyses aim to: 1) Understand and model learning in ST Math–a large-scale, drill-and-practice game that introduces and reinforces math skills through various problem-solving scenarios. 2) Derive actionable feedback to help game designers better design game content to support learning. 3) Provide data-driven insights on fraction learning and the knowledge transfer between problem-solving scenarios. 4) Suggest future research that analyzes and models learning in serious games.

1.1   Spatial Temporal Math (ST Math)

ST Math is designed to act as a supplemental program to a school's existing mathematics curriculum [13, 19, 24]. In ST Math, mathematical concepts are taught through spatial puzzles within various game-like arenas. These games are structured at the top level by *objective*, which cover broad learning concepts. Within each objective, individual *games* teach more targeted concepts through presentation of *puzzles*, which are grouped into *levels* for students to play. Students begin by completing a series of training games on the use of the ST Math platform and features. They are then guided to complete the first available objective in their grade-level curriculum, such as "Multiplication Concepts." Within an objective, games represent scenarios for problem-solving using a particular mathematical concept, such as finding the right number of boots for X animals with Y legs. Each game contains between one and ten levels, which follow the same general structure of the game, but with increasing difficulty. Students must unlock the games and levels in their designed order.

As with many games, students are given a set number of 'lives' per level. Every time they answer a puzzle incorrectly they lose one life. Each answer attempt is followed by animated feedback. For example, in a game asking students to select a number of boots for two dogs, a wrong answer of six will show that one dog has two feet without boots. After the feedback, students must attempt the same puzzle again until they answer it correctly or they exhaust all lives. If all lives for a given level are exhausted, students must re-attempt the level. When students re-attempt a level there is a probability that they may encounter a previously attempted puzzle, as puzzles are either randomly generated following a template, or randomly selected from a pre-designed puzzle pool. Once students

pass a level, they can play the next level or any previously-passed level. A general description and additional figures of ST Math can be found in [13, 19, 24].

2   Literature Review

Learning Curves are derived from the cognitive theory of Newell and Rose-bloom [17]. This theory assumes that with more practice, a students' speed and accuracy at answering a question increase logarithmically. In other words, a good learning curve shows that students' accuracy increases, but the increase gets smaller over time. After enough practice, the increase will be negligible, and students can be considered to have reached their best performance.

Several learning models have been applied to fit learning curves. The Additive Factors Model (AFM) [4] is a logistic regression that assumes the probability of correctly answering a question depends on individual students' parameters, the skill difficulties, and the number of previous practices. For questions containing multiple skills, the difficulty and practices of these skills are summed together. This assumes that a student can correctly answer a question without knowing all of the skills involved, provided the summation of known skills passes a certain threshold. The Conjunctive Factor Model (CFM) [5] is similar to AFM, but assumes that the difficulties and practices of skills are multiplied together for questions with multiple skills. This means that a student can never answer a question correctly unless they know all the required skills. Another method is Performance Factor Analysis (PFA) [18], which is similar to AFM with an additional assumption that success and failure have different impacts on learning. Because we don't currently have evidence that this additional assumption holds in ST Math, we focused on applying AFM and CFM in this paper.

Despite the wide application of empirical Learning Curves on intelligent tutors and other e-learning platforms, there has been little application in serious games. As one exception, Harpstead and Aleven [8] applied the AFM model in a physics game. Through examining learning curves, they identified an unforeseen shortcut strategy with which students could pass the game without sufficiently mastering its underlying math concepts. Similarly, Baker et al. [3] fit learning curves in a action-based math game to model gains in speed and accuracy over time. They found that modeling accuracy helped to identify skills that needed extra support and scaffolding. However, modeling speed was difficult, because it was hard to separate gains in math fluency from familiarity with the game-play. Lomas et al. [15] applied learning curves to a game locating numbers on a number line. They found when students were allowed to pass the game with less accurate estimations, the learning rate was lower. In these cases, estimating learning curves revealed areas for modification and improvement within serious games. However, in all previous work, the games studied represented a single problem-solving scenario. In contrast, games such as ST Math practice math skills through various problem-solving scenarios. Applying learning curves to ST Math can yield not only game design insights, but help understand how students transfer knowledge across different problem solving scenarios.

3   Data

MIND Research Institute, the developer of ST Math, collected and provided
the researchers with sample data from 3rd grade students who played ST Math
from August 2016 to February 2017. We focused on the "Comparing Fractions"
objective. Performance on fractions has been found to predict future mathe-
matical achievement [2, 26, 28]. Thus, investigating this objective will allow us
to contribute suggestions for game design of instruction around a crucial math
concept. This objective contains 26 levels across seven games; 1,007 students
completed the first game, and 860 students completed the last game. ST Math
recorded students' IDs, answers, and response times for each puzzle attempt.
The data also included the correct answer for each puzzle, and the level, game,
and objective to which it belonged. We filtered out students' replay of previously
passed levels [13] to focus solely on their attempts to pass an unlearned case.
The final data contains 146,498 unique puzzle attempts.

4   Analyses

Our goal was to understand and model learning in ST Math, and to suggest
better game design that would foster greater learning. ST Math is structured at
the top level by objectives, where games representing different problem-solving
scenarios and levels are puzzle sets of increasing difficulty. Puzzles are either ran-
domly generated following a template, or randomly selected from a pre-designed
puzzle pool. We first fit learning curves to the puzzles to identify the learning
patterns at each level. We then combined the levels hierarchically within each
game. We sought to find similarities between the levels–modeling levels as con-
tinuations within a single learning curve. Lastly, we sought to find associations
between games. We used an expert-designed Q-matrix with knowledge com-
ponents describing the shared math skills and problem representations across
games. We fit learning curves using different assumptions to identify how these
knowledge components interacted and affected students' learning across games.

4.1   Fitting Learning Curves

We used the AFM [4] and CFM [5] models to fit learning curves. We decided
to model the probability that a student answered a puzzle correctly on the first
attempt only. This is because students receive animated feedback after each
answer, with some feedback enabling them to quickly correct a wrong answer
without having to redo the entire problem. Thus, subsequent attempts may not
truly reflect each student's knowledge of the mathematics content. Next, we
decided to fit a learning curve on the first N puzzles in each level, where N
is the number of puzzles that must be answered correctly in order to pass the
level. We chose N because students who passed the level without exhausting all
of their lives (which is the majority in most levels) will not need another attempt.
Attempts following N will only contain data from low-performing students who

had to attempt the level multiple times to pass. Thus, we only consider students' first N puzzles (practice opportunities) in order to fit our learning curve on the same population. Lastly, for student variables in AFM and CFM, we used students' average performance in the prior two objectives: "Fraction Concepts" and "Fractions on a Number Line." These two reflect students' knowledge of fractions prior to attempting this objective.

4.2   Analyzing Puzzles in Levels

In ST Math, each level stresses skills of increasing difficulty under a problem-solving scenario defined by the game. The increased difficulty can be introduced by changes in math content (e.g., using larger numbers), changes in problem representation (e.g., use of math symbols instead of an visual object), or other factors. However, in each case the problem-solving scenario (e.g., finding shoes for animals) remains the same. Thus, we started by assuming one-to-one mappings between levels and knowledge components, and fit learning curves with AFM.

Figure 1 shows the learning curve fit, with content of representative games described in the later text. We applied 10-fold cross-validation and reported the model's accuracy as the percentage of instances correctly predicted by the logistic regression models. The majority of puzzles were answered correctly by over 50% of students, therefore we include the AUROC to describe how well models can differentiate between true positive and false positive.

We looked for four specific patterns to derive game-design feedback. A ***good learning curve*** displays a logarithmic pattern indicating that students increased their accuracy with practice and thus that learning appears to be well-supported. An ***incomplete learning curve*** is similar to a good learning curve except it does not include a flat tail. This indicates that students can still improve with more practice. Next, a ***flat learning curve***, which we defined here as having a smooth curve with $e <= 0.05$, means that students' performance did not improve substantially with practice. It could be that the level is too difficult or that the game content is not well-designed thus students did not learn with practice. It could also be that the level is too easy: students started with near perfect performance. Lastly, a ***non-learning curve*** does not follow a logarithmic or flat pattern. Performance in this type of learning curve increased or decreased suddenly at specific attempts. When this happens, it means that there was a change in the puzzle's template that introduced what the students perceived to be a new knowledge component. For example, a level can have half of the puzzles randomly generated with odd denominators, and the other half randomly generated with even denominators. If students failed to transfer the knowledge when denominators changed, we will see two ***disjointed learning curves***.

The majority of levels (16 of 26) showed the logarithmic pattern of a learning curve, which means that the game design helped students learn (or improve their performance). A few learning curves are incomplete (e.g., Game 7 L3), suggesting that game designers should increase the number of puzzles required to pass the levels (L) as students still have room for improvement.
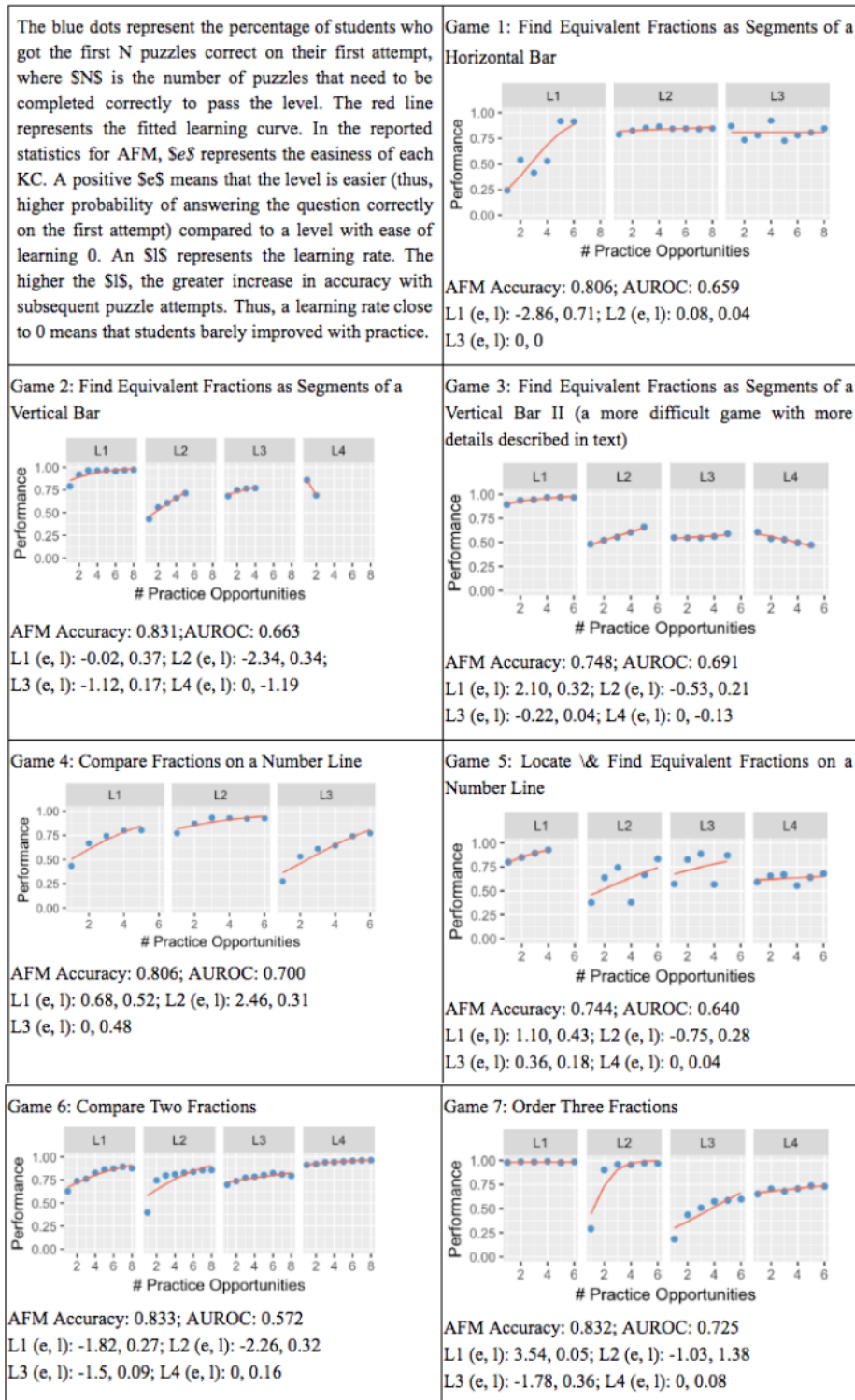
The blue dots represent the percentage of students who got the first N puzzles correct on their first attempt, where N is the number of puzzles that need to be completed correctly to pass the level. The red line represents the fitted learning curve. In the reported statistics for AFM, e represents the easiness of each KC. A positive e means that the level is easier (thus, higher probability of answering the question correctly on the first attempt) compared to a level with ease of learning 0. An l represents the learning rate. The higher the l, the greater increase in accuracy with subsequent puzzle attempts. Thus, a learning rate close to 0 means that students barely improved with practice.

Game 1: Find Equivalent Fractions as Segments of a Horizontal Bar

AFM Accuracy: 0.806; AUROC: 0.659
L1 (e, l): -2.86, 0.71; L2 (e, l): 0.08, 0.04
L3 (e, l): 0, 0

Game 2: Find Equivalent Fractions as Segments of a Vertical Bar

AFM Accuracy: 0.831; AUROC: 0.663
L1 (e, l): -0.02, 0.37; L2 (e, l): -2.34, 0.34;
L3 (e, l): -1.12, 0.17; L4 (e, l): 0, -1.19

Game 3: Find Equivalent Fractions as Segments of a Vertical Bar II (a more difficult game with more details described in text)

AFM Accuracy: 0.748; AUROC: 0.691
L1 (e, l): 2.10, 0.32; L2 (e, l): -0.53, 0.21
L3 (e, l): -0.22, 0.04; L4 (e, l): 0, -0.13

Game 4: Compare Fractions on a Number Line

AFM Accuracy: 0.806; AUROC: 0.700
L1 (e, l): 0.68, 0.52; L2 (e, l): 2.46, 0.31
L3 (e, l): 0, 0.48

Game 5: Locate \& Find Equivalent Fractions on a Number Line

AFM Accuracy: 0.744; AUROC: 0.640
L1 (e, l): 1.10, 0.43; L2 (e, l): -0.75, 0.28
L3 (e, l): 0.36, 0.18; L4 (e, l): 0, 0.04

Game 6: Compare Two Fractions

AFM Accuracy: 0.833; AUROC: 0.572
L1 (e, l): -1.82, 0.27; L2 (e, l): -2.26, 0.32
L3 (e, l): -1.5, 0.09; L4 (e, l): 0, 0.16

Game 7: Order Three Fractions

AFM Accuracy: 0.832; AUROC: 0.725
L1 (e, l): 3.54, 0.05; L2 (e, l): -1.03, 1.38
L3 (e, l): -1.78, 0.36; L4 (e, l): 0, 0.08

Fig. 1. Learning Curve Plots and AFM Statistics.

Four of the 26 levels had flat learning curves. Among these, Game 3 L3&4 appear too difficult and need to be re-designed to support learning. L3 presents a fraction and requires students to find two different ways of dividing a vertical bar by selecting the number of segments that equals the given fraction, as shown in Figure 2. However, the denominator of the given fraction is not allowed as a choice. For example, if the fraction is 3/4, the option to divide the bar into fourths is grayed out, forcing students to divide the bar into eighths and choose 6/8, and then 9/12. L4 concerns a similar skill with more difficult content. A longitudinal study by Hansen et al. also found questions where the denominator of the fraction did not correspond directly to the pieces in the presented model posed the most difficulties for low-achieving students [9]. Thus, designers may wish to offer extra scaffolding on these two levels. On the contrary, Game 1 L2 and Game 7 L1 appeared to be too easy because performance is consistently near perfect. For example, in Game 7 L1 students order fractions with the same denominator and different numerators, with visualizations showing the size of these fractions as the widths of bars. This level is too easy because students can, without understanding fractions, visually compare the bar widths. However, this level serves to teach the game mechanics for the subsequent levels in this game. Thus, we suggest designers either reduce the number of puzzles, or use one puzzle in this level as a tutorial for L2, instead of as an independent level.
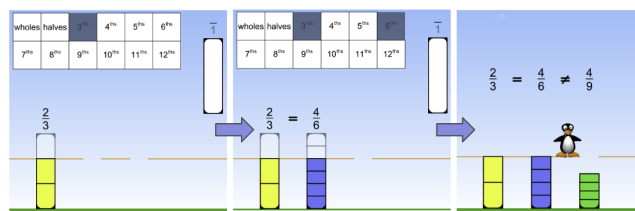

Fig. 2. An Example of a Too Difficult Level with Flat Learning Curve.

Six levels follow a non-learning curve pattern. Two levels (Game 1 L1, Game 2 L4) present the same puzzles at the same attempt for all students to make a specific point, such as understanding fractions equal to 1. These non-randomized puzzles are easier, causing a jump in performance. However, we do not suggest changing them due to their educational value. Four levels showed disjointed learning curves. These learning curves revealed cases where students failed to transfer between specific number content. For example, the first three puzzles of Game 5 L2 require students to locate a fraction X/8 on a number line divided into one fourths, and the last three require locating X/4 on a number line marked with one eighths. Game 5 L3 has a similar setup, with X/6 and X/3. Although these puzzles cover the same concept, the four disjointed learning curves show that students failed to transfer between puzzle sets of different fixed denominators (the four puzzles types shown in Figure 3). It could be that some students do not understand the underlying concepts, but learned pattern matching based

on specific number content. Another possible explanation is that the number of partitions prevented transfer between the two puzzle sets. Mitchell and Horne found that some students may incorrectly count the number of lines to determine where a fraction falls on the number line instead of considering the spaces between the lines [10]. Thus, game designers may consider providing practice with randomized numbers instead of presenting fixed numbers separately to reduce the ease of content-specific pattern matching or counting.
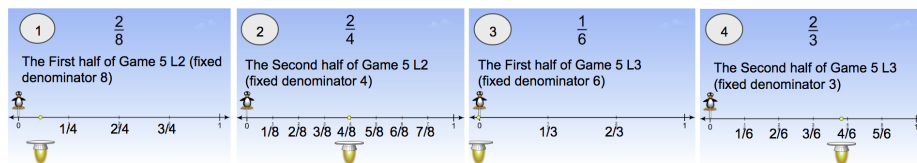
Fig. 3. An Example Where Students Failed to Transfer. The Below Four Types of Puzzles Showed Four Disjointed Learning Curves.

To summarize, fitting learning curves to each game's level derived specific game design insights, and helped us better understand how learning is structured in the ST Math environment. However, a general trend was that a number of the learning curves were disjointed. In many games, each level seemed to form its own learning curve instead of forming a single learning curve with the other levels. The lack of connection and the cases of disjointed learning curves implied that 3rd graders may rely on content- and problem-specific procedural knowledge or pattern matching strategies to solve puzzles, instead of transferring the understanding of underlying math concepts [12, 22]. When new content or problem representations were introduced to practice the same math concept, students treated them as new knowledge components.

4.3   Analyzing Levels in Games

In this section, we sought to find similarities between levels by looking for level combinations that would form a continuous learning curve. Based on the previous analyses, we started by considering each level as a separate Knowledge Component (KC), and applied a bottom-up approach to hierarchically combine levels within a game based on learning curve fitting. We then searched for KC pairs that, once combined as the same KC, yielded AFM models with the lowest Bayesian information criterion (BIC) values as compared to other combination choices. A lower BIC means that the model fit was comparatively better considering both the fit (maximum likelihood) and complexity (number of parameters). This approach is similar to Cen and Koedinger's work [4], except that they used a top-down approach that searches to split one KC into multiple KCs to improve the model. We applied this method to levels within a game instead of across games. We did so because the conjunction of learning curves only makes sense

if the different levels involve practicing the same skill (KC). We excluded levels with disjointed or flat learning curves. This is because disjointed learning curves, if split into multiple learning curves, have too few puzzles to study a pattern. Flat learning curves, especially those with high performance, may be appended to the tail of any previous learning curves that reached high performance in the last attempts. Such conjunctions may not have empirical meanings.

Figure 4 shows the hierarchical combinations of levels in Game 6. The algorithm first combined L1&4, which resulted in a lower BIC than the original model in which each level is a different KC. Both L1&4 ask students to compare two fractions with the same denominator and different numerators, with one requiring students to answer with a ladder and the other with math symbols. Similarly, L2&3 require comparing fractions of the same numerator with ladders or math symbols, and were suggested to be combined next. This hierarchy showed that students can transfer easily between ladder and math symbols, but not as easily from comparing fractions with the same denominator to comparing fractions with the same numerator and different denominators. This is likely due to the simplicity of comparing fractions with the same numerator (or denominator)–students only have to compare one number rather than considering the relationship between the numerator and denominator [6]. In other words, comparing fractions was still a difficult skill and it wasn't the symbolic representation of greater than/less than that tripped students up. Thus, we suggest re-positioning L4 to come after L1 or even removing ladders and using math symbols only. However, when introducing comparing fractions with the same numerator and different denominators in the next level, designers should provide other scaffolding to make connections between the two skills.
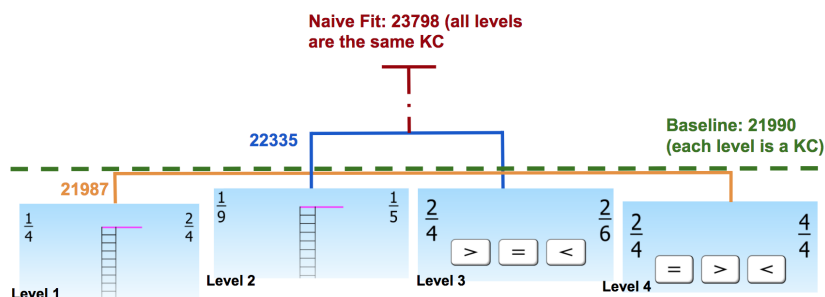

Fig. 4. Hierarchical Combinations of Game 6 Levels that Led to Models with Different BICs. BIC Similar to Baseline Indicates that the Combined Levels Share a Similar KC.

In the rest of the games, the algorithm suggested combining the following levels without much increase in BIC: L2&3 of Game 2 (BIC 14138 compared to baseline 14127), L1&2 of Game 4 (BIC 14731 compared to baseline 14702), and L3&4 of Game 7 (BIC 14425 compared to baseline 14352). The combination in Game 4 shows that by locating fractions one by one on a number line, stu-

dents can easily transfer from comparing fractions with the same denominator and different numerators to comparing fractions with the same numerator and different denominators. This implies that locating fractions on a number line is a strategy that facilitates such a transfer (see the integrated theory of numerical development in [21,25]). However, for comparing fractions with the same denominator, the transfer between fractions smaller than 1 to larger than 1 is much more difficult. It could be that students rely on the numbers in the fractions rather than their magnitude [9,11], so when the fraction is greater than 1, they do not have a conceptual understanding of where the fraction is located on a number line in order to use the number line to help with comparisons. Thus, designers may consider more scaffolding to help students make this transfer.

The other suggested combinations are pairs of levels concerning the same skill with the same problem representation. To summarize, hierarchically combining learning curves helped us identify where students may need extra support to transfer between math skills and problem representations.

4.4   Analyze Games in Objective

From previous analyses, we learned that the performance in ST Math is influenced by both targeted math skills and problem representations. Thus, in this subsection, we sought to further investigate how math skills and problem representations interact with each other and influence students' learning across games. We started with designing an expert Q-matrix (mapping from puzzles to knowledge components) with two types of knowledge component: math skill (KC-S) and problem representation (KC-R). Each level was mapped to at least one KC-S and one KC-R, but not all KCs were mutually exclusive, which means a level could contain multiple KC-Ss and/or KC-Rs. We constructed a total of four KC-Rs, including: number line, vertical bar, horizontal bar, and representation containing visual cues that help students solve the puzzle through pattern-matching. We constructed 6 KC-S, including: presenting fractions (e.g., as segments of a bar); finding equivalent fractions; comparing three types of fractions: same numerators, same denominators, and different in both numerators and denominators; and comparing fractions greater than 1.

Next, we fit learning curves on the expert-designed Q-matrix with three assumptions. The **first** assumption was that each KC contributed to performance *additively*, as assumed by the AFM model. The **second** was that each KC contributed to performance *conjunctively* (modeling the conjunctivity as a multiplication of skill parameters), as assumed by the CFM model. We used R's optim package with BFGS optimization method to estimate the parameters of CFM through maximizing the likelihood function. The **third** assumption was that the KCs interacted neither additively nor conjunctively. This means the same math skill presented in two separate levels would be viewed as two distinct skills that students learned under different representations. Thus, a *combination* of KCs forms a new KC. In other words, each level would be mapped to only one KC, and levels with the same KC-R and KC-S combinations shared the same KC.

When each level had one KC, AFM and CFM were equivalent. This assumption yielded 15 KCs (15 different combinations of KC-R and KC-S) across 26 levels.

Table 1. Learning Models Under Different Assumptions of KC Interactions.

KC Assumptions	Model	Accuracy	AUROC	BIC
High Baseline: Each level is a KC (26 KC, each level has 1 KC)	AFM/CFM	0.782	0.622	134831 (+0.0%)
Combined (15 KC, each level has 1 KC)	AFM/CFM	0.775	0.602	137425 (+1.9%)
Additive (9 KC, each level has >1 KC)	AFM	0.768	0.58	141607 (+5.0%)
Conjunctive(9 KC, each level has >1 KC)	CFM	0.761	0.54	149002 (+10.5%)
Low Baseline: Each game is a KC (7 KC, each level has 1 KC)	AFM/CFM	0.756	0.54	150787 (+11.8%)

As shown in Table 1, our models have low discrimination ability based on AUROC from 10-fold cross-validation. Thus, there are limitations when applying traditional learning models to serious games, where it is easier for children to guess or pattern match with visual cues in specific game environments. The model based on the third assumption had the best fit, with only 1.9% increase in BIC as compared to the most ideal model assuming each level is a KC. Our result implies that ST Math's targeted skills and problem representations do not contribute to performance simply through additive or conjunctive relationships. Instead, the same skill would be treated as different skills when combined with other skills and problem representations. It could be that when students play ST Math, they do not use math skills alone. Rather, students develop new skills that are content- and situation-specific, based on the combinations of targeted math skills and problem representations.

5 Discussion & Conclusion

In this paper, we demonstrated using learning curves as a simple, efficient way to evaluate how well an educational game supports learning. Our results pinpointed problematic levels and cases where students needed extra support to transfer between math skills, content, and problem representations. We derived actionable feedback for ST Math and general insights on fraction learning.

This work has several limitations. First, ST Math is designed as a curriculum-integrated game, but our data does not capture factors in classrooms. Future research will include teacher interviews and classroom observations to better assess the impact of classroom factors. Second, we limited the data to only the number of puzzles required to pass a level which excluded some attempts by low-performing students. Future research may explore methods to separate learning curves for student sub-populations [16] in order to increase external validity.

Our results suggest that students developed new 'skills' based on the combination of targeted math skills and problem representations, rather than simply combining them as assumed in the additive or conjunctive factor models. The variety in ST Math's problem-solving scenarios may improve students' understanding of math concepts. However, this variety could distract learning if students focus more on content- and situation-specific practice than the underlying math concepts. The literature review by Lehtinen and Hannula-Sormune [12] argues that in cases of transfer failures, the (new) situations are not necessarily interpreted as mathematical by children. For example, students may see Game 1-3 as 'selecting divided bars and the number of segments to match a given bar's height/width,' instead of 'understanding fractions as proportions and finding equivalent fractions by multiplying or dividing the numerator and denominator by the same number.' Thus, when bars are replaced with a number line to practice the same math skill, it becomes a different task. Similarly, work by Rau et al. [20] found that providing multiple representations can promote better learning than a single graphical representation, but only when students are prompted to self-explain how the graphics relate to the symbolic fraction representations. With the increasing popularity of mini-games collections, it is important to design scaffolding that facilitates transfer by focusing students on the underlying math concepts instead of reinforcing simple strategies like pattern matching. Such scaffolding should also be considered in other e-learning platforms that offer multiple problem-solving scenarios for young children.

We learned several lessons from applying learning curves to this serious game environment. Learning in game environments is inseparable from the games' mechanisms, structures, and designs. Researchers should consider starting analyses at a low granularity, such as the individual level we used here. Understanding students' learning at a low granularity would help illuminate factors that contribute to learning, and help structure analyses at higher granularities where these factors may combine or evolve. Moreover, game performance does not solely comprise learning. This means traditional learning modeling methods may have limited power in serious games. Thus, researchers should be flexible with different models and assumptions to work within specific game environments. Regardless, researchers should triangulate results with human interpretations and the literature to make sure the results do not derive from unforeseen game scenarios or the large amount of data. For example, Harpstead and Aleven [8] used both data and human judgment to examine the fit of learning curves in a physics game and identified an unforeseen pattern matching strategy. Liu et al. [19] mined predictive relationships between ST Math objectives, and used both human interpretation and literature to suggest game design feedback. Thus, when analyzing serious game data, it is extremely important to not solely focus on the performance of models, but also consider the models' interpretation and practical value.

References

1. Adkins, S. (2016). The 2016-2021 Global Game-based Learning Market. In Serious Play Conference.
2. Bailey, D. H., Hoard, M. K., Nugent, L., and Geary, D. C. (2012). Competence with fractions predicts gains in mathematics achievement. Journal of experimental child psychology, 113(3), 447-455.
3. Baker, R. S., Habgood, M. J., Ainsworth, S. E., and Corbett, A. T. (2007, July). Modeling the acquisition of fluent skill in educational action games. In proceedings of the International Conference on User Modeling (pp. 17-26). Springer, Berlin, Heidelberg.
4. Cen, H., Koedinger, K., and Junker, B. (2006, June). Learning factors analysisa general method for cognitive model evaluation and improvement. In proceedings of the International Conference on Intelligent Tutoring Systems (pp. 164-175). Springer, Berlin, Heidelberg.
5. Cen, H., Koedinger, K., and Junker, B. (2008, June). Comparing two IRT models for conjunctive skills. In proceedings of the International Conference on Intelligent Tutoring Systems (pp. 796-798). Springer, Berlin, Heidelberg.
6. Rittle-Johnson, B., Siegler, R. and Alibali, M. (2009). Students' fraction comparison strategies as a window into robust understanding and possible pointers for instruction. Educational Studies in Mathematics, 72(1), 127-138.
7. El-Nasr, M. S., Drachen, A., and Canossa, A. (2016). Game analytics. Springer London Limited.
8. Harpstead, E. and Aleven, V. (2015, October). Using empirical learning curve analysis to inform design in an educational game. In proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play. (pp. 197-207). ACM.
9. Hansen, N., Jordan, N. C., and Rodrigues, J. (2017). Identifying learning difficulties with fractions: A longitudinal study of student growth from third through sixth grade. Contemporary Educational Psychology, 50, 45-59.
10. Mitchell, A. and Horne, M. (2008). Fraction number line tasks and the additivity concept of length measurement. In proceedings of the 31st Annual Conference of the Mathematics Education Research Group of Australasia. (pp. 353-360).
11. Jordan, N. C., Resnick, I., Rodrigues, J., Hansen, N., and Dyson, N. (2017). Delaware longitudinal study of fraction learning: Implications for helping children with mathematics difficulties. Journal of learning disabilities, 50(6), 621-630.
12. Lehtinen, E., and Hannula, M. M. (2006). Attentional processes, abstraction and transfer in early mathematical development. Instructional psychology: Past, present and future trends. Fifteen essays in honour of Erik De Corte, 49, 39-55.
13. Liu, Z., Cody, C., Barnes, T., Lynch, C., and Rutherford, T. (2017). The Antecedents of and Associations with Elective Replay in An Educational Game: Is Replay Worth It. In proceedings of the 10th International Conference on Educational Data Mining (EDM).
14. Loh, C. S., Sheng, Y., and Ifenthaler, D. (2015). Serious Games Analytics. Edited by Christian Sebastian Loh, Yanyan Sheng, and Dirk Ifenthaler. Cham: Springer International Publishing. 10, 978-3.
15. Lomas, D., Patel, K., Forlizzi, J. L. and Koedinger, K. R. (2013, April). Optimizing challenge in an educational game using large-scale design experiments. In proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 89-98). ACM.

16. Murray, R. C., Ritter, S., Nixon, T., Schwiebert, R., Hausmann, R. G., Towle, B., Fancsali, S., and Vuong, A. (2013, July). Revealing the learning in learning curves. In proceedings of the International Conference on Artificial Intelligence in Education (pp. 473-482). Springer, Berlin, Heidelberg.

17. Newell, A., and Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. Cognitive skills and their acquisition, 1(1981), 1-55.

18. Pavlik, P. I., Yudelson, M., and Koedinger, K. R. (2015). A measurement model of microgenetic transfer for improving instructional outcomes. International Journal of Artificial Intelligence in Education, 25(3), 346-379.

19. Peddycord-Liu, Z., Cody, C., Kessler, S., Barnes, T., Lynch, C. F., and Rutherford, T. (2017, October). Using Serious Game Analytics to Inform Digital Curricular Sequencing: What Math Objective Should Students Play Next?. In proceedings of the Annual Symposium on Computer-Human Interaction in Play (pp. 195-204). ACM.

20. Rau, M. A., Aleven, V., and Rummel, N. (2009, July). Intelligent Tutoring Systems with Multiple Representations and Self-Explanation Prompts Support Learning of Fractions. In AIED (pp. 441-448).

21. Rinne, L., Ye, A., and Jordan, N. C. (2017). Development of fraction comparison strategies: A latent transition analysis. Developmental Psychology, 53(4), 713-730.

22. Rittle-Johnson, B., Siegler, R. and Alibali, M. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. Journal of Educational Psychology, 93(2), 346-362.

23. Romero, C. and Ventura, S. (2010). Educational data mining: a review of the state of the art. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 40(6), 601-618.

24. Rutherford, T., Farkas, G., Duncan, G., Burchinal, M., Kibrick, M., Graham, J., Richland, L., Tran, N., Schneider, S., Duran, L., and Martinez. M.E (2014). A randomized trial of an elementary school mathematics software intervention: Spatial-temporal math. Journal of Research on Educational Effectiveness, 7(4), 358-383.

25. Siegler, R., Thompson, C., and Schneider, M. (2011). An integrated theory of whole number and fractions development. Cognitive psychology, 62(4), 273-296.

26. Siegler, R. S., Duncan, G. J., Davis-Kean, P. E., Duckworth, K., Claessens, A., Engel, M., Ines Susperreguy, M., and Chen, M. (2012). Early predictors of high school mathematics achievement. Psychological science, 23(7), 691-697.

27. Siemens, G., and Baker, R. S. D. (2012, April). Learning analytics and educational data mining: towards communication and collaboration. In proceedings of the 2nd international conference on learning analytics and knowledge (pp. 252-254). ACM.

28. Torbeyns, J., Schneider, M., Xin, Z., and Siegler, R. S. (2015). Bridging the gap: Fraction understanding is central to mathematics achievement in students from three different continents. Learning and Instruction, 37, 5-13.