Using Serious Game Analytics to Inform Digital Curricular Sequencing: What Math Objective Should Students Play Next?

Zhongxiu Peddycord-Liu
North Carolina State University
zliu24@ncsu.edu

Christa Cody
North Carolina State University
cncody@ncsu.edu

Sarah Kessler
North Carolina State University
skessle@ncsu.edu

Tiffany Barnes
North Carolina State University
tmbarnes@ncsu.edu

Collin F. Lynch
North Carolina State University
cflynch@ncsu.edu

Teomara Rutherford
North Carolina State University
taruther@ncsu.edu

ABSTRACT
This paper applied serious game analytics to inform digital curricular sequencing in a longitude, curriculum-integrated math game, ST Math. When integrating serious games into classrooms, teachers may have the flexibility to change the order of math objectives for student groups to play. However, it is unclear how teacher decisions, as well as the sequencing of the original curricular order affect students. Moreover, few researchers have applied data-driven methods to inform content ordering in educational games, where the nature of educational content and student behaviors are different from many e-learning platforms. In this paper, we present a novel method that suggests curricular sequencing based on the prediction relationship between math objectives. Our results include specific design recommendations for ST Math, and general data-driven insights for digital curricular design, such as the pacing of objectives and the ordering of math concepts. Our method can potentially be applied to data from a wide range of games and digital learning platforms, enabling developers to better understand how to sequence educational content.

ACM Classification Keywords
H.5.2. User Interfaces: Evaluation/methodology; K.3.1. Computer-assisted instruction(CAI); K.8.0. Personal Computing: Games

Author Keywords
Serious Games; Serious Game Analytics; Digital Curricular

INTRODUCTION
The rapid growth of the serious game industry allows us to collect longitudinal gameplay data at large scale. These data capture learners' in-game behaviors and choices that go beyond information collected from traditional, shorter-term experimental trials. Such user-interaction data can offer new insights for HCI research, as it help us understand large-scale user interactions, and enable data-driven methods that inform game design in a scalable, efficient, and automatic manner.

Serious game analytics [21] (also called Game Learning Analytics [10]) is an emerging field that captures the potential of user-interaction data in serious games. This research grounds game design in students' learning and performance on targeted skills, whereas traditional game analytics [9] focuses on player enjoyment. In learning analytics we analyze data to understand and improve learning in a wide range of digital environments beyond games. These fields share similar and potentially transferable methods, but have different primary purposes. More research is needed to verify the applicability of these methods across these research domains.

This paper applies serious game analytics to evaluate the content sequencing in a year-long supplemental digital mathematics curriculum–Spatial Temporal (ST) Math. Educational research has shown that mathematics is inherently hierarchical–early mathematics skills are needed for later mathematics [8, 28]. Due to this hierarchical nature, children tend to follow a certain trajectory as early skills build upon each other to form a more complex understanding of mathematics [6]. Thus, in serious games that aim to help children develop complex interdependent skills of this type over a long timespan, math objective sequencing is an important design consideration to evaluate.

We designed a novel method to understand curricular sequencing by modeling the potential predictive relationships between objectives. We demonstrated this method on gameplay logs from 1,565 3rd graders from 17 schools and 78 classrooms during the 2012-2013 school year. Students completed different sets of objectives and often did so in unique orders depending upon their classroom structure and individual needs. Based on our new data-driven predictive models, we determined specific recommendations for curricular improvements, including the appropriate location of objective groups, the content that could

CHI-PLAY '17, October 15–18, 2017, Amsterdam, Netherlands

ACM ISBN 978-1-4503-4898-0/17/10...$15.00
DOI: https://doi.org/10.1145/3116595.3116620

be used to prepare for objectives, the pacing and the ordering of objectives, and re-design of game puzzles. Our method can potentially be applied to data from other serious games and e-learning environments that are intended to be used over a long timespan.

RELATED WORK

Serious Game Analytics and Game Design

In recent years, there has been increasing interest in applying data mining and data analytics to serious games. Increasingly, researchers have showed that serious game analytics is powerful, and can help discover how gameplay patterns relate to game design. Horn et al.[16] clustered players' action traces, and gained insight on how game mechanics can impact students' reflection on their gameplay, which impacted learning. Harpstead and Aleven [13] applied learning curve analysis from educational data mining to a physics game. By predicting students' error rates, they discovered an unforeseen short-cut strategy where students succeeded in using a game mechanism, but without a deep understanding of the physics principles. Similarly, Hicks et al.[15] applied a zero-inflation model to compare user-created content in a programming game. They found that the mechanisms of the game's different content creation tools directly impacted the aesthetics, affordances, and educational value of user-created content. Hicks and colleagues [14] also applied survival analysis and interaction networks to evaluate data from a physics game, and found problem spots where students frequently dropped out due to rough difficulty/complexity progression. These papers demonstrate that serious game analytics can help game designers better achieve their educational goals. Such data-driven insight may not be gained from traditional control-trial studies.

Data-driven Method for Curricular Design

There has been a growing interest in using data-driven approaches to inform curricular design, especially the sequencing of educational content.

At the skill granularity, researchers have informed educational content sequencing through mining prerequisites between skills. Chen et al. [4] applied Structural Expectation Maximization to learn the optimal Bayesian network structures that represent skill prerequisites using students' performance from textbook exercise for an English Proficiency test. Chen et al. [5] applied Association Rule Learning on the same educational context, which assumed that one skill is a prerequisite for another if the probabilities of mastering skill A given that B is mastered, or not mastering skill A given that B is not mastered, were both above a certain threshold. However, these methods require a significant amount of effort to map skills to educational content. This mapping process also faces the risk of expert blind-spots–pedagogical organization based on "the structure of the domain rather than the learning needs of novices." [25]

At the problem set granularity, Doroudi et al. [7] assigned each student's activity trajectory to violations of sequencing constraints, and compared how different sequencing constraints affected performance. However, it is difficult to design similar constraints for curricula with longer time spans and more varied content. Vuong et al. [30] applied binominal tests in all possible pair-wise relationships among curriculum units in an intelligent tutor. The binominal test compared the performance on unit A between students who mastered and had not mastered unit B, and deemed unit B as a prerequisite of unit A when significance was found in the test. Although this work created a heatmap describing the existence and strength of relationships, it lacked analysis of the meaning and implications of these relationships.

At the course granularity, Ochoa, Yang et al. and Moretti [26, 32, 24] applied data-driven methods on high-level data such as enrollment, web-mined course descriptions, and online ratings to inform the selection and sequencing of college courses. At the system granularity, Pechenizkiy et al. and Pechenizkiy and Toledo [27] designed conceptual frameworks for the cyclic process of curriculum mining and designing.

However, little work has applied data-driven methods to the curricular sequencing in serious games. In this paper, we focused on the problem-set granularity. ST Math implemented a pre-designed curriculum, but the sequence of students playing math objectives (problem sets) varied in practice. Our challenges were to analyze objective trajectories that varied both in length and sequence over long time-spans, and to derive meaningful data-driven insights for this serious game environment, where the nature of learning and play may differ from other e-learning technologies.

ST MATH AND DATA

ST Math

ST Math is designed to act as a supplemental program to a school's existing mathematics curriculum. In ST Math, mathematical concepts are taught through spatial puzzles within various game-like arenas. ST Math games are structured at the top level by objectives, which are 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 start 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." Students can only see this objective. and must complete a pre-test before beginning the content. Once students have completed the pretest, they can start the first sub-objective (game). 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. Figure 1 illustrates the hierarchy of ST Math content and examples.

As with many games, students are given a set number of 'lives' per level. Every time they fail to complete a puzzle correctly, they lose one life. If all of their lives for a given level are exhausted, then they will be required to replay the level before they can move on to new content. Once a student has passed a level, they can elect to replay it at any time. Once a student has passed every level within an objective, then they

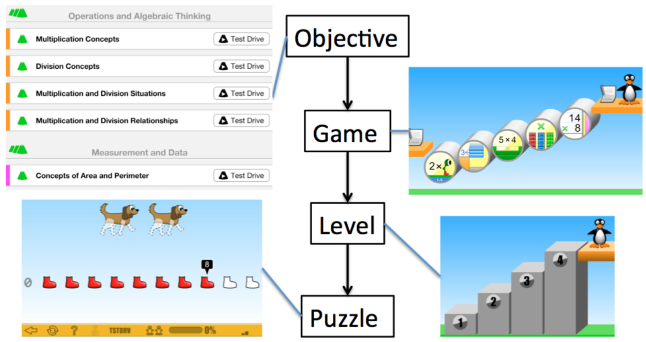

Figure 1. ST Math Content and Examples

will be permitted to take the objective-level post-test. The objective pre- and post-tests consist of 5-10 multiple choice questions related to the objective, and are parallel in both the question format and the content's difficulty. Students cannot progress to the next objective until they have completed the latest objective's post-test.

Participants and Data

MIND Research Institute, the developer of ST Math, collected and provided to the researchers gameplay data from 1,565 3rd grade students during the 2012-2013 school year. These students came from 17 schools and 78 classrooms. Among these students, 50.7% were male (2.9% NA), 80.7% were eligible for free/reduced lunch program (2.8% NA), 84.8% were hispanic or latino (2.8% NA), and 66.2% were English Language Learners (ELLs) (2.8% NA), and 10.9% were noted as having a disability (2.1% NA). For each attempt at a level (level attempt), ST Math logged the student's ID, timestamp, and the number of puzzles completed.

For the purpose of this analyses, we filtered out three types of level attempts: students' elective replay on previously passed levels (1.69%), level attempts from default objectives where all students completed less than 25% of the objectives' levels (0.54%), level attempts from 15 optional objectives provided by ST Math, and the last default objective (3.12%). We filtered out optional objectives because very few students attempted them. We filtered out the last default objective because it was a "challenge" level with spatial puzzle games that were not designed to teach math concepts. After filtering, students attempted(completed 25% levels) an average of 15.82 default objectives (sd=8.07, 95% CI: [15.42, 16.22]), with 371 students from 51 classes completing at least 25% of levels from each of the 26 default objectives. All of the remaining data (94.65%) from 1,565 students were included in the analyses.

Objective Sequencing in ST Math

The 3rd grade's pre-designed curriculum contains 27 default objectives in a particular order, as shown in Table 2. Teacher activities are the root cause of the variance in the order of objectives played by students. A teacher can separate his/her class into groups, and reorder objectives in the curriculum for each student group. For example, if some students missed the first class, a teacher can put them in a separate group, and assign what the rest of the class are currently playing first, and the missed objectives later. A teacher can also allow certain students to 'skip' an objective, which means playing the next objective before completing the current objective. Meanwhile, students can revisit skipped (allowed by teacher) objectives at any time. This causes variance in the order of objectives played by individual student.

METHOD

We first extracted objective sequences as the order of objectives that students played throughout the year. Each objective occurred only once in the sequence, representing where most of the students' effort (level attempts) occurred. For example, if a student attempted the first level of objective A, and was 'skipped' by the teacher to objective B, completed all levels in objective B, and went back to complete the rest of the levels in objective A, then objective B is ordered before objective A because most of the gameplay happened before objective A. Note that without teacher-allowed 'skipping,' a student cannot play the next objective until they have completed the latest objective's post-test. Thus, except for the rare skipping case, a student cannot complete two objective simultaneously.

Table 1 reports the statistics from objective sequences. Our data have over a hundred different objective sequences, with few sequences occurred in high frequencies. This was largely due to the nature of sequencing in ST Math–players followed a pre-designed curriculum order, with variance introduced by individual activities.

Next, we sought to understand the relations between math objectives, and to provide data-driven insight on the role and ordering of objectives. The first step was to select a single feature to measure student learning from an objective. We used percentage of levels completed within an objective, multiplied by the objective performance. We defined ***objective performance*** to be the average performance across the given objective's levels, and a level's performance as the average ***level attempt performance*** prior to and including the first attempt that passed the level. A ***level attempt performance*** is the percentage of puzzles a student completed before losing all lives on a single attempt of a level. The number ranges from two to four lives depending upon the number of questions in the level and their difficulty. Prior research has shown that this measurement can be used to predict normalized gains on the students' 2012 and 2013 state math test scores [20]. These tests provide standardized measurements for students' math ability outside this game environment. We chose to not use the

Table 1. Statistics of the Sequence of Objectives 3rd Grade Students Played in ST Math

N	#Students Played$\geq$N Objs	#Diff. Orders Students Played the First N Objs	#Students in the Top 5 Most Frequent Order
5	1414	107	550,214,89,69,37
10	1104	194	322,131,64,54,28
15	876	205	263, 74, 48, 39,22
20	595	156	186,47,41,24,20
26	371	84	129,34,30,20,19

Table 2. A Description of Objectives in the Curriculum and Objectives with Predictive Power Over Each From Regression Analyses

SO1E-Equal Group	Figure out N, X, Y for scenarios that split N objects or building block into X groups of Y items
SO1F-Multiplication Concept	Complete a multiplication formula with given objects; Represent a multiplication formula as addition using item, number line, addition formula; Calculate result of simple multiplication with help of building block
SO1O-Place Value Concept	Represent three digit numbers as flower petals in group of 1,10,100
SO1P-Place Value Bundles	Represent three digit numbers as flower petals in group of 1,10,100; Addition of numbers to form building blocks in unit of 10s and Hundreds
SO2-Order & Compare Whole Number	Locate number on number line; Compare number with/out help of items and number line; Round numbers up and down on number line
SO1Y-Add & Subtraction w Regrouping	Add and subtract three-digit numbers represented as flower petals, as numbers with flower petals, and as numbers with flower petal animations; Regroup flower petals in calculation.
SO28-Division Concept	Divide countable objects into X groups equally
SO29-Division	Calculate simple division; Present answer through objects, building blocks, and number line
SO5-Lines & Angles	Match types of angles with English name; Identify parallel and perpendicular lines from given lines
SO2I-Shapes	Match 2D shape objects with English name; Identify edge numbers for shapes, objects, and shape's English name; Draw shapes with pins and a rubber band; find shape pairs with same number of edges
SO2J-Shape Attributes	Identify number of edges, vertices, faces in 3D geometry
SO8-Algebraic Expressions & Equations	Represent equation with objects; separate numbers and units objects into equal groups; Identify missing operations from equation
SO9-Functional Relationship	Find X when split N into X groups of Y objects; Fill in missing number of linear pattern (e.g. 2,4,?,8) with/out help of countable objectives
SO2S-Fraction Concepts	Split bar into equal areas; Adding parts of pie chart; Represent fraction with bar or pie chart; And represent pie chart as fraction
SOB-Fraction Addition & Subtraction	Represent fraction's denominator and numerator as parts of robot, slices of pie chart; Add and subtract with these countable objects; Locate calculation result on number line with unit 1
SOC-Fraction Addition & Subtraction L1	Add and subtract fractions without countable objects; calculate on fractions greater than 1
SOD-Money & Decimals	Recognize money unit and coins; Compare money amount; Add, subtract, and find combinations of paper money and coins equal to certain amount
SOE-Measurement	Compare weight of items using balance and scale; measure with ruler; Translate between hours and minutes, and between inch and foot, and specify the answer on number line
SO3C-Concepts of Area and Perimeter	Calculate perimeter and area of 2D shapes presented with/out square grid or fill in calculation formula; Identify 2D shapes for given perimeter and area; Select grids based on difference of 2D shape area and given number
SO3D-Volume & Weight	Order objects' weight using balance and scale; Indicate volume of a given 3D object in cubic
SOG-Addition & Subtraction to 10,000	Calculate multi-digit addition and subtraction; Represent answer on number line or unit block; Fill in missing digit given the result of a multi-digit addition and subtraction
SO3M-Number Patterns	Identify or fill in linear pattern (e.g. 10, 20, 30, 40) with help of building block;Use data-table representation for addition, subtraction, multiplication test
SOI-Multiplication of Multiple Digit	Use building block (grid) to represent multiplication; Calculate multi-digit multiplication with 2 dimensional representation of building block
SOJ-Fraction & Decimal Equivalence	Add and subtract fraction&decimal, present result in 10-by-10 grid; Locate fraction&decimal on number line of unit 1; Represent decimal as formula of fraction addition; Use coins to present dollars in decimal number
SOK-Outcomes	Distinguish whether countable object is "impossible", "unlikely", "likely" or "certain" to be picked from a jar and from lucky wheel
SOL-Using Data & Graph	Present numbers of objects in bar chart and pie chart, and read number from the charts

pre- and post-tests in ST Math in this analysis due to a their lack of precision. The pre- and post-tests contain 5-7 multiple choice questions, so a lucky guess and a mistake could cause a 14-20% change in score. The pre and post-tests also have missing data (11.8% in pre- or post- test) and ceiling effects (19.0% achieved full pre-test scores).

In the second step, it was necessary to focus on the most influential objectives. To do so, for each objective (we refer to as goal objective), we conducted linear regression analyses to find which prior objectives were most predictive of its objective performance. To avoid over-fitting, the first step was to filter the objectives included in the regression. To do so, we applied Spearman Correlation for each goal objective, against all the objectives that happened before it. To ensure there were enough observations, we applied a power test, and filtered out objective pairs with fewer than 30 observations. This step left us 430 correlations for 26 goal objectives. Given the fact that all objective performances were influenced by a student's general math ability, our correlation table had an average value of 0.39. Thus, we decided to apply linear regression with only those objectives with a correlation of at least 0.39. This resulted in considering only 189 of the 430 possible objectives-goal pairs.

In the third step, we applied stepwise mixed effect linear regression for each goal objective, as:

$$P_g = \beta_0 + \beta_1 M + \beta_2 I_g + C \sim N(0, \sigma^2) + \sum_{i=1}^{k} \beta_i P_{gi} \quad (1)$$

Each regression included only the students who attempted the goal objective. The dependent variable P_g is the percentage of levels completed, multiplied by the averaged objective performance of the goal objective g. Betas are coefficients for independent variables, derived from the regression. M is the state math test in 2012, which measured students' general math ability prior to playing ST Math. It was also necessary to include classroom C as a random effect to account for students' being taught by different teachers, especially given the importance of teachers in deciding upon the objective ordering of their classes and groups in classes. The position of objective j in each student's objective sequence is denoted by I_g. We assumed there is a linear relationship between how many objectives students played in ST Math, and the ability to perform well in the same game environment. $P_{g1} - P_{gk}$ describe the k objectives found from the previous step: happened before goal objective g in a sufficient number of sequences (at least 30 and passed the power test for correlation), correlated with g ($r > 0.39$), and were independent of each other ($VIF < 1.5$). Based on our outcome measurement of learning, P_{gi} is 0 if the student did not do an objective i before the goal objective. We also used a stepwise process to automatically add or remove variables based on AIC (Akaike information criterion). This process helped select variables that had the most predictive power for goal objectives, under the influence of student's general math ability, classrooms, and the sequential position of the goal objective. This step also avoided overfitting through cross-validation.

After applying the Benjamini-Hochberg Correction for multiple statistical tests, a total of 96 predictor objectives were found to be statistically significant predictors for the 26 goal objectives. The coefficients had an average standardized beta of 0.12, with 92 coefficients being positive. Figures 2 and 3 display the results, as discussed in the following subsections.

RESULTS AND INTERPRETATION

This section presents the results from two angles: categorization of objectives, and categorization on objective pairs. These categorizations were interpreted in relation to our research goal–what data-driven insight we can provide to inform the design of digital curricular sequencing.

Categorization of Objectives

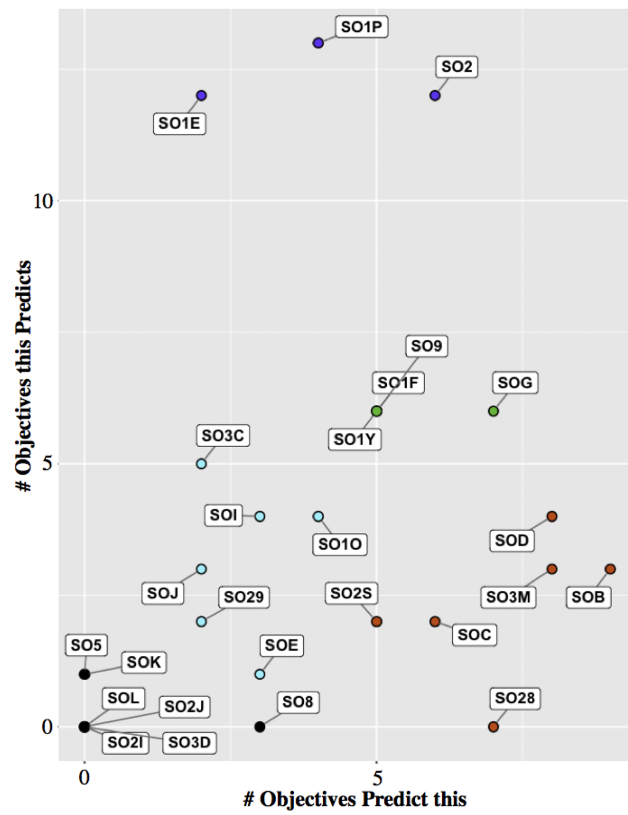

Figure 2. Objectives, and the # of objectives they predict, and the # of objectives that predict them.

Figure 2 is a plot of the predictive power and predictability of objectives, categorized by k-means clustering. Objectives in Group Dark Blue predicted a high number of other objectives. These objectives also occurred early in the pre-designed curriculum, which suggests that they captured basic skills needed in the later objectives. Objectives in this group could be used as early indicators of later performance. For example, if a student skipped or performed poorly on these basic objectives, the teacher may consider intervening to prevent poor performance in later objectives.

Objectives in Group Red were predicted by a high number of other objectives. This implies that playing objectives in this group required math skills learned from many other objectives. For example, games in SOD:"Money and Decimal" required students to compare, add, and subtract money in the form of both paper money and coins. Performing these tasks required a variety of math skills including unit transformation, calculation, number comparison, and a preliminary understanding of fractions and decimals, because coins represent fractions of a dollar. Additionally, three objectives in this group (SO3M:"Number Patterns," SOD:"Money and Decimals," and SOB:"Fraction Addition & Subtraction") were among the top five most difficult, in that students took the most attempts to pass levels within these objectives. This indicates that foundational skills may be especially important for these objectives, and struggle within them may indicate to teachers that students lack prerequisite knowledge.

Objectives in the Group Green both predicted a high number of objectives and were themselves predicted by a high number. This implies that the math skills covered in these objectives are more advanced (depend on many prerequisites), and useful (as other objectives depend on them). For example, both SO1Y and SOG involved multi-digit addition and subtraction tasks. These tasks required a solid understanding of numbers and units, and strong calculation skills. Playing puzzles in these objectives, in return, enhanced understanding on numbers and provided practice in the calculation skills needed for many other objectives. This indicates that the Green Group objectives were difficult, but should be completed before moving on because they could affect later performance. In other words, if some students weren't doing well on these objectives, we would not recommend that teachers skip these students forward. Instead, the teachers should help the students learn this material rather than allowing them to advance in the game.

The majority of objectives in Group Black (SOL, SO5, SO2I, SO2J, SO3D, SOK) were not predicted by any objectives, and were also not predictive of any objectives. This implies that these objectives practiced math skills that were isolated from other objectives. For example, SO5, SO2I, SO2J involved tasks such as identifying shapes, faces, edges and types of angles. This suggests that Group Black objectives can be moved around more flexibly in the curriculum, and may be used at the beginning of the curriculum to familiarize students with the ST Math system.

Categorization of Objective Pairs

Figure 3 presents the prediction relationship between objective pairs, symmetrically along the X=Y line to show complete information of an objective when reading from a row or column. A grey color shows a ***symmetrical relationship***: when X is attempted first, X predicts Y; when Y is attempted first, Y predicts X. A blue or red color shows an ***asymmetrical relationship***: one objective predicts another if attempted first, but not vise versa.

Within asymmetrical relationships, a light blue or red shows a ***curriculum-order asymmetrical relationship***: the objective attempted first and which predicts the other, is also ordered earlier in the curriculum. This relationship may exist because either the later-ordered objective was not chosen as a predictor of the earlier-ordered objective, or because we simply do not have enough data (failed power-test) to tell so. For either reason, this relationship suggests keeping the current ordering of the objective pairs in the curriculum.

On the other hand, a dark blue or red shows a ***non curriculum-order asymmetrical relationship***: the objective attempted first and which predicts the other is ordered later in the curriculum. This relationship implies that students who completed more levels and performed better in the later-ordered objective before attempting the earlier-ordered objective performed better in the earlier-ordered objective. These relationships are the most interesting as they may suggest potential changes to the existing curriculum. Below sections categorize the 20 ***non curriculum-order asymmetrical relationships*** we found based on their implications for curriculum design.

Suggesting more Practice on Certain Skills (8/20)

This group includes: SOG: "Addition & Subtraction to 10,000," predicted SO1F, SO9, SOB, and SOD; SOI: "Multiplication of Multiple Digit" predicted SOB and SOC; SO1Y: "Addition & Subtraction with Regrouping" predicted SO2; and SOJ: "Fraction & Decimal Equivalence" predicted SOE: "Measurement." One explanation is that these predictor objectives aimed to intensively practice a specific skill, such as multi-digit addition. These skills were needed in the earlier objectives they predicted. For example, SOJ:"Fraction & Decimal Equivalence" predicted SOE:"Measurement," where the predictor SOJ practiced the understanding and transformation between fractions and decimals. This skill was needed in SOE, where many games required students to place certain minutes on a number line of unit hours, or certain inches in a number line of unit feet. A solid understanding of fractions and decimals gained from SOJ may help transforming minutes and inches as fractions of hours and feet. However, the relationship is asymmetrical because the predicted objectives were not designed for enhancing the specific calculation skills in the predictor objectives. Thus, they were not the major influencers of the performance and completion of their predictors.

Although relationships in this group do not necessarily suggest that these calculation-based objectives should be moved earlier in the curriculum, these relationships imply that the performance and completion of many objectives could be improved if there were more practice on certain calculation skills beforehand.

Suggesting more Applicable Examples (4/20)

This group includes: SOD: "Money & Decimals," predicted SO8, SO2S; and SO3C: "Concepts of Area and Perimeter," predicted SO29, SOB. One explanation is that tasks in SOD and SO3C contained examples and practical scenarios for the more conceptually challenging objectives they predicted. For example, SOD practiced the recognition, comparison, and calculation of money units such as dimes, quarters, and dollars. Because coins are fractions of a dollar, these tasks provided examples for SO2S: "Fraction Concepts." Another example is SO3C practiced calculating area and perimeter for a given square grid or highlighting areas in a square grid for a given area and perimeter. These tasks were practical scenarios to practice addition, multiplication, and division at small-scale, with the help of countable squares. Thus, SO3C predicted

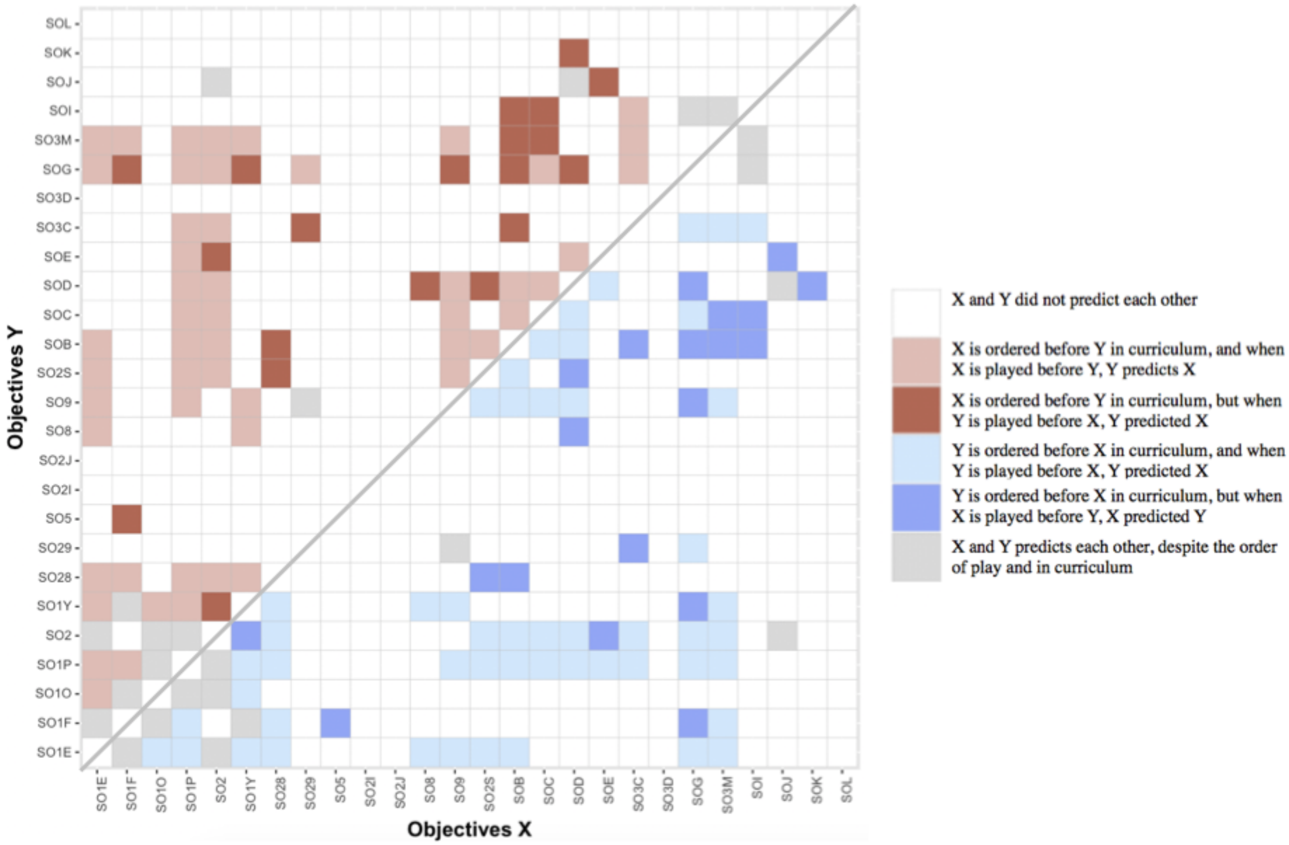

Figure 3. Pairwise Relationship between Objectives. This figure is symmetric across the X=Y line (so light pink = light blue; dark pink = dark blue) Light pink/blue relationships are expected because the curriculum is sequenced purposefully. Dark pink/blue relationships are surprising, since a later curricular item predicts performance on an earlier item if attempted early

SO29 and SOB. Relationships in this group suggest that performance and completion for some conceptually difficult objectives can be improved with more applicable, simplistic examples beforehand [2].

One interesting result is that although SOD: "Money / Decimals" predicted SO2S:"Fraction Concepts," the addition and subtraction of fractions predicted SOD (curriculum-order symmetrical relationship). This may be because the later levels in SOD required more complicated calculations. For example, some tasks asked students to add dimes and quarters to a double-digit amount in dollars, which could be helped by practice on fraction addition, and on the transformation between fractions and decimals (as SOD and SOJ predicted each other). These levels contained much harder tasks than other levels in SOD, as we found that these four levels in SOD cost students an average of 5.14 attempts to pass, whereas the other 12 SOD levels cost an average of 1.71 attempts. Thus, there may be a benefit in splitting SOD. An example would be moving the earlier levels for recognizing and comparing coins before fractions so these levels could serve as practical scenarios for using fractions, and moving the harder levels after or as levels within fraction calculation objectives.

Suggesting Reordering of Teaching Math Concepts (2/20)
SO2S: "Fraction Concept" and SOB: "Fraction Addition and Subtraction" predicted SO28: "Division Concept." Although it is conventionally believed that division should be taught before fractions, with prior research finding that division predicts fractions skills (e.g., [19, 1], there may be reasons why the specific actions within these objectives of ST Math produced the opposite order. SO2S and SOB presented fractions as countable parts of a divisible object, such as slices of a pie or groups of bricks. Manipulating these objects' parts may help students develop a better understanding of division through partitioning items into groups in a familiar way with smaller numbers [22]. This relationship is asymmetrical because games in the fraction objectives generally involved the addition and subtraction of divisible objects, which were too simple to involve division calculations.

Suggesting a Change in Game Design (2/20)
SOG: "Addition and Subtraction within 10,000" predicted SO1Y: "Addition and Subtraction with Regrouping." It could be that if a student practiced SOG and became more proficient with calculations, s/he may have more working memory to perform the extra counting and regrouping tasks in SO1Y [17]. In SO1Y, digits were presented as petals in groups of one, 10, and 100. In the majority of tasks, students needed to count and

regroup petals during calculation, with some counting above 10. In certain tasks, students also needed to present their calculation results in the format of petals. On the other hand, SOG practiced calculation with more traditional presentations, such as the numbers themselves, number lines, or building blocks marked with one, 10, or 100. SOG also did not require students to regroup objectives during calculation or to count over 10.

This relationship suggests that counting and regrouping petals during the calculation process may be ineffective and distracting. This is evidenced by the pre- and post-test scores. Recall that the pre- post-tests in ST Math consist of parallel questions (same problem type with different numbers). In both SO1Y and SOG, The tests ask student to choose the correct answer when adding or subtracting multi-digit numbers. In SO1Y: "Addition and Subtraction with Regrouping," the average pre- and post-test scores were 0.64 and 0.66; in SOG: "Addition and Subtraction within 10,000," they were 0.66 and 0.73–a much higher learning gain. It could be that students experienced the split-attention effect defined by Chandler and Sweller[3], because students spent their limited working memory resources mentally integrating two sources of informations (e.g., number and petals). Also, this relationship is not symmetrical (SO1Y does not predict SOG), which could be because the performance and completion of SO1Y reflected more about the difficulties encountered in counting and regrouping petals, than in the calculation shared with SOG. This difficulty and ineffectiveness of regrouping and counting petals is also evidenced by statistics from SO1P (Place Value Bundles using petals). SO1P had the second lowest level attempt performance among all 26 objectives, and only a 0.03 gain from pre- to post-test.

Another relationship that suggests a change in game design is SOE: "Measurement," which negatively predicted SO2: "Comparing Whole Numbers." One likely explanation is that the number line was presented differently in these two objectives. In SOE, the number line was represented by a ruler, scaled for time units and for imperial and metric length units. This means that in SOE, a number line can be one centimeter, 60 minutes, or 12 inches. On the other hand, SO2 presented the number line as metric units. Because young students often do not have an accurate understanding of the number line [12], introducing conflicting number lines may create misconceptions and interfere with their number line knowledge. This hypothesis is evidenced by the performance statistics in SO2. The first level of SO2, which requires students to locate an integer on a number line, took students an average of 5.5 attempts to pass, whereas the rest of the levels in SO2 (some contained similar tasks of a locating number on number line) only took students an average of 1.5 attempts to pass, with the max being 2.6 attempts. This relationship suggests that using different units on the same presentation of a number line may be confusing, and similar practices on unit conversion should be moved after students have developed a concrete understanding of the number line.

Suggesting a Change in Pacing (2/20)

SO3M: "Number Patterns" negatively correlated with objectives SOB: "Fraction Addition Subtraction L1" and SOC: "Fraction Addition and Subtraction I and II." It could be due to the pacing in the most common sequence, in which SO3M was played before SOB and SOC. The majority, 293 out of 360 students who played SO3M before SOB, played in a sequence "SO28 SO3M SO29 SO9 SO2S SOB SOC." SOB and SOC required calculation in conceptually difficult topic in third grade curricular–fractions. SO3M contained difficult tasks, such as asking students to find or complete number patterns of 3-6-9-12, or performing multi-digit multi-step addition, subtraction, and multiplication using a new data-table representation. As compared to the pre-designed curriculum in Table 2, these three relatively difficult concepts, division, number patterns, and fractions, were spaced too close to each other. Thus, it could be that students get mentally overloaded when too many difficult concepts and different number representations are packed together. This relationship suggests that teachers should vary the pacing accordingly and take breaks between difficult concepts [31]. For example, teachers can allow students to play simpler objectives that offer more concrete examples or practical scenarios to prepare for the upcoming difficult concepts. Teachers can also assign objectives in the Group Black described in in previous, as they practiced concepts that are isolated from other objectives.

Unclear (2/20)

SO5: "Lines and Angles" predicted SO1F: "Multiplication Concepts." One possible explanation is that SO1F was introduced very early in the curriculum, with few objectives beforehand to explain its variance in performance and completion. SO5 required selecting the right names for angle types and distinguishing parallel and perpendicular lines; it is a simple objective that could be captured by general skills, such as memorization, that helped predict SO1F. Another unclear relationship is that SOK: "Outcomes" predicted SOD: "Money." SOK dealt with probability problems. It could be that the presentations in SOK closely resembled lucky wheels and lottery jars, which both relate to money, but the relationship is still unclear.

DISCUSSION AND CONCLUSION

This paper presents a novel method that informs curricular sequencing in a serious game, taking advantage of the variance in the order of objectives that students played. We applied this method to ST Math gameplay data from 1,565 3rd graders. Our results include specific design recommendations for ST Math, and general data-driven insights for digital curricular design, such as the pacing of objectives and the ordering of math concepts. The method itself, and some of these data-driven insights can inform design for digital curricular design in other serious games and e-learning platforms.

From a technical perspective, the growth of educational games and other e-learning platforms enables a large amount of data collection. This in turn has facilitated the use of innovative analytical methods that reveal user behavior patterns, which go beyond traditional experimental trials. However, data analytics, although informative, can make it easier to find statistically-significant results that may not be grounded in theory, and simply arise because of the sheer volume of available data. Therefore, when analyzing any large data set such as ST Math, we want to ensure that our analyses help uncover educationally relevant results. From the education literature, researchers have identified some early skills that predict later mathematics skills. For example, the general knowledge of the meaning and magnitude of numbers, is highly predictive of later mathematics achievement, including number line representation and calculation fluency [18]. These skills are in turn predictive of skills such as fractions [1, 19]. We believe that our data-driven work on discovering predictive relationships can be leveraged to discover more learning trajectories[29], as well as to make practical sequencing recommendations that may not have been apparent from a purely top-down approach.

From a practical perspective, this analysis helps guide one crucial decision when integrating serious games into a classroom: what objective should students play next? When applying digital curriculum in classrooms, teachers may be given the flexibility of reordering and skipping math objectives. However, it is unclear whether teachers are aware of the impact of objective selection and sequencing[11]. Molin [23] also suggested that teachers not only have a limited time to prepare and play a game for game-based learning, but are seldom engaged with game designers in making decisions on content ordering. Thus, our method can help teachers to become more aware of how their local decisions may affect students' longer-term performance. This motivates future work on designing a user interface to notify teachers of the likely impact of individual ordering decisions. It is possible that there exists some optimal pathway through which students should experience a given game. Unfortunately, our method could not be used to find the optimal path in present context, as our data have too much variance both in objective selected and those completed. In future work, we plan to develop quantitative metrics to compare individual paths against better orderings and to relate those variances to performance.

Serious game designers can use our method and its data-driven insights to evaluate the sequencing and predictive relationships within their games. Many of our results are applicable beyond ST Math. For example, we found that counting objects during calculation distracted students, which implies that game designers should avoid designing 'extra' interactions that split students' attention during problem solving. Other results may be limited to ST Math's unique design characteristics. For example, fraction puzzles used countable parts of a divisible objects, such as divided pie charts, which helped students to understand division. Such a predictive relationship may or may not hold in games that use different representations. To assess the impact of differences between game environments, game designers can follow the same analytical procedure, and redefine performance variables and performance-associated mixed effects based upon their user-interaction data. Our method can potentially be applied to similar data outside of the serious games domain, where content is presented under varying sequences over a long time span.

One limitation of this work is that there are many factors outside of the system, such as varying instructional methods, that may affect learning and performance. We sought to control these factors by adding random effect distributions for individual classrooms. We also worked with domain experts to interpret the results, and examined the candidate predictive relationships by playing the games. Future work could collect qualitative data on teachers' understanding of these data-driven insights and the effects of their applications in classrooms.

ACKNOWLEDGEMENT

This work was supported by NSF grant IUSE #1544273 "Evaluation for Actionable Change: A Data-Driven Approach" Teomara Rutherford, Tiffany Barnes & Collin F. Lynch Co-PIs.

REFERENCES

1. D. Bailey, R. Siegler, and D. Geary. 2014. Early predictors of middle school fraction knowledge. *Developmental Science* 17, 5 (2014), 775–785.
2. S. Bokosmaty, J. Sweller, and S. Kalyuga. 2015. Learning geometry problem solving by studying worked examples: Effects of learner guidance and expertise. *American Educational Research Journal* 52, 2 (2015), 307–333.
3. P. Chandler and J. Sweller. 1992. The splitâĂŘattention effect as a factor in the design of instruction. *British Journal of Educational Psychology* 62(2) (1992).
4. Y. Chen, J. Gonzalez-Brenes, and J. Tian. 2016. Joint discovery of skill prerequisite graphs and student models. In *Proceedings of the 9th International Conference on Educational Data Mining (EDM)*.
5. Y. Chen, P. H. Wuillemin, and J. M. Labat. 2015. Discovering Prerequisite Structure of Skills through Probabilistic Association Rules Mining. In *Proceedings of the 8th International Conference on Educational Data Mining (EDM)*.
6. D. Clements and J. Sarama. 2014. *Learning and teaching early math: The learning trajectories approach*. Routledge.
7. S. Doroudi, K. Holstein, V. Aleven, and E. Brunskill. 2016. Sequence matters, but how exactly? A method for evaluating activity sequences from data. In *Proceedings of the 9th International Conference on Educational Data Mining (EDM)*.
8. G. Duncan, C. Dowsett, A. Claessens, K. Magnuson, A. Huston, P. Klebanov, L. Pagani, L. Feinstein, M. Engel, and J. Brooks-Gunn. 2007. School readiness and later achievement. *Developmental Psychology* 43, 6 (2007), 1428.
9. M. S. El-Nasr, A. Drachen, and A. Canossa. 2016. *Game analytics*. Springer London Limited.
10. M. Freire, A Serrano-Laguna, B. M. Iglesias, I. Martinez-Ortiz, P. Moreno-Ger, and B.

Fernandez-Manjon. 2016. Game learning analytic: Learning analytics for serious games. *Learning, Design, and Technology* (2016).

11. J. Galant. 2013. Selecting and sequencing mathematics tasks: Seeking mathematical knowledge for teaching. *Perspectives in Education* 31(3) (2013), 34–48.

12. D. Geary, M. Hoard, L. Nugent, and J. Byrd-Craven. 2008. Development of number line representations in children with mathematical learning disability. *Developmental Neuropsychology* 33, 3 (2008), 277–299.

13. E. Harpstead and V. Aleven. 2015. 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 (CHI PLAY)*. ACM, 197–207.

14. D. Hick, M. Eagle, E. Rowe, J. Asbell-clarke, Edwards T., and Barnes T. 2016. Using game analytics to evaluate puzzle design and level progression in a serious game. In *Proceedings of the 6th international conference on learning analytic and knowledge (LAK)*.

15. D. Hicks, Z. Liu, and T. Barnes. 2016. Measuring gameplay affordances of user-generated content in an educational game. In *In Proceedings of the 9th international conference on educational data mining (EDM)*.

16. B. Horn, A. K. Hoover, J. Barnes, Y. Folajimi, G. Smith, and C. Harteveld. 2016. Opening the black box of play: Strategy analysis of an educational game. In *In Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY)*. ACM, 142–153.

17. I. Imbo, A. Vandierendonck, and E. Vergauwe. 2007. The role of working memory in carrying and borrowing. *Psychological Research* 71, 4 (2007), 467–483.

18. N. Jordan, J. Glutting, and C. Ramineni. 2010. The importance of number sense to mathematics achievement in first and third grades. *Learning and Individual Differences* 20, 2 (2010), 82–88.

19. N. Jordan, N. Hansen, L. Fuchs, R. Siegler, R. Gersten, and D. Micklos. 2013. Developmental predictors of fraction concepts and procedures. *Journal of Experimental Child Psychology* 116, 1 (2013), 45–58.

20. Z. Liu, C. Cody, T. Barnes, C. Lynch, and T. Rutherford. 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)*.

21. C. S. Loh, Y. Sheng, and D. Ifenthaler. 2015. *Serious game analytics. Methodologies for performance, measurement, assessment, and improvement*. Springer, Cham, Switzerland.

22. T. Martin and D. Schwartz. 2005. Physically distributed learning: Adapting and reinterpreting physical environments in the development of fraction concepts. *Cognitive Science* 29, 4 (2005), 587–625.

23. G. Molin. 2017. The role of the teacher in game-based learning: A review and outlook. In *In serious games and edutainment applications*. Springer International Publishing, 649–674.

24. A. Moretti, J. Gonzalez-Brenes, and K. McKnight. 2014. Data-driven curriculum design: Mining the web to make better teaching decisions. In *Proceedings of the 7th International Conference on Educational Data Mining (EDM)*.

25. M. Nathan, K. Koedinger, and M. Alibali. 2001. Expert blind spot: When content knowledge eclipses pedagogical content knowledge. In *Proceedings of the 3rd International Conference on Cognitive Science*.

26. X. Ochoa. 2016. Simple Metrics for Curricular Analytics. In *Proceedings of the 1 st Learning Analytics for Curriculum and Program Quality Improvement Workshop*.

27. M. Pechenizkiy and P. A. Toledo. 2014. Learning to teach like a Bandit.. In *Proceedings of the 7th International Conference on Educational Data Mining (EDM)*.

28. G. Ramani and R. Siegler. 2011. Reducing the gap in numerical knowledge between low-and middle-income preschoolers. *Journal of Applied Developmental Psychology* 32, 3 (2011), 146–159.

29. R. Shavelson and A. Kurpius. 2012. Reflections on learning progressions. In *Learning progressions in science*. Springer, 13–26.

30. A. Vuong, T. Nixon, and B. Towle. 2011. A method for finding prerequisites within a curriculum. In *Proceedings of the 4th International Conference on Educational Data Mining (EDM)*.

31. B. Witzel and P. Riccomini. 2007. Optimizing math curriculum to meet the learning needs of students. *Preventing School Failure: Alternative Education for Children and Youth* 52, 1 (2007), 13–18.

32. Y. Yang, H. Liu, J. Carbonell, and W. Ma. 2015. Concept graph learning from educational data. In *Proceedings of the 8th International Conference on Web Search and Data Mining*. ACM.