

Improving User Interface and User Experience of MathSpring Intelligent Tutoring System for Teachers

Master's Thesis

By Neeraj S. Menon (nsmenon@wpi.edu)

Advisor: Prof. Ivon Arroyo (iarroyo@wpi.edu)

Reader: Prof. Joseph Beck (josephbeck@wpi.edu)

Computer Science Department

Worcester Polytechnic Institute

100 Institute Rd

Worcester, MA 01609

Table of Contents

Abstract.....	6
Introduction.....	8
Background Research	9
Method of Evaluation	11
Participants	12
First Iteration: The Old Teacher Tools	12
Detail per Student Report.....	14
Class Summary Per Student Per Topic	15
Class Summary Per Problem.....	17
Class Summary per Common Core Report.....	18
Second Iteration: Prototype I of the New Teacher Tools	20
Teacher Response to Prototype I	21
Class Summary Per Student.....	22
Class Summary Per Student per Problem Set	24
Class Summary Per Problem.....	26
Class Summary per Common Core.....	28
Third / Final Iteration: Prototype II of the New Teacher Tools	29
Class Summary Per Student.....	30
Class Summary Per Problem.....	37
Class Summary per Common Core.....	37
Discussion and Conclusion.....	44
References.....	63

Table of Figures

Figure 1: Iterative and Incremental software development process.....	12
Figure 2a. Problems with the interface of the old teacher tools, according to a senior math teacher.....	13
Figure 2b. Teacher comments regarding information displayed by the Report per Student in the Old Teacher Tools.....	15
Figure 3a. Class Summary Per Student Per Topic report in the Old Teacher Tools, showing mastery and problems completed.....	16
Figure 3b. Class Summary Per Student Per Topic report in the Old Teacher Tools, showing mastery and problems completed, for a student that appears to be learning.....	17
Figure 4. The Class Summary Per Problem Report in the Old Teacher Tools.....	18
Figure 5. The Class Summary Per Common Core Report, highlighting standards that a class appears to be weaker at.....	19
Figure 7a. Teacher comments regarding information displayed in the Class Summary per Student in Prototype I of the New Teacher Tools.....	23
Figure 7b. Teacher perceptions of the Class Summary per Student Report in Prototype I of the New Teacher Tools; a. Perceived Usefulness; b. Sample qualitative comments.....	24
Figure 8. Teacher comments regarding information displayed in the Class Summary per Student per Problem Set in Prototype I of the New Teacher Tools.....	25
Figure 9. Teacher comments regarding information displayed in the Class Summary per Student per Problem Set in Prototype I of the New Teacher Tools.....	26
Figure 10. Teacher comments regarding information displayed in the Class Summary per Student per Problem in Prototype I of the New Teacher Tools.....	27

Figure 11. Teacher comments regarding information displayed in the Class Summary per Student per Problem in Prototype I of the New Teacher Tools. 28

Figure 12. Teacher usefulness ratings regarding the Class Summary per Student per Problem in Prototype I of the New Teacher Tools..... 29

Figure 13a. Emotions in the Class Summary per Student Report in Prototype II of the New Teacher Tools: a. Initial View when opening this report, including engagement and emotion chart icons; b. Detailed charts of students’ emotions, and their reported reasons they attribute to those emotions..... 31

Figure 13b. Student History in the Class Summary per Student Report in Prototype II of the New Teacher Tools: a. Clicking on arrow reveals allow different time slices of history per student; b. Clicking on Problems solved most recently reveals the student history for the last day MathSpring was used by individual students. 32

Figure 13c. Student Engagement in the Class Summary per Student Report in Prototype II of the New Teacher Tools. Clicking on the “Effort Chart” icon reveals a stacked bar chart that shows percent of specific behaviors on problems..... 33

Figure 14a. Teachers’ Perception of the Class Summary per Student Report in Prototype II of the New Teacher Tools..... 34

Figure 14b. Teachers’ Comments about the Class Summary per Student Report in Prototype II of the New Teacher Tools..... 35

Figure 15a. Teacher usefulness ratings regarding the Class Summary per Problem Set Report in Prototype II of the New Teacher Tools. 36

Figure 15b. Teacher comments regarding the Class Summary per Problem Set Report in Prototype II of the New Teacher Tools. 37

Figure 16a. The Per Common Core Report in Prototype II of the New Teacher Tools. a. Common Core Clusters may be expanded via arrows, and cells that indicate poor performance on a problem are highlighted; b. Descriptions of the standards of a problem come up on roll-over; c. The web page that comes up from MASS DOE when clicking on the question mark for “see similar problems”. 39

Figure 16b. Aggregate Effort Chart in the Per Common Core Report in Prototype II of the New Teacher Tools. 40

Figure **17a**. Teacher usefulness ratings regarding the Class Summary per Student per Problem in Prototype II of the New Teacher Tools..... 41

Figure **17b**. Teacher comments regarding the Class Summary per Student per Problem in Prototype II of the New Teacher Tools. 42

Figure **18**. Teacher comments regarding the Class Reports Prototype II of the New Teacher Tools. 42

Abstract

Common goals of Educational Data Mining are to model both student knowledge as well as student affect. While research continues along these lines of gathering data and building models of students' changing knowledge and affect states, little is being done to transform this collected (Raw) data into meaningful entities that are more relatable to teachers, parents and other stakeholders i.e. people who aren't researchers. This research entails design and development of Teacher Tools created with input from teachers and other experts. Teacher Tools is a web application designed within MathSpring Intelligent Tutoring system.

In our study, we redesigned the existing version of the Teacher Tools in three iterations based on the feedback gathered during each of those phases. The feedback isolated from the first iteration clearly suggested for multiple design level changes with respect to math content organization, the interface, and complexity level of the existing class-level reports. Responses to Prototype I during the second iteration, designed on the basis of responses from the first iteration, were met with mixed to positive responses regarding the usability and interface aspects. Experts at this point suggested further areas of improvement from a usability standpoint, which resulted in Prototype II of the Teacher Tools. Prototype II was then subjected to the third and final iteration; it was well received by a set of teachers and other experts, who thought that Prototype II was useful to them. Teacher comments were positive in general with this final iteration, as they saw the use that they could give to these Teacher Tools to understand their students better and guide future action plans that would alter their teaching based on information about their students' behavior, performance and affect.

We have created a product that summarizes rich information from data logs into visualizations and other representations that are useful to teachers in Middle Schools in Massachusetts, so that they would use this information to change their teaching plans.

Keywords: Teacher Tools, Mastery, Emotion Charts, Teacher Dashboards, Learning Technologies, Mathematics Education

Introduction

MathSpring is a personalized intelligent tutoring system that assists 5-10th grade students' in studying mathematics. Students obtain proficiency in different mathematical concepts by practicing various problems and benefiting from several features such as virtual character tutoring, receiving hints and obtaining personalized math problems of the "right" level of challenge. The system tracks and records all of these activities in a central database. There is great potential for teachers to understand and interpret summaries of this data to help understand their students better, and feedback this information to alter their instruction. For instance, general class-wide performance or item based performance for each class could greatly help teachers determine how hard the students are finding the material, where are they lagging behind and so on. However, while learning scientists and other researchers already use this data to understand behavioral patterns as students' knowledge and affect changes and develops, these datasets tend to be more complex for teachers or parents without any expertise in data analysis to comprehend. This is part of a more general data science problem of how to summarize and visualize rich event logs of user behavior as understandable and valuable sources of information, so that informed action plans can be generated by users.

Background Research

Data Science and Data Analytics are current major trend in all fields that collect some sort of data involving human subjects (Donoho, 2015). In the case of education (Learning Analytics), computer based tutoring systems that are web-based allow the collection of vast amounts of data on their student – users (e.g. ASSISTments, Heffernan & Heffernan, 2014).

In a report released by US department of Education on “Enhancing Teaching and Learning through Educational Data Mining and Learning Analytics”, education stakeholders are encouraged to use educational data mining and learning analytics to improve student learning. The report suggests for development of a mindset of using more student data strategically (Bienkowski, Feng, Means 2012). The report says that student data provides the opportunity for educators to obtain something useful and actionable towards teaching and learning. This suggests that real time access and easy to understand visual representations of student learning metrics should be made available to teachers, to help them make informed decisions regarding their learning instructions, with the potential to be truly helpful in their teaching decisions. This is in opposition to the current reality of teachers, where scores on an achievement test taken six months ago does not tell them how to help a particular student tomorrow. On the other hand, no real research has been done yet on understanding what questions teachers may want to answer via Data Analytics, nor what could be most useful towards affecting their practices. This research is a first step in that direction.

ASSISTments is a tutoring system with a Teacher Dashboard (Heffernan & Heffernan, 2014). Teachers receive feedback on the performance of each individual student and of the class as a whole. With such knowledge, teachers can identify weak students needing additional help

and support understanding the concept while also identifying areas where the class as a whole is lagging. The latter area can be addressed during class time when the instructor can respond to questions and address student misconceptions and lack of comprehension. For the former areas, teachers may choose to intervene with the system to adjust student learning pace or may assign additional learning materials targeting the skills that are not yet mastered. MathSpring.org is also web based online tutoring system focusing on helping middle/high school students studying math and preparing for standardized math tests, such as the Massachusetts Comprehensive Assessment System (MCAS). MathSpring is used in public schools in Massachusetts, offering students various features including but not limited to posing questions for students to solve, offering hints, and providing custom feedback to students. Different to ASSISTments, the system is designed to track and capture the cognitive, metacognitive and motivational aspects of a student's state of mind as they progress through the various problems encountered, and personalizes the difficulty of math problems based on an "effort-based tutoring" approach (Arroyo et al., 2014). While this math practice software is "intelligent" enough to statically estimate the degree of understanding and knowledge of student and sequence/align the problems accordingly for individual students, teachers do not necessarily understand what has happened –what students have learned, what material they have seen exactly, or what major shortcomings students may have had as they used MathSpring and solved math problems. The "teacher tools" in MathSpring play an important function in this regard. As teachers log in with their own usernames and passwords to MathSpring, they can:

1. Set up MathSpring to deliver certain kinds of content (configuration component)
2. Obtain reports about what students have done.

This is especially important because teachers might make decisions (alter their teaching practices) based on this information, creating a loop of information and decisions that feeds back into their teaching and student learning. The goal of this research is to re-design and develop a sub-system within Mathspring that teachers may use to understand students' state of knowledge as well as trace students' affect (mood and emotion) based on all the data gathered while the students are using MathSpring. This research, attempts to seek answers to the following questions:

1. How can we generate dashboards for teachers that are useful to teachers' teaching, by presenting valuable visualizations about their students' performance, and affect?
2. How can we generate data visualizations so that are understandable to teachers? Can teachers faithfully interpret what the data is revealing?

Method of Evaluation

The proposed design was modeled to follow Iterative and incremental software development plan. The system underwent three iterations. Quantitative and qualitative data collected from surveys and interviews with the participants was analyzed at the end of each iteration. The responses received at the end of each iteration became guidelines for the following iteration.

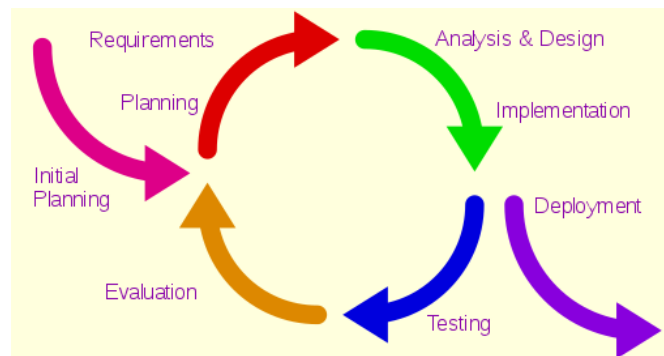


Figure 1: Iterative and Incremental software development process

Participants

Grade level Math teachers comprised our main subject matters of research. The software design and development model followed principles of an iterative and incremental design (Figure 1), where $N=10$ expert math teachers were invited to participate and provide feedback at each iteration, including a senior expert math teacher who had committed time to advise on the development of this software.

First Iteration: The Old Teacher Tools

Researchers designed the existing Teacher Tools based on general notions, empirical studies and state of art. An expert senior math teacher reviewed this first version of the Teacher Tools, after using the software with her seventh and eighth grade intervention classes for about 4 or 5 times in each of the classes. The teacher then evaluated the class reports and usefulness of the Teacher Tools in general against data available for students for this class. A qualitative analysis of the feedback given was collected in written form, which suggested that the current form of the Teacher Tools did not meet the needs of math teachers in K-12 public schools. It also suggested specific needs of teachers that should be met by this tool. This detailed analysis of teacher tools by the senior teacher was the inspiration for the research and became the principal guideline for the next iterations. That summary is included in Appendix A, and only the main points are included into this section.

Appendix A shows that the teacher talked about issues regarding math content organization and other aspects that go beyond the design of the teacher tools. After isolating comments regarding the design and presentation of the teacher tools, we found that main themes

of concern regarded: a) the interface, and b) complexity in understanding the existing class reports.

Figure 2a shows some main problems with the general look and feel of the user interface, according to the senior teacher.

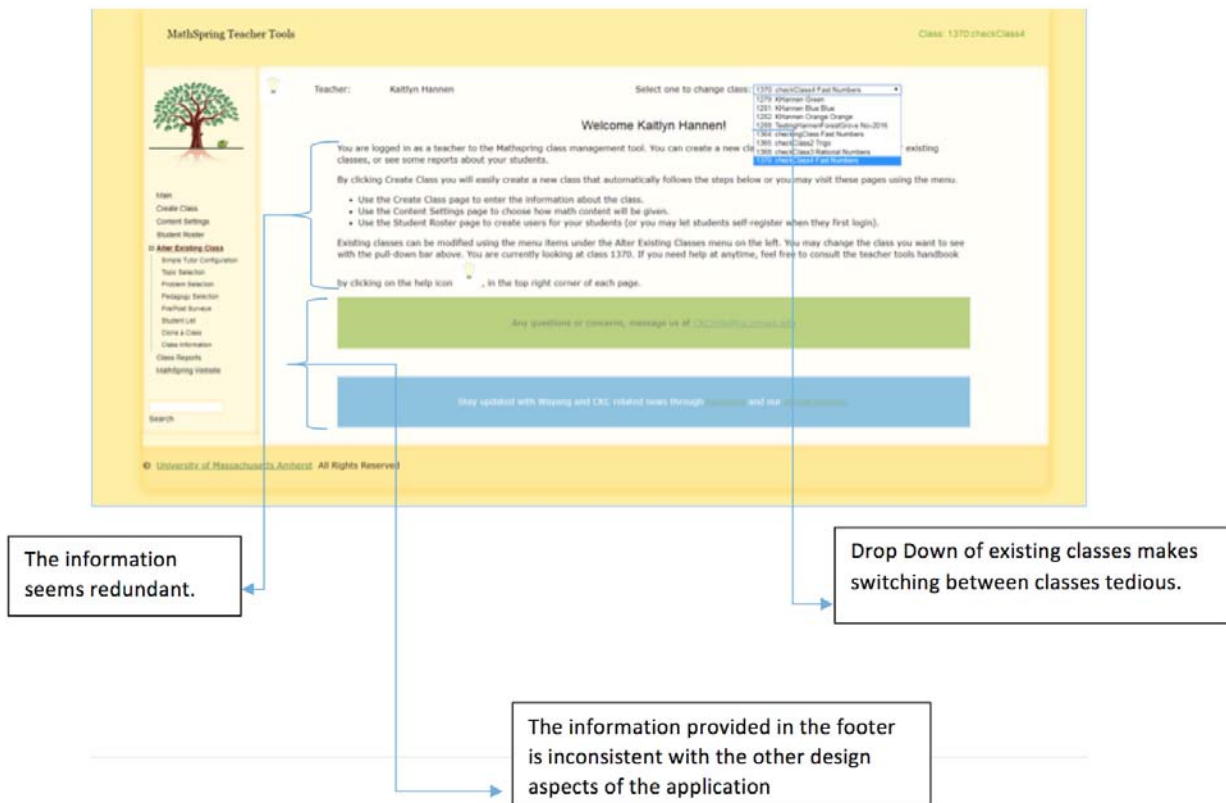


Figure 2a. Problems with the interface of the old teacher tools, according to a senior math teacher.

One general comment given by the expert teacher regarding all of the class reports summarizes the general feel of this teacher regarding the old teacher tools: “I am having trouble

seeing how I can take the reports and get a snapshot of what my students know and don't know. Some of the data may be more helpful to the people running the study but not to me as a teacher" and "The data in the class reports I feel benefits those analyzing the usefulness of the system and not the teachers". The teacher was of the opinion that the old Teacher Tools were serving a purpose of gathering data that was more suitable for researchers and not teachers. The remaining comments themes that follow regard specific details of each individual report presented to the teacher.

Detail per Student Report

The Detail per Student showed a summary of all the problems seen over the course of all student logins, for an individual student. The feedback we got by the expert teacher on this report was "I don't believe that as a teacher this report is very useful". We analyzed this report of the old teacher tool version and agreed ourselves with many aspects of the feedback given by the teacher. Specific comments are in Figure 2b, which associate the individual comments to each section of the report. In general, information is too broad to understand what is going on with a student.

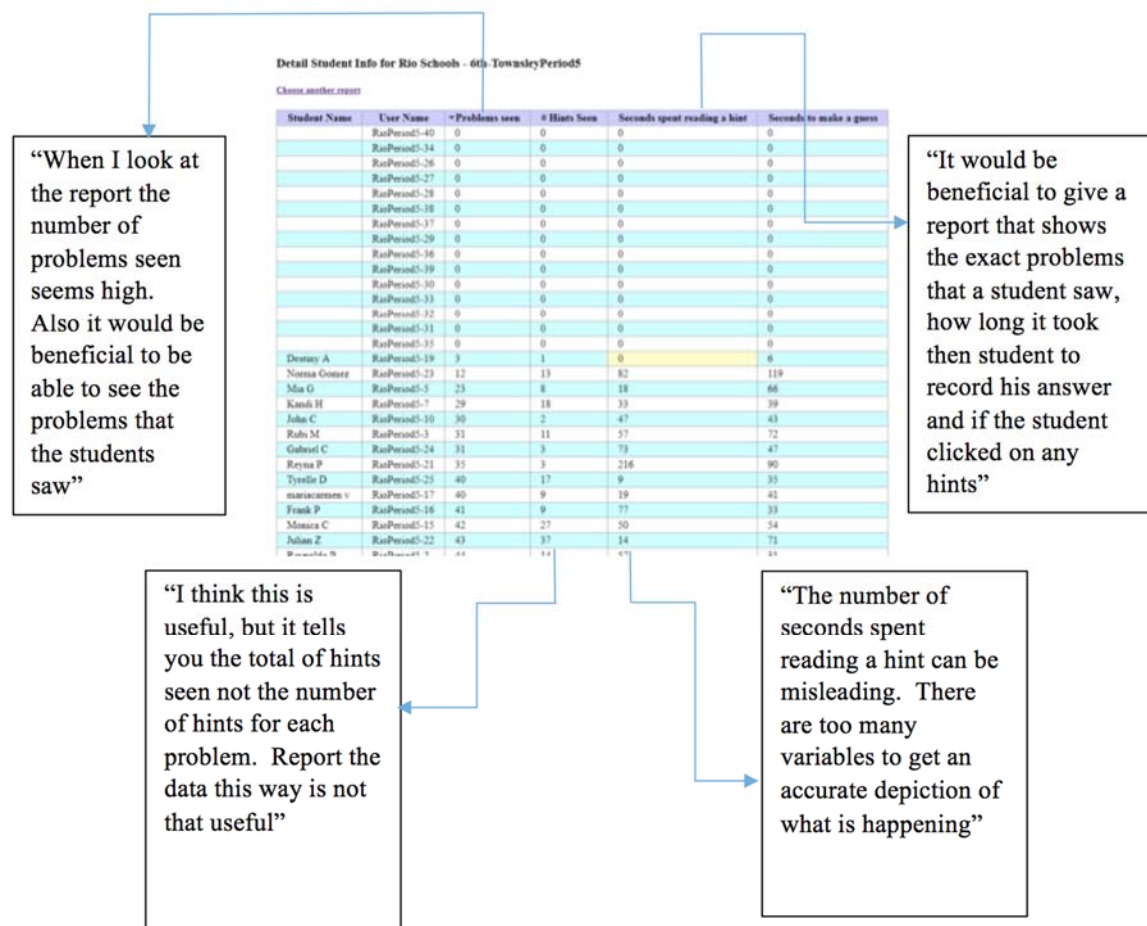


Figure 2b. Teacher comments regarding information displayed by the Report per Student in the Old Teacher Tools.

Class Summary Per Student Per Topic

The Per Student Per Topic Report showed how the students are doing on each topic along with the “mastery” acquired on each topic. The feedback from the teacher suggested for “...more information on how the “mastery” component is calculated.” The feedback also suggested for more understandable information on the mastery projection or line graph projected when a topic is clicked. This can be inferred from the feedback comment “When I click on the topic, a line graph pops up that shows the number of problems seen and the average process mastery. I am not sure what this graph is telling me”.

Figure 3a shows the old teacher tools version of this report, which highlights that it is not clear what “mastery” concretely means, nor how it could be beneficial for teacher’s assessment and knowledge about their students. Figure 4 shows the graph that pops up when clicking on an individual cell of the 3b report.

■ Mastery level < 0.25
■ Mastery level < 0.50
■ Mastery level < 0.75
■ Mastery level >= 0.75

Each cell contains the number correct out of the number of problems given in the topic.

Student Name	User Name	Fractions	Number Sense	Whole numbers	Segments and Distances	Expressions	Decimals and Percents	Perimeter	Expre wi Vari
Student DT_200	DT_200	0.88 [19/21]		0.08 [12/15]			0.17 [8/47]		
Student DT_213	DT_213	0.86 [28/30]		0.47 [20/21]		0.12 [56/64]	0.74 [11/13]		
Student DT_208	DT_208	0.84 [26/27]		0.36 [15/18]	0.05 [7/9]			0.17 [7/9]	
Student DT_201	DT_201	0.81 [21/24]		0.85 [7/10]	0.27 [15/18]				
Student DT_217	DT_217	0.78 [30/33]		0.83 [16/18]	0.46 [9/11]				
Student DT_209	DT_209	0.66 [22/24]		0.86 [7/12]					
Student DT_215	DT_215	0.61 [24/26]		0.82 [20/21]	0.14 [3/9]				
Student DT_223	DT_223	0.60 [9/13]	0.69 [16/20]		0.30 [15/9]				
Student DT_218	DT_218	0.51 [13/17]		0.05 [1/20]	0.36 [18/19]			0.20 [6/7]	
Student DT_207	DT_207	0.46 [9/15]		0.17 [13/16]	0.27 [14/17]			0.20 [5/8]	
Student DT_211	DT_211	0.41 [7/9]		0.74 [20/21]					
Student DT_204	DT_204	0.29 [30/31]	0.17 [55/64]	0.08 [11/14]			0.07 [7/11]		
Student	-- --	0.24	0.06	0.13			0.65		

Figure 3a. Class Summary Per Student Per Topic report in the Old Teacher Tools, showing mastery and problems completed.

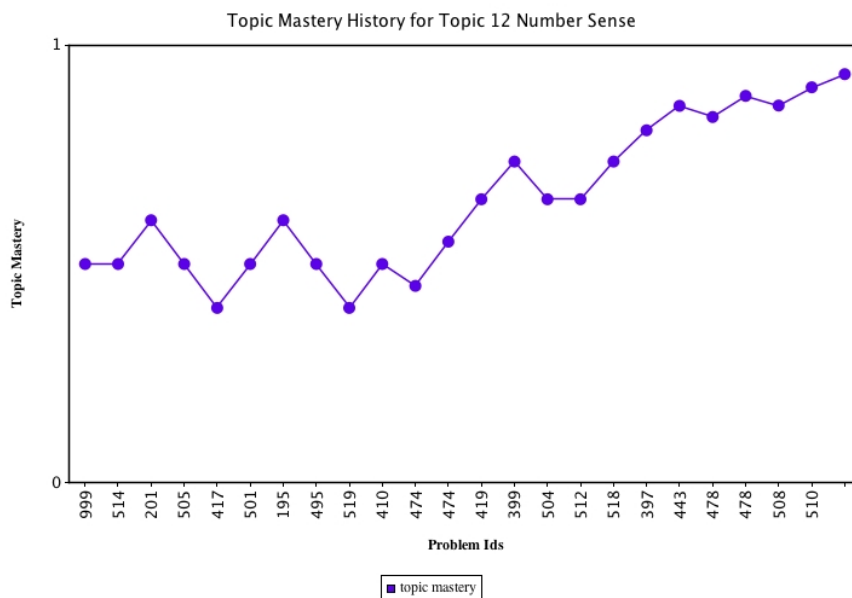
Topic Mastery history for student DT_267[Choose another report!](#) [Choose another class](#)

Figure 3b. Class Summary Per Student Per Topic report in the Old Teacher Tools, showing mastery and problems completed, for a student that appears to be learning.

Class Summary Per Problem

The Per Problem Report summarizes class-wide performance across students for each individual problem. The teacher's feedback on this report suggested that many of the columns present in the original report were redundant or not useful. The columns such as "Percentage of students who eventually solved the problem", "Average number of hints on the last encounter" and "Average comprehension time solving" were some of the columns that the feedback deemed as redundant. However, other columns such as "Percentage of students who solved it on the first try/second try", "Most frequent incorrect response" and "Percentage of students who skipped/gave up the problem" were found to be useful pieces of information from a Teacher's perspective.

Figure 4 shows the old teacher tools version of the report.

Problem Difficulty Report for Rio Schools - 6th-TownleyPeriod5

Problems that less than 4 students saw

Choose another report

#	#	Problem	Number of Students Who have seen it	Percentage of Students Who've repeated the problem	Percentage of Students who skipped	Percentage of Students who gave up	Percentage of Students who eventually solved it	Percentage of Students solved on first try	Percentage of Students solved on second try	Most frequent incorrect response	Avg Number of hints on last encounter	Avg comprehension time on solving encounter
0	1	OrderOfOps_2_oo2	6	16.7	0.0	33.3	66.7	50.0	0.0	-	0.0	43.9
0	0	problem_018: Trigonical/parameter #1	1	0.0	0.0	0.0	100.0	0.0	0.0	-	0.0	4.2
0	0	problem_013: RectParamFromSquares	2	0.0	50.0	0.0	50.0	0.0	0.0	-	0.0	6.5
0	1	problem_020: PolygonParameter L 44	5	20.0	20.0	20.0	60.0	40.0	0.0	-	0.0	32.4
1	1	problem_022: ShadedRectsWithinSquare	5	20.0	0.0	0.0	100.0	0.0	0.0	-	0.0	23.0
2	2	problem_023: AreaSquarePart	5	0.0	40.0	20.0	40.0	0.0	20.0	A	2.2	60.4
0	0	problem_043: RectParamFromSquares	1	0.0	0.0	0.0	100.0	0.0	100.0	D	0.0	4.7
0	0	problem_050: PolygonParameter	2	50.0	0.0	50.0	50.0	0.0	0.0	-	0.0	73.5
0	0	problem_053: AreaSquarePart	3	0.0	33.3	33.3	33.3	33.3	0.0	-	0.0	9.2
0	2	problem_079: Parallelogram Parameter	5	0.0	0.0	40.0	60.0	20.0	20.0	C	0.2	7.5
3	3	problem_092: Cost renting office for 6 months	10	10.0	90.0	0.0	10.0	0.0	0.0	-	0.0	4.0
0	0	problem_107: perimeter	3	0.0	33.3	0.0	66.7	0.0	33.3	-	0.0	77.2
0	0	problem_112: Perimeter rectangles	3	0.0	0.0	0.0	100.0	66.7	0.0	-	0.0	24.2
0	0	problem_113: rectangle area	9	0.0	22.2	0.0	88.9	22.2	33.3	B	0.3	-65.1
0	0	problem_116: sufficiency	23	26.1	4.3	13.0	82.6	30.4	8.7	-	0.1	40.1

Figure 4. The Class Summary Per Problem Report in the Old Teacher Tools.

Class Summary per Common Core Report

The class report per common core summarizes the performance of a class per Common Core Standard. The feedback received regarding this report was generally positive but requesting further information, such as *“It is nice that the class summary per Common Core gives a summary for each cluster but as a teacher I would like to see it broken down further”*.

The teacher feedback also pointed towards the fact that the report would allow teachers to prioritize standards to work on, if shown appropriately. The teacher also commented *“I think MathSpring would be very beneficial for my students to practice and assess their knowledge on each objective within the standard”*.

Figure 5 shows the old teacher tools version of the report.

Detail Problem Info for Rio Schools - 6th TownsleyPeriod5

■ Clusters that your students found really hard
■ Clusters that your students found challenging

Choose another report

Cluster name	# problems	% solved in the first attempt	Avg # of hints requested per problem	Seconds "thinking" out the problem – time to first action (hint or attempt)	Incorrect attempts per Problem
Grade-6: Applying Multiplication and Division Concepts, Apply And Extend Previous Understandings Of Multiplication And Division To Divide Fractions By Fractions	2	0.0	0.0	16.1	3.5
Grade-6: Equations and Inequalities, Reason About And Solve One-Variable Equations And Inequalities	13	15.3	0.1	33.0	1.2
Grade-6: Area, Perimeter, Area, and Volume, Solve Real-World And Mathematical Problems Involving Area, Surface Area, And Volume	6	16.6	1.6	23.6	0.6
Grade-4: Decimals and Fractions, Understand Decimal Notation For Fractions, And Compare Decimal Fractions	24	16.6	0.0	7.6	1.2
Grade-5: Multi-Digit Whole Numbers and Decimals, Perform Operations With Multi-Digit Whole Numbers And With Decimals To Hundredths	10	20.0	0.3	50.3	1.3
Grade-4: Lines and Angles, Draw And Identify Lines And Angles, And Classify Shapes By Properties Of Their Lines And Angles	5	20.0	0.0	15.8	2.4
Grade-3: Problems with Measurements, Solve Problems Involving Measurement And Estimation Of Intervals Of Time, Liquid Volumes, And Masses Of Objects	11	27.2	0.0	22.6	0.9
Grade-5: Place Value System, Understand The Place Value System	16	31.2	0.0	28.1	1.0
Grade-4: Building Fractions, Build Fractions From Unit Fractions By Applying And Extending Previous Understandings Of Operations On Whole Numbers	193	33.1	0.2	36.3	0.7
Grade-6: Ratio, Understand Ratio Concepts And Use Ratio Reasoning To Solve Problems	40	35.0	0.1	14.7	0.6
Grade-5: Multiplication and Division of Fractions, Apply And Extend Previous Understandings Of Multiplication And Division To Multiply And Divide Fractions	323	35.6	0.3	41.7	0.7
Grade-6: Algebraic Expressions, Apply And Extend Previous Understandings Of Arithmetic To Algebraic Expressions	59	35.8	0.4	55.8	0.6
Grade-3: Fractions as Numbers, Develop Understanding Of Fractions As Numbers	179	36.8	0.2	24.1	0.8
Grade-5: Add and Subtract Fractions, Use Equivalent Fractions As A Strategy To Add And Subtract Fractions	209	39.7	0.3	41.0	0.6
Grade-5: Unit Conversion, Convert Like Measurement Units Within A Given Measurement System	112	41.0	0.4	45.1	0.5
Grade-5: Numerical Expressions, Write And Interpret Numerical Expressions	60	45.0	0.2	56.2	0.4
Grade-4: Fraction Equivalence and Ordering, Extend Understanding Of Fraction Equivalence And Ordering	69	46.3	0.2	29.5	0.5

Figure 5. The Class Summary Per Common Core Report, highlighting standards that a class appears to be weaker at.

Based on the gathered feedback, the conclusion was that a new version of the teacher tools should be designed. In general, the teacher’s feedback suggested that:

1. Specific fields/columns in the reports were unnecessary and confusing to teachers.
2. The set of existing class reports needed to be redesigned and revamped from a teacher’s usability standpoint. Hence, based on that feedback collected we decided to:
 - Improve the general user interface such a way that it is easier to use and relatable to Teachers (Non-Researchers);
 - Improve the class reports based on the specific input and critique provided by the expert teacher to make it more understandable and usable to Teachers.

Screenshots for the initially designed prototype built upon these responses are shown as part of the next section, which comprises what we call the “second iteration”.

Second Iteration: Prototype I of the New Teacher Tools

The newly designed **Prototype I** of new Teacher Tools was designed to be more user friendly and intuitive to Teachers. This version was designed with the notion that teachers could configure a new class and control other aspects such as Manage Students and their user information, clone an existing class etc. with much better ease. **Prototype I** also included several design and implementation level changes to all the available reports based on data received during iteration one.

During the second iteration we also evaluated Prototype 1 of new Teacher Tools based on quantitative and qualitative responses collected from N=10 teachers on this release via a detailed closed and open question survey. Teachers were part of a 3-day MathSpring Professional Development workshop during the summer of 2018. A quantitative data analysis of the collected survey responses was carried out, to understand understandability and usefulness, as well as weaknesses and strengths of the new version of Teacher Tools, in order to identify further areas for improvement.

Figure 6 shows the first page that teachers encounter after login in **Prototype I**. This page was revamped to a dashboard-like design with an improved look and feel as well as a rather self-explanatory layout. Teachers are now allowed to immediately see all of their classes after login, giving the possibility to choose a specific one for further inspection. At the same time, the menu on the left allows the creation of new classes or surveys, which are independent from one of the existing classes. It also allows to create a new class by “cloning” the configurations of one of the old classes.

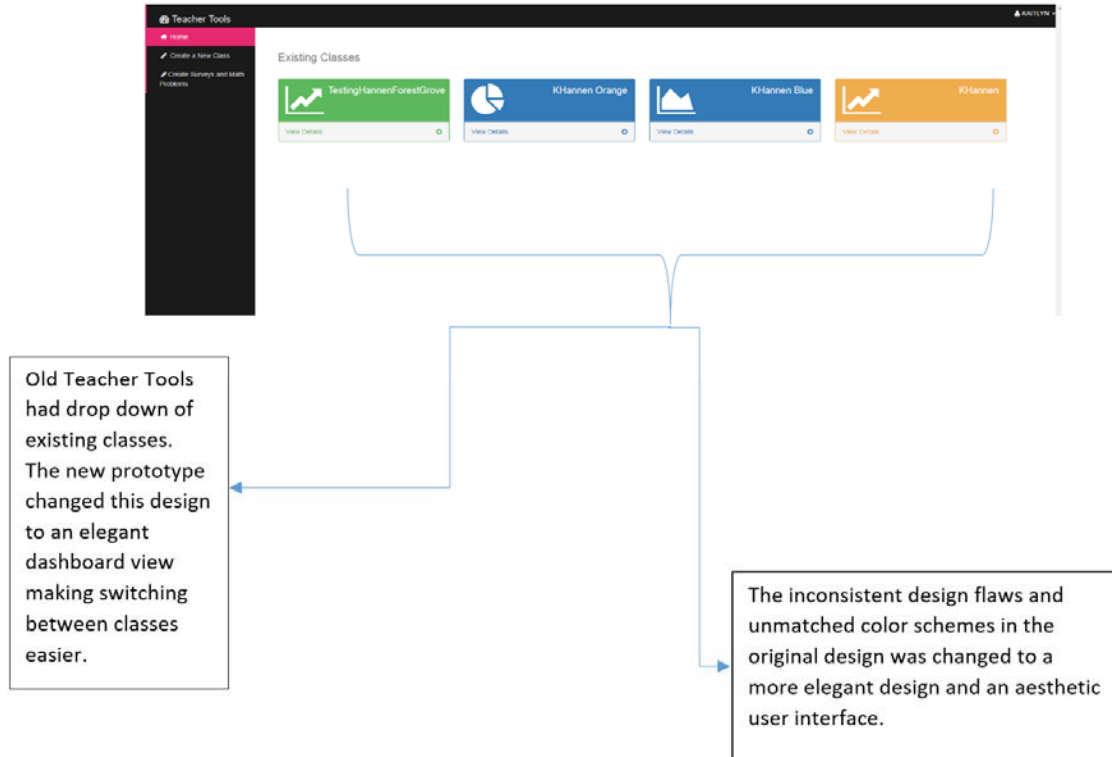


Figure 6. The Teacher Dashboard in the New Teacher Tools (Prototype I).

Teacher Response to Prototype I

As we analyzed the data from the iteration one which called for a serious redesigning of Teacher Tools as far as parameters such as usability and understandability of class reports from teachers' standpoint, it was decided to make the class reports more accessible and readable to teachers at the end of first iteration.

As mentioned earlier, a quantitative data analysis of the collected survey responses during this iteration was carried out. As part of this we asked the participants (teachers) to rate the usefulness of each of the reports while also drawing on any comments they have about the same.

Class Summary Per Student

The report that used to be called “Detail per Student” report now changed to be called Class Summary Per Student. This report shows a summary the problems seen by each student over the course of all student logins.

Figure 7 shows the new Look-and-feel of this report, which now includes charts of detailed statistics of different forms of engagement of each student (e.g. what percent of problems the student quick-guessed, or solved on first attempt, or gave up on, or solved with hints).

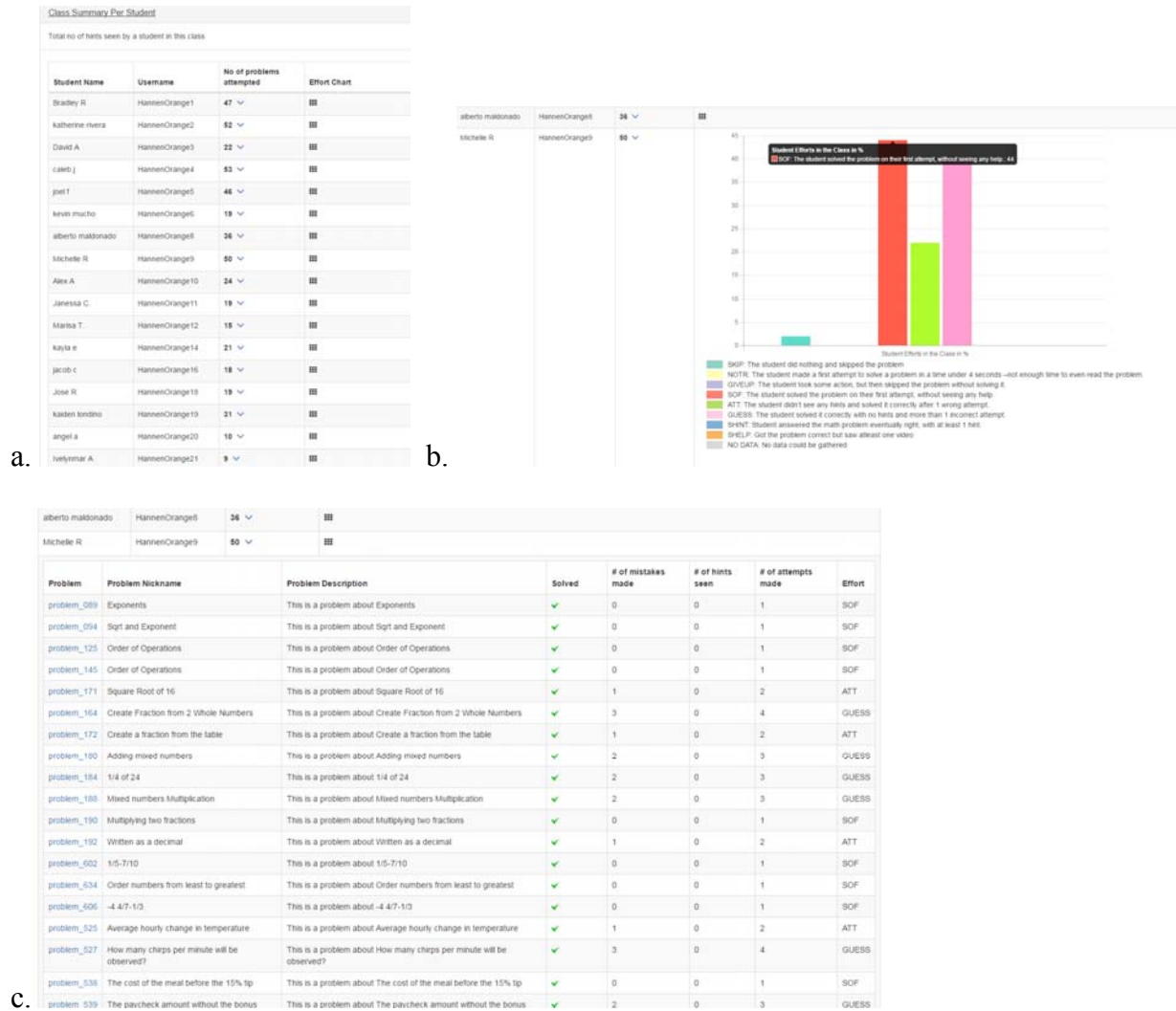


Figure 7a. Teacher comments regarding information displayed in the Class Summary per Student in Prototype I of the New Teacher Tools.

When we evaluated the quantitative responses from the teachers regarding the usefulness of this report, we found that **70%** of the participants found the report to be very useful, **30%** of the participants found this report to be somewhat useful, and nobody found it not useful. Unfortunately, no comments were included to suggest how it might be improved further.

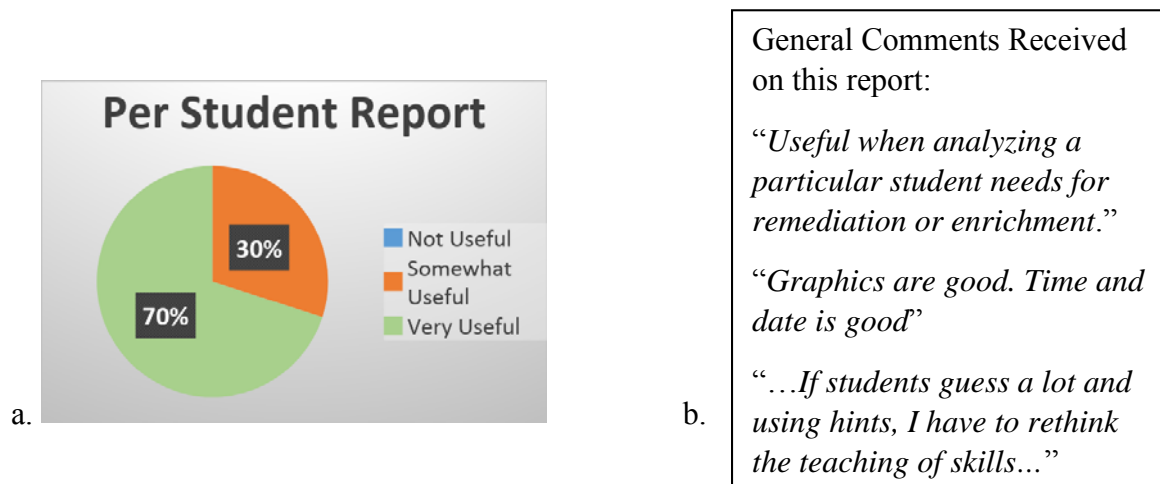


Figure 7b. Teacher perceptions of the Class Summary per Student Report in Prototype I of the New Teacher Tools; a. Perceived Usefulness; b. Sample qualitative comments.

Class Summary Per Student per Problem Set

The Class Summary Per Student Per Topic report was renamed to Class Summary Per Student Per Problem Set, as “topics” are actually bags of problems that span a few different standards and across grades with a common theme that teachers and students easily relate to (e.g. “Triangles Problems”), so the term “Problem Set” seemed like a more intuitive name to talk about them. The Class Summary Per Student Per Problem Set Report showed how students were doing on each problem set, along with the “mastery” acquired on each of these. As suggested by teachers, more details on how “mastery” of a particular problem set gets calculated was included, along with more elaborate information on mastery development of each student.

Class Summary Per Student Per Problemset

This table shows problem set-wise performance of students of this class.

Mastery Range	Grade/Color Code
0.75 or Greater	Grade A (Excellent)
Between 0.5 and 0.75	Grade B (Good)
Between 0.25 and 0.5	Grade C (Needs Improvement)
0.25 or Less	Grade D (Unsatisfactory)

Showing 1 to 4 of 4 entries

- Each cell shows [number solved on first attempt / number problems solved] along with highest recorded "Mastery" 0.
- Cell wherein students have attempted 10 or more problems are color coded.
- Click on the cell to get the complete "Mastery Trajectory" for given student and problemset

What is Mastery?

- Student "Mastery" is MathSpring's estimation of how much the student has demonstrated to know the topic, ranging from zero (0) to one (1)
- Example: If a student solves 4 problems correctly in a row without mistakes and no help requests, they get a mastery of 0.85 and mastered the topic.
(Note: also, these 4 problems would be increasingly harder, because as students continue to solve these problems correctly, problems get harder). Mastery would increase slower if the student asks for hints. Mastery would decrease if students make mistakes

Student Name	Username	Fractions Review	Numerical Expressions and Operations	Decimals and	Rates Ratios and Proport
randi gil	RioPeriod5-1	[5/11]0.92	[3/13]0.21	[0/4]0.10	[3/5]0.59
Reynaldo P	RioPeriod5-2		[3/9]0.44		[5/7]0.94
Rubi M	RioPeriod5-3		[5/5]0.84		[2/4]0.24
Axel F	RioPeriod5-4	[0/0]0.00	[4/0]0.38		[5/0]0.84
Mia G	RioPeriod5-5		[2/7]0.21		[5/7]0.36

a.

Mastery Trajectory Report

Problem ID	Problem Name	Student Effort
453	problem_184	ATT
453	problem_184	SQE
1200	problem_366	NO DATA

The student SOLVED the problem correctly on the FIRST attempt, without any help.

b. c.

Figure 8. Teacher comments regarding information displayed in the Class Summary per Student per Problem Set in Prototype I of the New Teacher Tools.

When we evaluated the response from the teacher's regarding the usefulness of this report, we found that **30%** of the participants found this report to be *somewhat useful* and **50%** of the participants found the report to be *very useful* while **20%** found it be *not useful*.

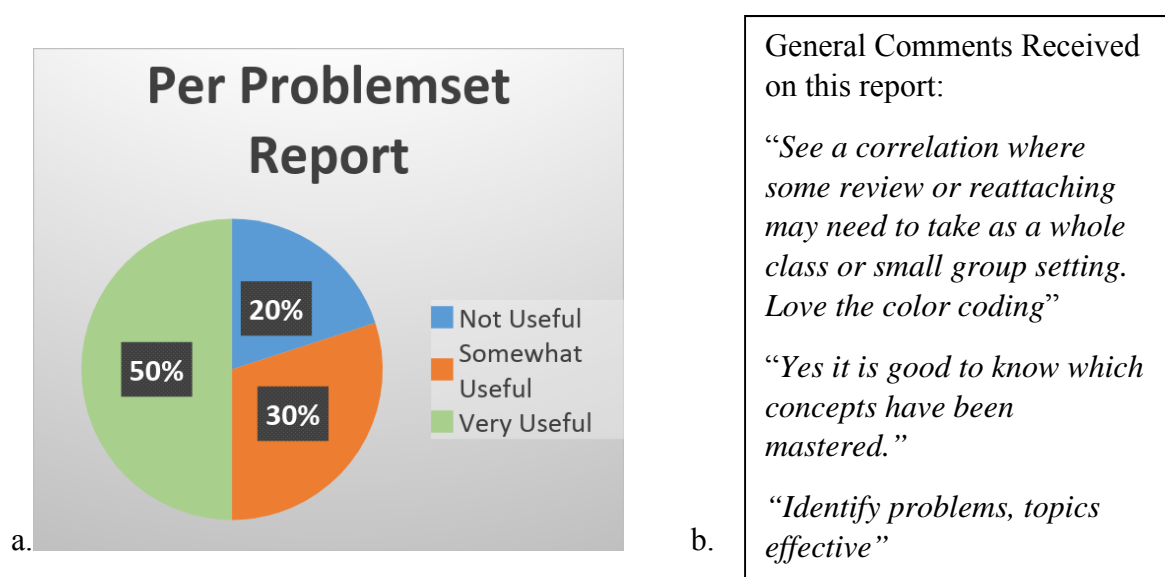


Figure 9. Teacher comments regarding information displayed in the Class Summary per Student per Problem Set in Prototype I of the New Teacher Tools.

Class Summary Per Problem

The Per Problem report summarized class-wide performance of the all the students in a single report, where each row corresponded to an individual problem, regardless of which problem set it belonged to (and note that some problems may show up in several problem sets, as a problem regarding sum of the internal angles of a triangle may be part of the *Triangles* problem set, the *Triangles and Angles Review* problem set, and the *Equations (solve for x)* problem set, depending on the characteristics of the problem. The Expert Teacher’s feedback during the First Iteration suggested removal of certain columns, which seemed to be more useful for researchers than teachers. That feedback also suggested for association of each problem to particular standards, thus standards information for a problem was included. The new version of this report, also retained some of the columns of the original report such as *“Most frequent incorrect*

response”, which were believed to be excellent data points from a teacher’s perspective. The result of the modifications of this report is shown in Figure

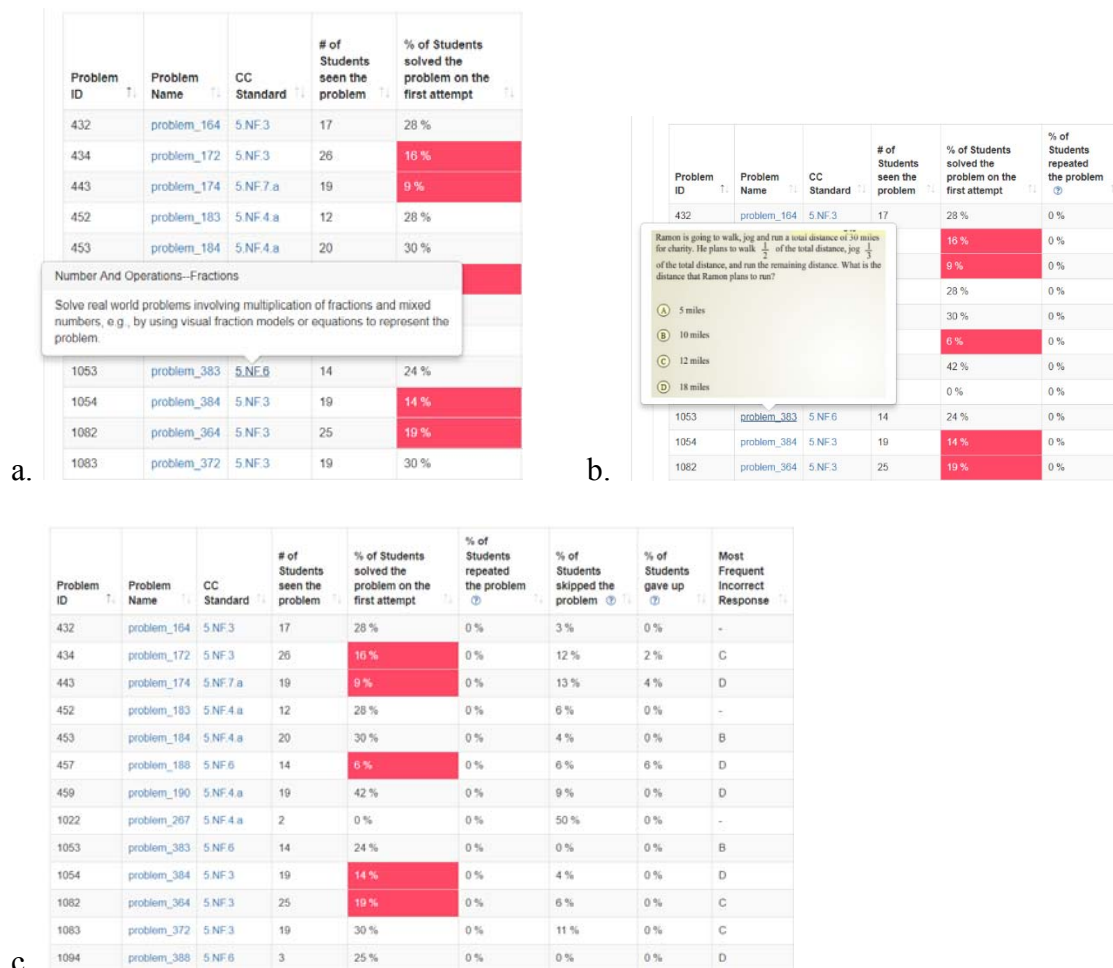


Figure 10. Teacher comments regarding information displayed in the Class Summary per Student per Problem in Prototype I of the New Teacher Tools.

When we evaluated the response by the 20 teachers regarding the usefulness of this report, we found that **10%** of the participants found this report to be somewhat useful and **60%** of the participants found the report to be very useful, while at the same time, **30%** found it be not useful. Unfortunately, people who said the report was *not* useful did not provide reasons why.

Figure NEWTTPERPROBLEMDATA shows these statistics as well as sample qualitative comments received.

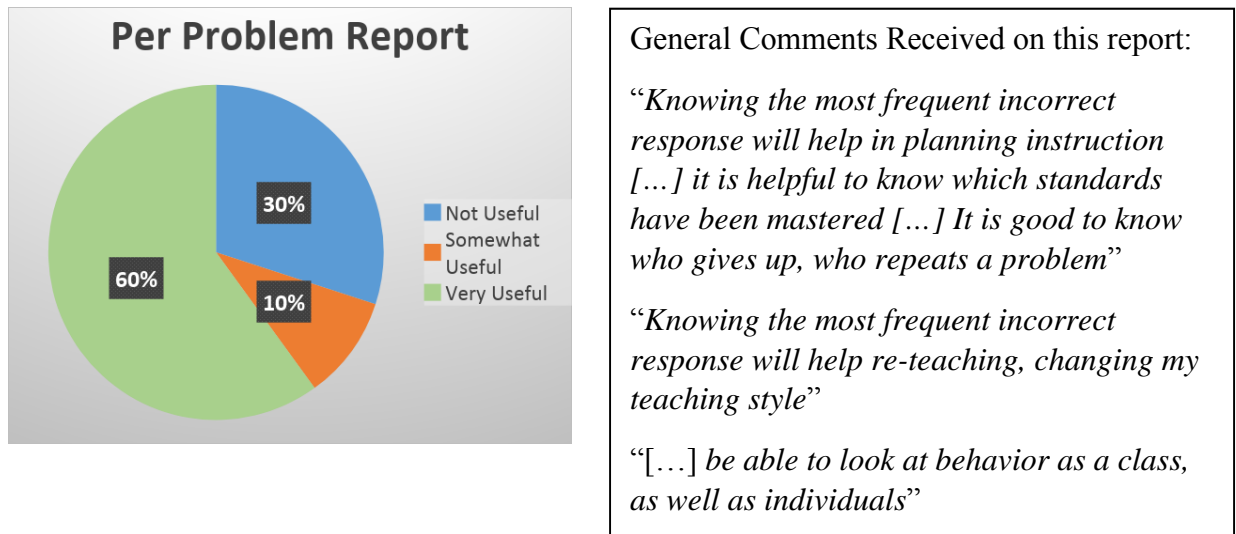


Figure 11. Teacher comments regarding information displayed in the Class Summary per Student per Problem in Prototype I of the New Teacher Tools.

Class Summary per Common Core

The class report per common core summarizes the performance of a class per standard. Since the feedback received on this report was generally positive, we did not make changes to this report. However, we did redesign this report to incorporate a more relatable color coding within each standard that would help teachers identify and assign priority within these standards.

When we evaluated the response from the 20 teacher’s in the summer 2017 regarding the usefulness of this report, we found that 20% of the participants found this report to be somewhat useful and 60% of the participants found the report to be very useful while 20% found it be not useful, as shown in Figure NEWTTPERCOMMONCOREDATA.

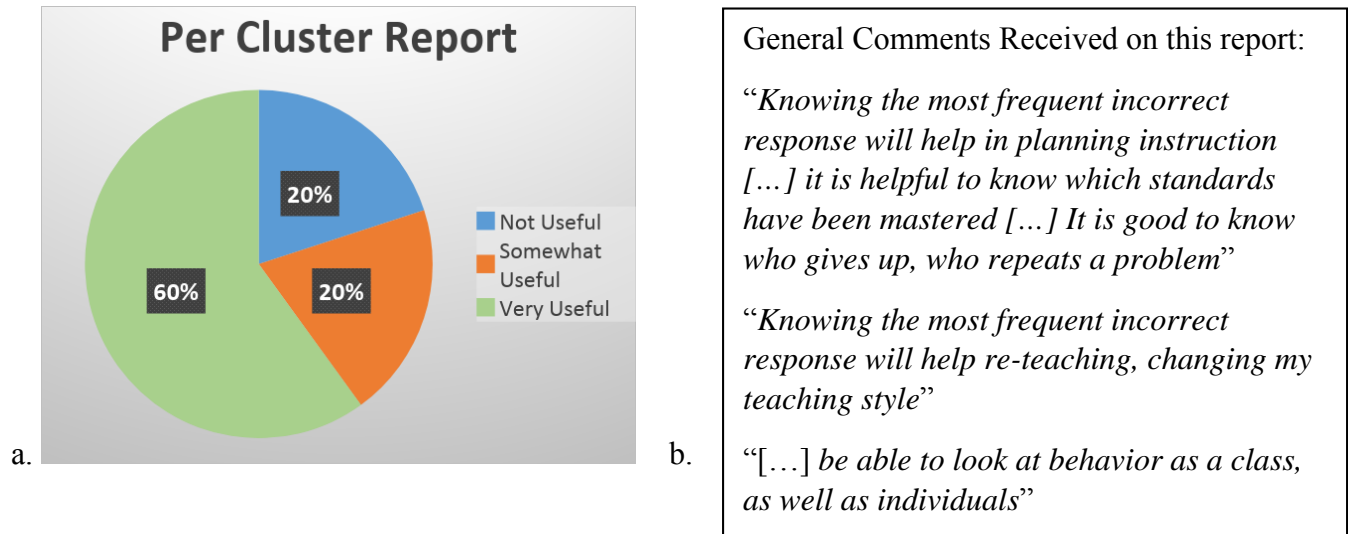


Figure 12. Teacher usefulness ratings regarding the Class Summary per Student per Problem in Prototype I of the New Teacher Tools.

In conclusion, we found out that Prototype I was well received by the majority of teachers and found very useful most of the time. However, a reasonable percent of teachers found the reports not useful, and it was unclear what the reason was for their answers.

As a consequence, we decided to pull a small expert panel of a senior teacher, an HCI expert, and the rest of the MathSpring development team to decide on changes that would improve each of the reports further. The decisions made are explained in the next section.

Third / Final Iteration: Prototype II of the New Teacher Tools

The improved Prototype I design - Prototype II, of new Teacher Tools was designed based on changes, features suggested by a panel consisting of a senior teacher (Participant of Iteration one and Iteration Two), an HCI expert, and the rest of the MathSpring development team. This new Prototype II essentially implied a redesign of the class reports based on strategic inputs received from the panel regarding Prototype I.

The third iteration concluded with an evaluation of Prototype II by N=10 teachers who were part of a MathSpring workshop this Spring 2018, being able to collect quantitative and qualitative responses via a detailed closed and open question survey, similar to the previous teacher feedback collected. Teachers were part of an Action Research Training course during the spring of 2018, and their inputs were quite detailed. A quantitative data analysis of the perceived usefulness of the reports was carried out, to understand understandability and usefulness, as well as weaknesses and strengths of the improved new version of Teacher Tools, in order to identify future areas for improvement.

As mentioned earlier, Prototype II of new Teacher Tools was designed based on changes, features suggested by a panel consisting of a senior teacher (Participant of Iteration one and Iteration Two), an HCI expert, and the rest of the MathSpring development team. The changes incorporated into the class reports are specified in detail in each of the remaining sections.

Class Summary Per Student

For this report, the expert teacher and other content experts on the panel initially suggested to display students affect data that was already being tracked by MathSpring, to portray a better picture of the student. Based on the aforesaid input, a new attribute called “*Emotion Chart*” was created as part of this report. This chart tried to provide an insight into students’ affect based on survey responses given by students during their MathSpring Session. These emotions self-reported by students were defined as “*Frustration*”, “*Excitement*”, “*Interest*”, and “*Confidence*”. The general comments / reasons stated by students for their affect (emotion) were also shown and are displayed in Figure 13a.

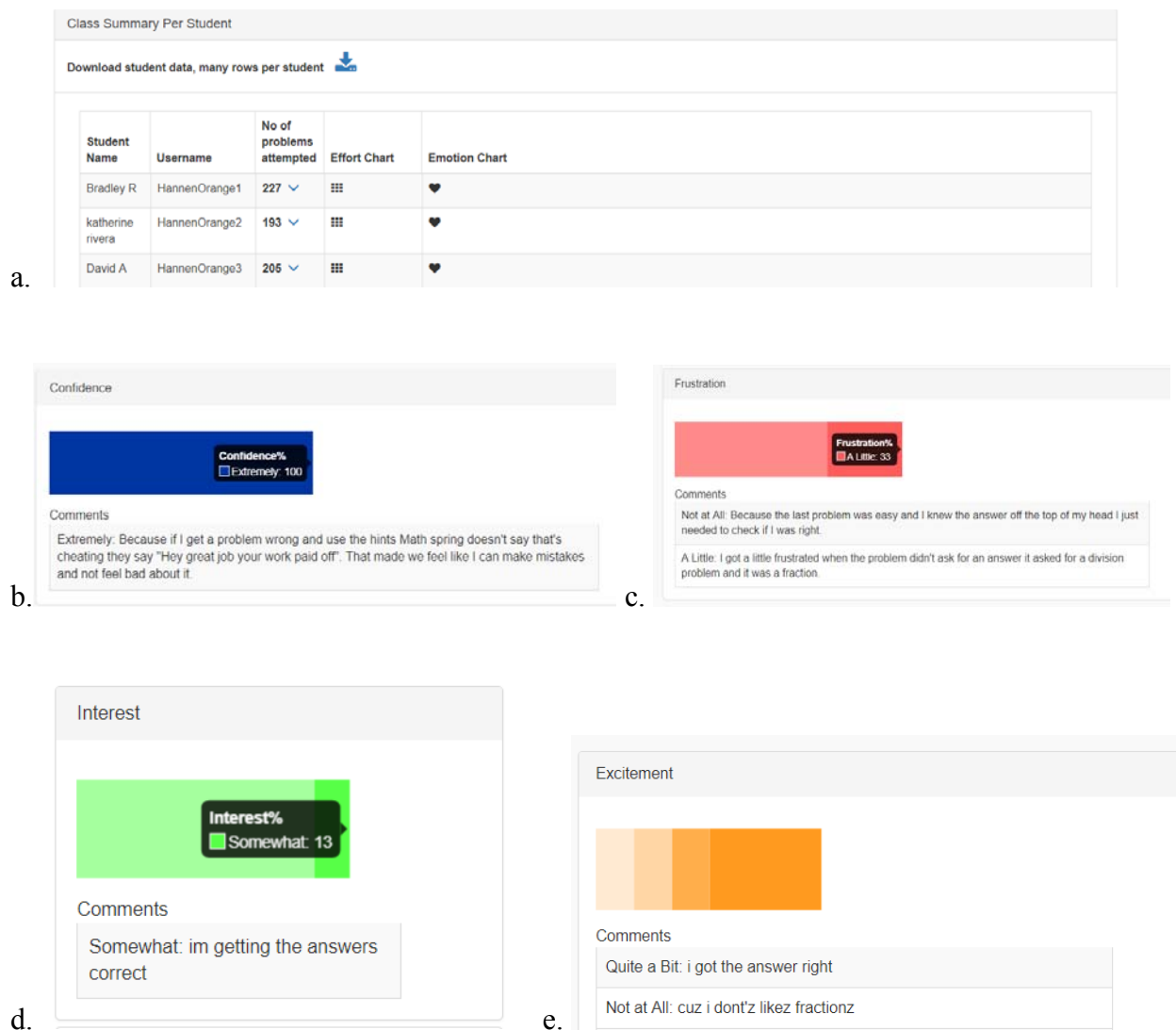


Figure 13a. Emotions in the Class Summary per Student Report in Prototype II of the New Teacher Tools: a. Initial View when opening this report, including engagement and emotion chart icons; b. Detailed charts of students' emotions, and their reported reasons they attribute to those emotions.

Secondly, the math experts on the panel asked to improve the navigation of the detailed logs of students' activity, organizing these logs by some criteria that would make it easier to parse. After discussions, we decided to group the number of problems attempted per student based on timing. These categories included "Problems Solved Today", "Problems Solved Most Recently" (i.e. the last day used) and "All Solved Problems". This would allow teachers to

immediately get a snapshot of what students had just finished doing after students finished working with MathSpring (or while students were working with MathSpring), as opposed to seeing the full log of data. These changes are shown in Figure 13b.

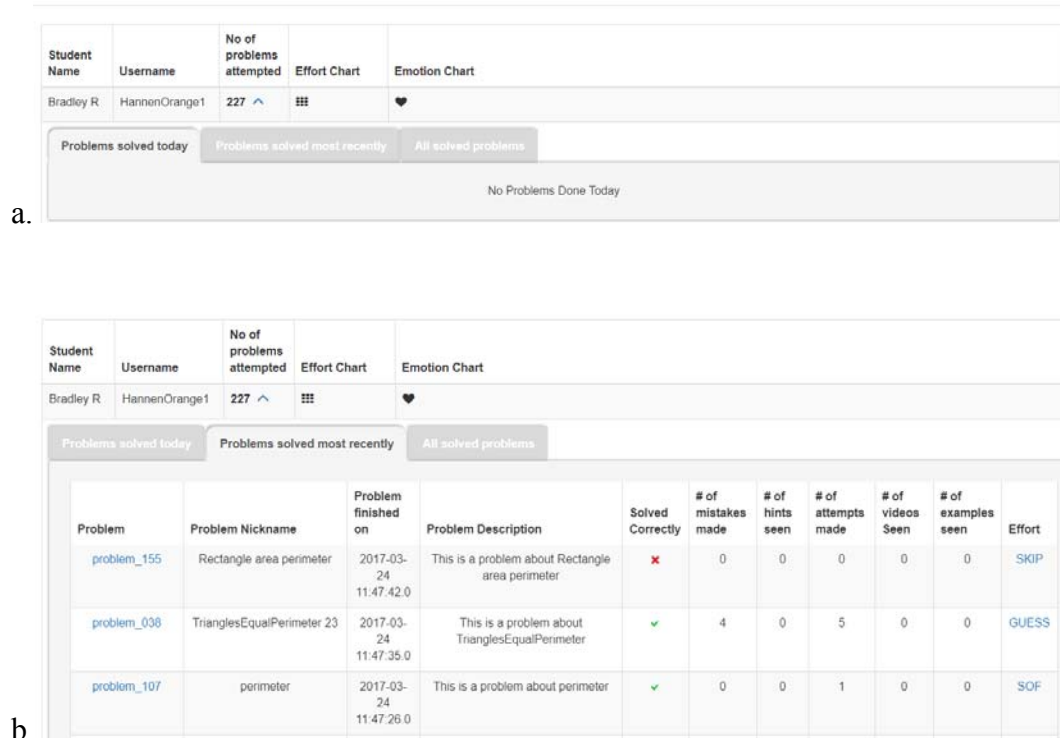


Figure 13b. Student History in the Class Summary per Student Report in Prototype II of the New Teacher Tools: a. Clicking on arrow reveals allow different time slices of history per student; b. Clicking on Problems solved most recently reveals the student history for the last day MathSpring was used by individual students.

Finally, the HCI expert on the panel gave elaborate feedback on how the effort chart should be modified from and readability perspective, and his full report is included in Appendix B. He commented that “...*The effort chart, while useful, is presented in a very inefficient way; there are nine effort categories, and each is shown as a percentage in a bar graph, with the scale showing all the way up to 100%. This wastes most of the space, as typically each bar is less than*

50%...Finally, once expanded, the only way to collapse it is to expand another part of the table; as it takes up so much space, one must often scroll down just to be able to expand the next student's effort chart. I believe this could be much more compactly and effectively conveyed with a single horizontal stacked bar,... as Stacked Bar Chart". This idea was adopted into the new design concept and the original bar chart for each of the individual effort was modified to a more compact and sign elegant stacked chart layout. Figure 13c shows the resulting chart.

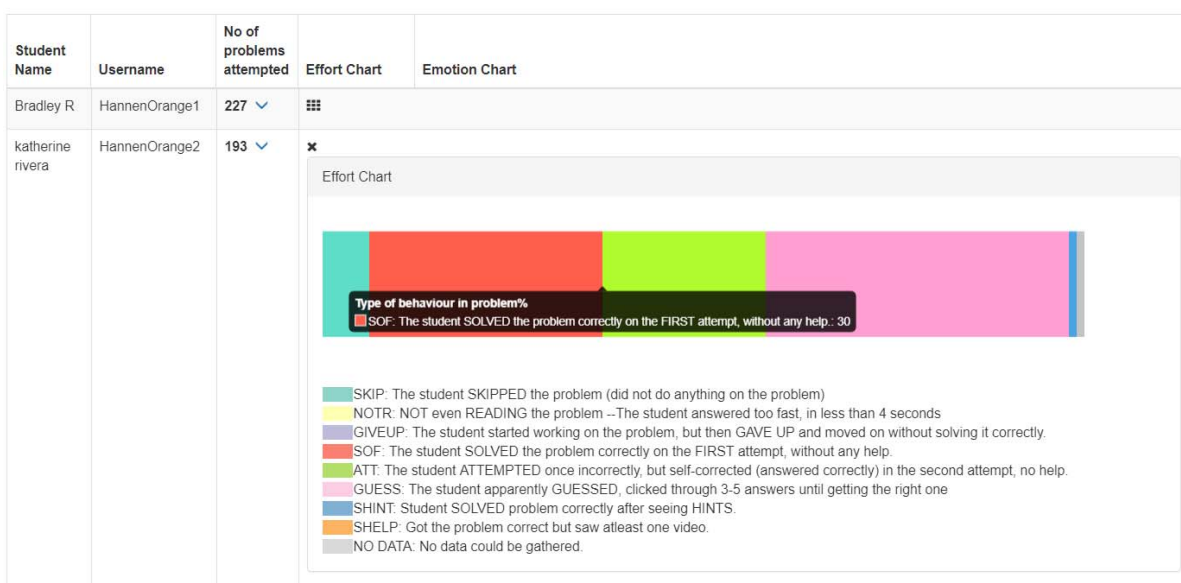


Figure 13c. Student Engagement in the Class Summary per Student Report in Prototype II of the New Teacher Tools. Clicking on the “Effort Chart” icon reveals a stacked bar chart that shows percent of specific behaviors on problems.

When we later had teachers provide ratings of the usefulness of this report, we found that **57%** of the participants thought the report was very useful, **34%** of the participants found this report to be somewhat useful, and **9%** found it not useful. There was sharp decline in the number of people who found this report very useful during second iteration from **70% to 57%**. While we

did not expect this to happen, something important to note is that the participants of second iteration were different from participants of third iteration, and that were half the amount. Moreover, many teachers' marked this report as somewhat useful commenting "*I am only saying that because I see potential in having this data but navigation and clarity are lacking in this. I personally wouldn't use this in my class but I do see how someone could make it useful*". More details of teacher comments are shown in Figure 14a.

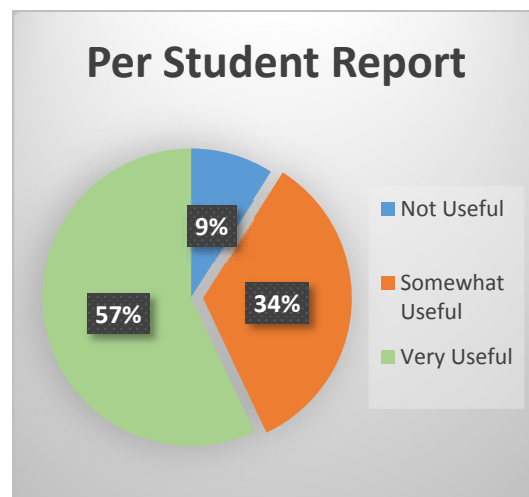


Figure 14a. Teachers' Perception of the Class Summary per Student Report in Prototype II of the New Teacher Tools.

General Comments Received on this report:

“I like the effort chart. I don’t feel that I know enough about the questions for it to be truly helpful but I think it will. I wonder about the conversation before students use the emotion chart”

“The various ways that I can look at how the student is interacting with the program are very useful. I can easily pull up the question types that the student answered and look at the ways in which the students answered the questions (skipping, guessing, first try, self-corrected, etc...). This provides a great basis for a conference with a student like Andres (one of the two I looked at) to ask him why he answered the way he did to the emotional questions and yet guessed or skipped half of the questions.”

“I find the effort chart to be the most interesting because it gives me quick insight into how the student is doing with the program. I feel like I could use this information for conferencing with students. For example, with Ruby I could say, “I see that you are frequently using hints to help you and you are getting many questions right with the support of hints. This is working for you so keep it up.”

“I think it’s great to be able to see more than just right/wrong. The report gathers information about emotion as well as more about how the student answered the question. Right and wrong doesn’t always give us a good picture of a student’s understanding.”

Figure 14b. Teachers’ Comments about the Class Summary per Student Report in Prototype II of the New Teacher Tools.

The panel did not suggest for any design level changes for this report. Hence we subjected the new panel of teachers to same version of this report for the third iteration and gathered responses for the same.

When we evaluated the quantitative responses from the teachers regarding the usefulness of this report, we found that 66% of the participants found the report to be very useful, 34% of

the participants found this report to be somewhat useful, and no one found it not useful. Here, there was sharp increase in the number of people who found this report very useful during second iteration from 50% to 66%. The comments we received for this received were generally positive with a few suggestions for future work. Figure 15a shows these results in graphical form, and Figure 15b highlights very positive qualitative comments about this report.

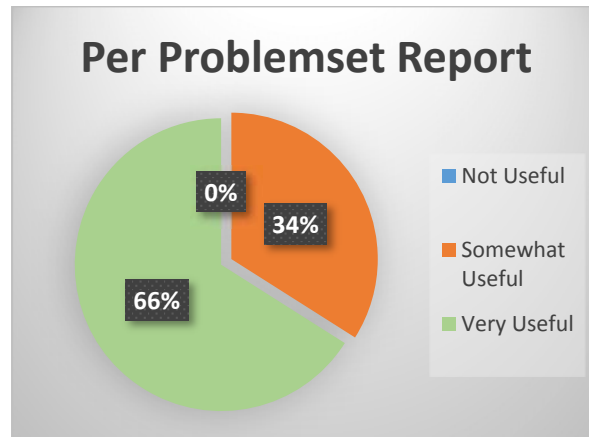


Figure 15a. Teacher usefulness ratings regarding the Class Summary per Problem Set Report in Prototype II of the New Teacher Tools.

General Comments Received on this report:

“I love the way the questions are mapped and quantified so that student progress can be tracked over time and progress (or lack thereof) can be clearly seen. This is great data to be able to share with students during conferencing as well.”

“I was able to draw a lot of conclusions quickly from the data. It would be even better if I could see the effort chart when I click on the mastery trajectory report. I found myself going through the list of problems manually to get an idea of the student effort from this” view”.

“I think the data will allow teachers to see their students’ true colors on how they work on certain questions and concepts. This really does a good job at breaking the data down into smaller images of the students’ performances. I think this could help diagnose what students need work on to boost their confidence levels before MCAS.”

“All teachers are comfortable with the standards. The standards are very similar through grade levels. I like the colors for the grades. It is very easy to see. The mastery trajectory report is interesting. I like looking at this information.”

Figure 15b. Teacher comments regarding the Class Summary per Problem Set Report in Prototype II of the New Teacher Tools.

Class Summary Per Problem

The experts of the review panel were of the opinion that the information on this report was redundant, and already shown as part of the “*Class Summary Per Common Core*” report. Based on the aforesaid observations we decided to remove this report altogether. The screenshots of the changes incorporated in this report are hence shown in the next section.

Class Summary per Common Core

The panel experts did not make any suggestions for changes on this report. However, since the “*Class Summary Per Problem*” was removed and was to be merged with this one (so

that information would not be lost in the process), we highlight the principal changes done to make this single report more complete and useful.

The two main suggestions given by the panel regarding this report regarded the inclusion of two additional attributes/columns. These attributes were “*Similar Problems*” and “*Collective effort on Problem*”. “*Similar Problems*” is a way to provide further material for practice in an easy way; it provides a web link to problems associated with the current standard in the **Massachusetts Comprehensive Assessment System Web Portal** which is essentially a website maintained and updated by the Massachusetts State Government dedicated towards MCAS testing. A specific feature allows searching for MCAS problems regarding a specific standard. This allowed providing teachers with a myriad of problems that were somewhat similar to the problem at hand.

6.RP (Grade-6: Ratio) 5 46 1

Problem ID	Problem Name	CC Standard	# of Students seen the problem	% of Students solved the problem on the first attempt	% of Students repeated the problem	% of Students skipped the problem	% of Students gave up	Most Frequent Incorrect Response	See Similar Problems	Collective effort on Problem
1059	problem_462	6.RP.3.b	8	18 %	0 %	18 %	9 %	C	?	⋮
1096	problem_394	6.RP.3.a	2	100 %	0 %	0 %	0 %	C	?	⋮
1236	problem_370	6.RP.1	4	60 %	0 %	0 %	20 %	C	?	⋮
1239	problem_393	6.RP.3.c	4	40 %	0 %	40 %	0 %	D	?	⋮
1318	problem_800	6.RP.3.c	4	14 %	0 %	0 %	0 %	-	?	⋮

a.

Problem ID	Problem Name	CC Standard	# of Students seen the problem	% of Students solved the problem on the first attempt	% of Students repeated the problem	% of Students skipped the problem	% of Students gave up	Most Frequent Incorrect Response	See Similar Problems	Collective effort on Problem
1236	problem_370	6.RP.1	4	60 %	0 %	0 %	20 %	C	?	⋮
1239	problem_393	6.RP.3.c	4	40 %	0 %	40 %	0 %	D	?	⋮
1318	problem_800	6.RP.3.c	4	14 %	0 %	0 %	0 %	-	?	⋮

Expressions And Equations

Understand the concept of a ratio and use ratio language to describe a ratio relationship between two quantities. For example, “The ratio of wings to beaks in the bird house at the zoo was 2

b.

Elementary & Secondary Education

Students & Families Educators & Administrators Teaching, Learning & Testing Data & Accountability Finance & Funding About the

Student Assessment Home
 Next-Gen MCAS
 Graduation Requirements
 MCAS
 Student Participation
 Statewide Testing Schedule
 Text Administration Resources
 Accessibility and Accommodations
 Test Design and Development
 2018 Assessed Standards/Transition
 Test Questions and Practice Tests
 Student Work/Scoring Guides
 Technical Reports
 Results
 MCAS Alternate Assessment
 Performance Appeals
 ACCESS for ELLs
 National/International Tests
 Growth Model
 Parent/Guardian Information
 Training Opportunities

Massachusetts Comprehensive Assessment System

MCAS Question Search

[Show Instructions](#) | [Send Feedback](#)

Basics

Year: Grade: Question Type: Question Set:

Massachusetts Curriculum Frameworks

Framework:
 Domain:
 Topic:

Extras

Keyword:

Allow Calculator: yes | no | all Answers: all | off Unreleased Items: yes | no [Get Results](#)

Showing 1 to 10 of 16
[Previous](#) | [Next](#)

2017 Spring Release, Mathematics - Grade 6
Question 4: Selected-Response [Show Questions](#)

Reporting Category: Ratios and Proportional Relationships
Standard: CCSS.Math.Content.6.RP.A.1 - Understand the concept of a ratio and use ratio language to describe a ratio relationship between two quantities. For example, "The ratio of wings to beaks in the bird house at the zoo was 2:1, because for every 2 wings there was 1 beak." "For every vote candidate A received, candidate C received nearly three votes."
Item Description: Determine which ratio describes a given ratio relationship in a real-world context.

Mr. Johnson's class has 24 students.

- There are 18 students in the class who study Spanish.
- The other students in the class study French.

In Mr. Johnson's class, which of the following is the ratio of students who study French to students who study Spanish?

A. 1:3
 B. 1:4
 C. 2:3
 D. 3:4

c.

Figure 16a. The Per Common Core Report in Prototype II of the New Teacher Tools. a. Common Core Clusters may be expanded via arrows, and cells that indicate poor performance on a problem are highlighted; b. Descriptions of the standards of a problem come up on roll-over; c. The web page that comes up from MASS DOE when clicking on the question mark for “see similar problems”.

Meanwhile, “Collective effort on Problem” as the name suggests depicts the collective effort of all students in the class towards a particular problem, in a visual manner. The resulting report is highlighted in Figure 16b.

Collective Students Effort on This Problem



Figure 16b. Aggregate Effort Chart in the Per Common Core Report in Prototype II of the New Teacher Tools.

When we evaluated the quantitative responses from the teachers regarding the usefulness of this report, we found that **66%** of the participants found the report to be very useful, **34%** of the participants found this report to be somewhat useful, and nobody found the report to not be useful. Here, there was an increase in the percent of people who found this report very useful during second iteration, from **60% to 66%**, as well as an increase from **20% to 34%** in the percent of people who found the report somewhat useful. The reduction from **20% to 0%** in non-usefulness is especially notable.

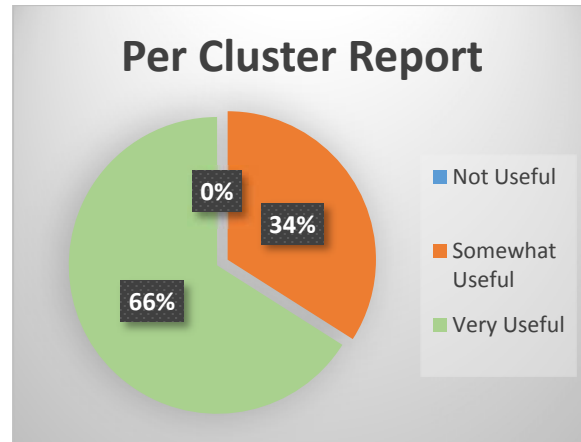


Figure 17a. Teacher usefulness ratings regarding the Class Summary per Student per Problem in Prototype II of the New Teacher Tools.

General Comments received on this report:

“would use this as a summarizer at the end of a mathspring session to show what we are good at and what we need to work on until we get those numbers to rise. This is also great for starting a whole group conversation.”

“I could really see this section helping me to decide what I might do at a class level to help a group with a skill set. I would be even better if I could see at the top level how many questions had been attempted for that skill. For example, the red rows caught by eye but because not many students actually attempted these skills, I’m not actually that interested in them right now. I wish I could see that at the top level before clicking on that cluster.”

“By clustering problems, it allows access into a child’s strengths and weaknesses and provides concrete feedback as to what domains needs to be revisited (or not).”

“The red and orange colors are fantastic. It is nice that you can see the questions, If there was more student data, I think one would see many more patterns and area where teachers would know they have to reteach something. I like that they had the most frequent incorrect response. I thought this was very helpful.”

Figure 17b. Teacher comments regarding the Class Summary per Student per Problem in Prototype II of the New Teacher Tools.

General Comments received on class reports in general:

“I think the Effort and Emotion charts would be very helpful on the Class Summary per Student chart and the Summary per Cluster reports. In particular, the skipped questions and not answered questions. I think those would tie into the “I don’t get it!” problem. I wonder if you could chart the number of times a question was skipped earlier in usage as opposed to later usage.”

“The class summary per student per problem set is my favorite of the three to track student progress. Realistically, my main wondering was about retention of previous concepts and with this tool I would be able to see (a) that my prediction is correct - they have not retained much from previous units and (b) track their progress in improving their scores with the previous concepts. Using the graphs, I’d be able to see how students answered (guess, skip, etc.) and watch how they continued to answer as they worked through the CCSS cluster - ideally looking for improvement”

“The “per student per problem set” data report will be most helpful for me as its layout allows me to easily compare my experimental group (students who are assigned a specific concept based on past performance) vs. the control group (those who are free to select a problem set). This report also will allow me to see which students are following through on a problem set vs. those who are cherry-picking easy problems from each set in order to grow their gardens”

“The effort charts will help me the most to see how often students actually used the available resources. Right now though I cannot tell when students looked at examples. I also cannot tell if they used a resource but still got it wrong. This information would help me to better understand how effectively my students are using the hint, example, and video resources.”

“I think the data will allow teachers to see their students’ true colors on how they work on certain questions and concepts. This really does a good job at breaking the data down into smaller images of the students’ performances. I think this could help diagnose what students need work on to boost their confidence levels before MCAS.”

Figure 18. Teacher comments regarding the Class Reports Prototype II of the New Teacher Tools.

As part of this final iteration, we also asked for a feedback as to how useful they thought the class reports are in general. The responses revealed a good liking for this feature from a teacher's standpoint. Some of the comments on the report feature in general is elaborated in Figure 18.

These results show that Prototype II is in general an improvement over Prototype I and the old teacher tools. Teacher comments are positive in general, they see the use that they could give to these Teacher Tools to understand their students better and guide future action plans that would alter their teaching based on information about their students' behavior, performance and affect.

Discussion and Conclusion

The feedback data isolated from Appendix A (Expert Teacher's Feed Back Responses), clearly suggested for multiple design level changes with respect to math content organization, the interface, and complexity in understanding the existing class reports.

Prototype I was constructed based on these feed-back responses. Responses to the initial prototype was met with mixed to positive responses in general regarding the usability and interface aspects. However, the experts on our team believed there are areas of further improvement from the usability standpoint. We came up a set of design updates to the existing reports to boost the usability quotient further. These changes were then subjected to further scrutiny by a different set of 10 teachers as part of the third and final iteration.

Prototype II was created based on the 10 teachers' feedback from the second iteration as well as the feedback of an HCI panel and again, the Expert Teacher's feedback. Prototype II was well received by a new set of 10 teachers, who thought that Prototype II was useful to them. Teacher comments were positive in general, they see the use that they could give to these Teacher Tools to understand their students better and guide future action plans that would alter their teaching based on information about their students' behavior, performance and affect.

Prototype II can still use some improvement. In the future, the Per Student report should be looked into again, so that teachers are better satisfied with the information provided.

Appendix A

Class Repots: The data in the class reports I feel benefits those analyzing the usefulness of the system and not the teachers.

- It is my understanding that the **detail per student info report** is a summary of all the problems seen over the course of all student logins. I don't believe that as a teacher this report is very useful.

Problems Seen: Does that mean all the problems seen including the hints within a problem? Does that include problems that they have seen but skipped because they didn't know how to them? When I look at the report the number of problems seen seems high. Also it would be beneficial to be able to see the problems that the students saw.

of hints seen: I think this is useful, but it tells you the total of hints seen not the number of hints for each problem. Report the data this way is not that useful. I want to know if my student is struggling on a certain topic or a certain type of problem. I want to be able to track my students progress as they use the hints. Did it take them one hint to get it right, two, etc.?

Seconds Spent Reading a Hint: The number of seconds spent reading a hint can be misleading. There are too many variables to get an accurate depiction of what is happening. For instance, what is the "right" number of minutes a student should spend reading a hint. If a student takes too long to read a hint does that mean that they are distracted, confused, or trying to reason it out.

Seconds to Make a Guess: This is helpful because you want to know if the students are really trying or just guessing. It would be extremely beneficial to see the actual problem that is linked to the number of seconds, instead of the total seconds for all problems.

It would be beneficial to give a report that shows the exact problems that a student saw, how long it took then student to record his answer and if the student clicked on any hints

- **Class summary per student per topic** gives an indication of how the students are doing on each topic. I would like to know how the percent master is calculated. On the line graph it would be nice to click on each point and get a detailed report of how the students answered the problem. For instance, when you click on the point it would show that student asked for a hint after x seconds. Then it would say whether the student made any guesses before clicking on another hint. The number of seconds students spent reading a hint could also be included.

When I click on the topic, a line graph pops up that shows the number of problems seen and the average process mastery. I am not sure what this graph is telling me.

- **Class summary per problem report**

Percentage of students who repeated the problem. If a student gets the problem right on the first try is there a chance that he will repeat the problem? Or are problems repeated only if a skips the problem or gets it wrong? If I look at the report I would think that the percentage of students who repeated a problem would be higher for the students that skipped a problem or gave up, but this is not the case.

Percentage of students who skipped the problem is a great data point It is easy to understand and tells me as a teacher that students didn't feel confident even attempting that type of problem. This is information that would guide classroom instruction. **The percentage of students who gave up** also is also an informative piece of data. Does giving up mean that they hinted out? Or that they started to solve the problem but then skipped it?

Percentage of students who eventually solve the problem is also a good data point. However, does this mean that they solved the problem after they hinted out? on the first, second, third try? Could you eventually click on the answers until you get the right one? If so then this data point tells me nothing.

Percentage of students who solved it on the first try/second try are excellent data points. They are not ambiguous and tells me as a teacher what concepts students know well. Given these two data points does that mean that if I add them together and subtract the, from the percentage of students who eventually solved it then I would get the percentage of students who solved it after the second try or more. Are they mutually exclusive events?

Most frequent incorrect response is a great data point. It would be nice to see what percent of the students answered that.

The last two columns I am not sure what they mean. Average number of hints on the last encounter and average comprehension time solving. I am also not sure what the red and yellow signifies.

- Class summary per problem report student is probably the least useful report to me as a teacher. It provided little information to me on how I can modify my instruction to meet the needs of my students on certain topics. The only thing it tells me if students were clicking through hints. But it doesn't tell me if they were clicking through hints on all problem sets, problems sets that were linked to a certain objective, etc.

- **Class summary per common core.**

The number of problems columns appear to be the total number of problems seen by all students in this learning objective. For instance, Gr 7 proportional relationships, students saw 112 problems. But when I click on the objective I am brought to the Problem Difficulty Report, If I add up the problems here there are a total of 66 number of students that have seen the problems. That means of those 112 problems students could have seen the problem twice. It is nice that the class summary per Common Core gives a summary for each cluster but as a teacher I would like to see it broken down further.

For example, there should be problems for each objective within the standard Ratios and Proportions 7.RP.

As a teacher I identify priority standards. I think MathSpring would be very beneficial for my students to practice and assess their knowledge on each objective within the standard.

CONTENT: Ratio and proportions contains the objectives below. A problem set for each objective should be developed.

1. Computing unit rates (including complex fractions)
2. Recognizing and representing proportional relationships
 - a. By a graph, table, equation. Students must know that if a graph goes through the origin and is a straight line then we can say that x and y are in proportional relationship or that x and y show direct variation or that x and y are proportional. All of these are synonymous. I think there should be a problem set that link all of these concepts.
 - b. They must know that if x and y are in a proportional relationship then the slope of the graph is a unit rate. They should be able to identify the constant of proportionality (unit rate) in tables, graphs, equations, diagrams, and verbal descriptions of proportional relationships. Again a problem set for this skill should be developed.
 - c. Represent proportional relationships by equations. For example, if total cost t is proportional to the number n of items purchased at a constant price p , the relationship between the total cost and the number of items can be expressed as $t = pn$.
 - d. Explain what a point (x, y) on the graph of a proportional relationship means in terms of the situation, with special attention to the points $(0, 0)$ and $(1, r)$ where r is the unit rate.
3. Use proportional relationships to solve multi-step ratio and percent problems. Examples: simple interest, tax, markups and markdowns, gratuities and commissions, fees, percent increase and decrease, percent error. Here is where there should a problem set just on percent. Finding the part, whole, and percent is simple problems and real life applications.

- **Class summary per skill.** I think skill reinforce problem sets should be offered. These are skills that students should master before attempting grade level problem sets. For instance, pre-requisite skills for proportional relationships would be those sixth grade skills under CCSS for ratios and proportional relationships. For instance, “the what you learned before” should be:
 - Writing and describing a ratio
 - Understanding the concept of a unit rate and how to calculate it
 - Understanding ratio tables and finding missing values in tables and plotting pair of values on a coordinate plane
 - Solving simple proportions, if it took 7 hours to mow 4 lawns, then, at that rate, how many lawns could be mowed in 35 hours? At what rate were lawns being mowed
 - Solve percent problems using “nice percent’s” 25% of 80? 50% of what number is 70?
- **Pre- and Post- Test Information.** This looks like a useful report but I would like to know how to read it. I am not sure what problems were given as a pre-test and posttest. When I click on the student name it links me to the detailed student info report which doesn’t tell me much (unless I am missing the connection). I still want to know what problems my students worked on that particular day. It is too bad there isn’t a report that breaks down the stats for each day a student works on MathSpring. The data is lumped together.
- I find some information in the reports of value, however I am having trouble seeing how I can take the reports and get a snapshot of what my students know and don’t know. Some of the data may be more helpful to the people running the study but not to me as a teacher.
- It would be awesome if the program can be used to assess students’ knowledge of the important skills identified in the CCSS for their grade level. It is not very helpful for a student to be working in a problem set, work on a seventh grade skill then drop to a 5th grade skills.

Appendix B

Visualizing Student Data in Mathspring -----Rafael Lizarralde - 2017 October 23

Introduction

Mathspring has accumulated a wealth of student data as successive groups of students have interacted with the system. It tracks hint usage, timing of student actions within a problem-solving attempt, correctness of the various attempts, as well as the difficulty and relevant Common Core Standards for each problem. However, the existing resources for presenting this data to teachers is still rather limited. The four reports currently available mostly present information in the form of simple tables, without ways of guiding teachers to useful conclusions. In this report I seek to review these existing reports and teacher feedback on them, as well as suggest ways of improving or reinventing the reports to better inform teacher decisions in the classroom or while customizing the system.

Visualizations Offered by Mathspring Currently

Currently, Mathspring provides four reports for teachers, all of which are in the form of tables with a small amount of interactivity.

Class Summary Per Student

This report provides information on each student's number of problems attempted, with the option to show a suitable with a particular student's per-problem breakdown of correctness, mistakes (incorrect attempts), hints, total attempts, and effort type. Effort Type can also be shown as a bar graph.

However, there are a few problems with this report. The top level of the report shows far too little information; the only thing it presents is how much work each student has done in the system. Most of the information is hidden in the sub tables, which are not adequately summarized. For example, the top level table could show a summary of the number of correct attempts, percentage of topics mastered, average time spent per problem (this may need to be normalized by each problem's time average), average number of hints seen, and so on; there's a wealth of per-student information that is missing.

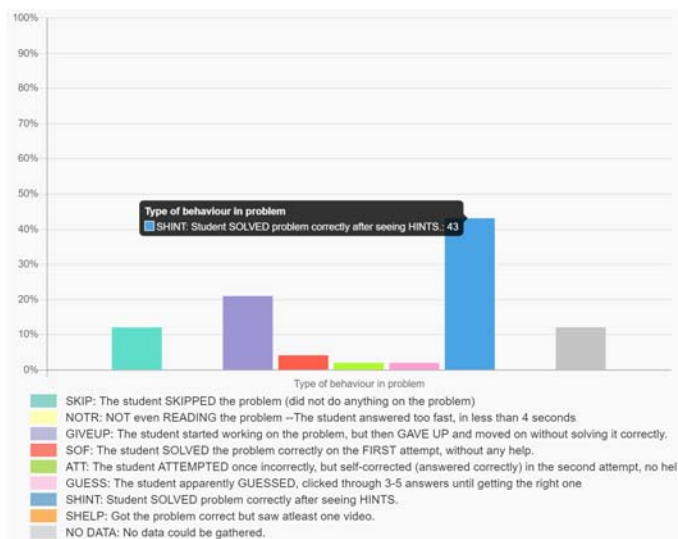


Figure 1: Current visualization of effort values as a bar chart.

The SHINT bar is showing its hover-over help text.

The effort chart, while useful, is presented in a very inefficient way; there are nine effort categories, and each is shown as a percentage in a bar graph, with the scale showing all the way up to 100%. This wastes most of the space, as typically each bar is less than 50% (Figure 1). Additionally, the key is a "wall of text", rendering it pretty useless. Fortunately, the graph is interactive, and will show you the corresponding key entry on hovering over a bar. Finally, once expanded, the only way to collapse it is to expand another part of the table; as it takes up so much space, one must often scroll down just to be able to expand the next student's effort chart. I believe this could be much more compactly and effectively conveyed with a single horizontal stacked bar, as discussed in the *New Visualizations: Stacked Bar Chart* section.

In addition to the design problems of the report, there are also several visual bugs. The effort chart expands off-screen without providing the ability to horizontally scroll, which cuts off some of the key. When expanding a student's per-problem suitable, it first creates the space for it, followed by an animation of the table expanding upward. This violates the physical intuition of expanding/collapsing elements; it should expand outward from the element that was clicked (in this case, downward rather than upward).

Class Summary Per Student Per Problem set

This report provides information on each student's performance in each topic area. It presents this in four ways in each cell: showing the fractional number of problems solved on the first attempt, a mastery value, a color coding of mastery value, and an expandable chart of mastery over time.

The numbers could be better separated; currently there is not even a space between the fraction and the mastery number. In most cases, the topic name is longer than the contents of any cell, so spacing the numbers out more would not increase the size of the table significantly. The table could be made more compact as well by allowing for wrapping in the column titles and giving each column a set width.

Mastery Range	Grade/Color Code
0.75 or Greater	Grade A (Excellent)
Between 0.5 and 0.75	Grade B (Good)
Between 0.25 and 0.5	Grade C (Needs Improvement)
0.25 or Less	Grade D (Unsatisfactory)
Fewer than 10 problems solved	Not enough data

Figure 2: A modified mastery key for the Class Summary Per Student Per Problem set report which indicates the lack of color means

Mastery Range	More than 10 problems solved	Fewer than 10 problems solved
0.75 or Greater	Grade A (Excellent)	Estimated Grade A (Excellent)
Between 0.5 and 0.75	Grade B (Good)	Estimated Grade B (Good)
Between 0.25 and 0.5	Grade C (Needs Improvement)	Estimated Grade C (Needs Improvement)
0.25 or Less	Grade D (Unsatisfactory)	Estimated Grade D (Unsatisfactory)

Figure 3: A modified mastery key for the Class Summary Per Student Per Problem set report which uses lower saturation to indicate partial

insufficient data.

data.

While the color-coding is useful, the lack of color in most cells is problematic. As the text above the table explains, colors are only used for cells with 10 or more problems. This information should really be in the key instead (Figure 2). However, it would be even better to use saturation to convey the level of confidence; as more problems are solved, the color gets darker. This would be straightforward to add to the key as another column (Figure 3), or added in a more continuous way as discussed in *New Visualizations: Heat map Tables*.

The mastery over time chart might be helpful, but could definitely benefit from some improvements. For example, the button to show it could summarize the most recent changes (over the last several problems), by having the icon show an up arrow, down arrow, or right arrow. This would allow teachers to see recent developments at a glance, without having to flip through each of the charts. Additionally, while the charts do inherit the color from their cell, they should really be showing the color as it changed over time as well.

Finally, unlike the previous report, this table is paginated, requiring clicks to see different arbitrary groups of students. This is unnecessarily limiting, as it is really no better than scrolling, and also prevents things like zooming out to see the whole table at once. It also has the undesirable quality of causing the layout of the page to jitter when flipping between pages that contain different numbers of students. This pagination issue is repeated for the next two reports.

Class Summary Per Problem

This report shows a variety of interesting information about overall class performance on each problem, such as the relevant Common Core Standard, number of students who have seen the problem, solved it on the first attempt, repeated, skipped, or given up on the problem, as well as the most frequent incorrect response. It also color-codes the percentage of students that solved the problem on the first attempt.

While color-coding is always nice, the coding here is very inconsistent. First, it uses color, absence of color, and a symbol to indicate different levels of the same thing; it would be much better to use color consistently for all of them. Second, it only applies to the "solved on first attempt" column. On the one hand, if it were to be extended to other columns, it could be confusing because each column has very different significance (for example, it is desirable for the percentage of solving on the first attempt to be high, but for the percentage of students who gave up to be low). On the other hand, the precise cutoffs of the colors shouldn't actually matter much to teachers; what is relevant is that positive or negative information is being highlighted. For a more consistent color-coding system that could be applied to the other columns, the key could simply read "Poor", "Okay", "Good", "Excellent", and use the same color coding as the mastery scale from the previous report.

The next issue with this is much subtler. When students are using the tutor, they are often encouraged to use hints to solve the problem, but this report presents that to the teacher as a negative behavior. In a typical instance of this report, almost all cells are red or uncolored, indicating that most problems had very few students solving them on the first attempt. To a teacher, that sounds terrible. Although this goes beyond visualization, it may be worth reexamining what this column really represents to a teacher and how it could be presented in a more accurate way. For example, perhaps a low value here really represents that most students are still learning this material, while a high value indicates that solving the problem was a mostly redundant exercise. This could be explained in a help hover-over like the other columns have.

Finally, like with the *Class Summary Per Student* report, there is a variety of information recorded by Mathspring that could be presented here, such as average time spent on each problem, average percentage of hints used, and problem difficulty rating.

An alternative view of this report is that most of the columns are essentially bars from the effort chart in *Class Summary Per Student*, but aggregated by problem instead of by student (specifically, the columns for solve on first attempt, skip, and gave up). In this sense, much of it could be replaced by a single horizontal stacked bar graph, as discussed in *New Visualizations: Stacked Bar Chart*.

Class Summary Per Common Core Cluster

This report shows the number of problems encountered, percentage of them that were solved on the first attempt, and average "ratio" of hints requested. The problems encountered can be expanded to show the relevant rows from the *Class Summary Per Problem* report.

The number of problems encountered and its expanded suitable are functional and somewhat useful, but the information contained there can already be found by sorting the *Class Summary Per Problem* by Common Core Standard. Additionally, the nested table pagination here is confusing and should be removed.

The other two columns appear to be broken or misleading. The percentage of problems solved on the first attempt frequently shows 100% when the per-problem information suggests otherwise, and the last column appears to be showing the average number of hints, not a ratio.

In an example class, the *Class Summary Per Problem* showed that most problems were difficult for students, while this report has not highlighted any clusters with either of the two provided levels of difficulty.

As it stands, this report should be removed, and ideally replaced with a more comprehensive one.

Teacher Feedback on Current Visualizations

Susan Tarallo provided a large amount of useful feedback on the previous set of class reports. Fortunately, some of this has already been incorporated, such as viewing problem previews alongside the data, showing hints per problem, and a better explanation of the mastery rating. However, there were several other pieces of information she liked and wanted more detail for:

- The most frequent incorrect response to problems
- Solve on first attempt; she would like to know about second, third, and so on
- Time to make a guess (this has since been removed!)

Tarallo was also interested in per-day versions of the reports. From the interface side, this could be relatively easily added by having reports initially open to a summary, then use pagination (as is currently used to show more results) to flip to the most recent day and continue backward (Figure 4).



Figure 4: *Pagination controls for viewing a summary of data over all days, as well as paging through individual days.*

Although not directly mentioned, it seemed like she was more interested in Common Core Standards than the topics system. It would make sense to have *Class Summary Per Student Per Problemset* matched or replaced by one with Common Core Clusters instead of Problemsets.

Teacher Use Cases

Before deciding on how to present the data Mathspring has, we must first examine what actions teachers might be able to take in response to this data. In most cases these actions could be done either

within Mathspring (should teacher customization of the system be improved), or outside of Mathspring in the normal classroom.

1) Adjust or review problems for the class

This is already possible in Mathspring, as teachers can change topics, the grade level of the class, and select individual problems to activate. In order to decide whether to do this, a teacher would need information on how problems or categories of problems are being solved by students. For example, if a particular Common Core Cluster is being quickly and accurately solved by all the students, the teacher may want to disable it in order to help students focus on clusters they find more challenging. Alternatively, if one topic or cluster is proving difficult for most students, the teacher could review the topic with the class.

Currently, the *Class Summary Per Problem* and *Per Common Core Cluster* reports facilitate making this decision, but could be improved.

2) Adjust or review problems for a particular student

While not currently possible in Mathspring, there are a number of similar actions a teacher could take in the classroom, such as going over a problem with the student or assigning similar problems for homework. The most helpful data for this decision is individual student performance and coverage across problems, topics, or Common Core Clusters. However, this data is very difficult to collapse into easily readable summaries (there is no way to escape having two dimensions of categorical data; students and topics/clusters), and may require automated analysis to be useful, such as discussed below in *New Visualizations: Insights*.

Currently, the *Class Summary Per Student* (when looking at the subtable for a particular student) and *Per Student Per Problemset* reports help teachers decide whether to do this. The *Per Problemset* report is fairly effective at communicating areas that might need work, as it color-codes areas of high and low mastery. However, it doesn't convey low-confidence information as well as it could, and the lack of an analogous *Class Summary Per Student Per Common Core Cluster* report limits usefulness for classroom actions. *Class Summary Per Student* is not particularly helpful, as it takes a lot of interaction and sifting through subtables to get anything meaningful from it. However, this report could be improved significantly as discussed above in *Visualizations Offered by Mathspring Currently*.

3) Adjust or review problems for a group of students

Like per-student customization, this is not possible in Mathspring, but is feasible in the classroom in the form of group exercises. While this requires the same data as per-student adjustment, there are ways to present it that could facilitate this decision, such as clustering students in a per-student report by their mastery or effort profiles, as discussed below in *New Visualizations: Heatmap Tables*.

4) Encourage or congratulate a student

In order to determine whether particular students need encouragement or have been behaving studiously, teachers need a more nuanced look at student behavior in the system, both in terms of the type of behavior and in how it changes over time.

Currently, the *Class Summary Per Student* report can shed some light on this through the effort charts, but it is very cumbersome to flip through and interpret effort charts. However, by improving the presentation of the effort chart (such as with a stacked bar displayed directly in the table) this could be made much easier. Additionally, adding the ability to view these results by individual days or weeks, it would be much easier for teachers to identify shifts worthy of attention-- it would even be possible to highlighting large changes in behavior, such as a "plus" icon on SOF to indicate that the student has been solving more of their problems on the first attempt this day, as compared to before.

New Visualizations

A key limitation in deciding how to visualize data in Mathspring is that it must be understood by teachers using the system. These teachers are incredibly time-constrained, so we must avoid any barriers to adoption, such as needing to learn how to read a new visualization. This limits us to presenting data through tables, common charts (bar, line, pie, etc), timelines, and text (ideally short and direct). Other more unusual visualizations, such as parallel coordinates, tree maps, and cone trees, (Khan & Khan 2011) are unlikely to be very accessible.

I propose four ways of presenting Mathspring's data to facilitate these teacher use cases. The first, *Heatmap Tables*, are a modification of the existing tables, but formalize the use of color and saturation to guide teachers to useful information. Second, *Stacked Bar Charts*, are a way to more compactly show data covering mutually exclusive categories, such as effort behaviors. Third, *Problem Timelines* provide a visualization of student behavior in a temporal context, by showing the student actions over the course of an individual problem. Finally, *Insights* are short text messages to teachers, which allow for highlighting specific scenarios that are otherwise difficult to identify.

Heatmap Tables

A heatmap table is a simple extension to the existing tables in the reports. *Class Summary Per Student Per Problemset* is already nearly a heatmap table; it only requires a coloring scheme that can cover all populated cells with values, rather than just the high-confidence ones that are currently covered. Typically a heatmap table also represents colors along a spectrum rather than categories. So this report could be easily converted to a heatmap table by assigning hue to the mastery level and saturation to the number of problems seen (Figure 5).

Student Name	Username	Area of Polygons	Angles	Triangles	Special Triangles	Pythagorean Theorem	XY Linear Functions and Relationships
name1	user1	[1/6] 0.21	[0/1] 0.10				[4/5] 0.21
name2	user2						
name3	user3	[2/5] 0.38	[2/5] 0.21	[5/17] 0.38	[3/10] 0.44	[0/6] 0.10	[2/9] 0.21
name4	user4	[3/7] 0.44	[3/6] 0.38	[5/13] 0.64			[3/8] 0.44
name5	user5	[0/3] 0.10					[0/8] 0.10
name6	user6	[3/5] 0.21	[2/6] 0.38	[5/10] 0.59	[1/4] 0.21	[1/7] 0.20	[2/4] 0.21
name7	user7	[1/6] 0.21	[1/7] 0.21	[5/19] 0.64		[1/6] 0.20	[3/8] 0.21
name8	user8	[2/6] 0.21	[3/7] 0.44				[1/7] 0.21
name9	user9	[0/1] 0.10	[0/3] 0.10	[1/6] 0.21		[0/2] 0.10	[4/8] 0.72
name10	user10	[2/6] 0.21	[2/5] 0.21	[3/12] 0.21	[3/9] 0.38	[0/7] 0.10	[0/9] 0.10

Figure 5: Class Summary Per Student Per Problemset, presented as a heatmap table. Hue represents mastery level, and saturation represents the number of problems supporting the mastery level.

Use of heatmap tables is most appropriate for addressing the use cases 1 and 2, where problems are adjusted for the class or individual students. For use case 1, adjusting problems for the whole class, the best table would be a combination of *Class Summary Per Common Core Cluster* and *Class Summary Per Problem*. Instead of restricting color coding to "challenging" and "really hard" clusters as the table currently does, all columns could be color-coded, and the columns used by both the subtable and the outer table should be the same. Additionally, adding a button to expand or collapse all clusters would help the teacher switch between per-cluster and per-problem modes (the current functionality of expanding and collapsing one at a time facilitates looking into the details of a cluster, but makes it difficult to get an overall picture of all the problems, which is why the second *Class Summary Per Problem* exists separately). However, a heatmap table may not be the best way to present the information relevant to *Class Summary Per Problem*. An alternative is presented below in *Stacked Bar Chart*.

For use case 2, *Class Summary Per Student Per Problemset* would be a good heatmap table if it colored all cells as suggested above, with hue for mastery level and saturation for the number of problems solved (Figure 5).

For use case 3, adjusting problems based on groups of students, an additional modification is required. Automatic clustering of students could be used to suggest groups with similar attributes, who could then be given customized group exercises or other classroom interventions. This could be represented in the table by placing the rows adjacent to each other, and adding a separator between the

cluster and other students. One class of clustering algorithms is hierarchical clustering (Murtagh 1983), which have the advantage of providing different levels of granularity. So while the default number of clusters might be set to 5, a slider could be added above the table to display the clustering into different total numbers of clusters. However, hierarchical clustering does not necessarily divide the class evenly into groups; this could either be left up to the teacher, or alternative clustering methods that prioritize class balancing could be explored, such as Regularized Information Maximization (Krause et al. 2010) or Same-Size K-Means (Schubert & Zimek 2017).

For the *Class Summary Per Student* report, clustering could use effort values as features, while the *Class Summary Per Student Per Problemset* would use mastery levels. However, in this case we must be more careful about how we define distance between two students. There are two potential issues: first, mastery can be undefined if no problems in the area have been solved, and low-confidence if few problems have been solved; second, because use case 3 is about finding groups of students who should be working on something similar, it will be most relevant to find shared areas of low mastery, while different areas of high mastery on their own may not be relevant on their own. One way to account for both of these issues would be to scale distance linearly with the minimum or average of both mastery levels' uncertainty, so that if one student has no data on a particular topic, then the distance in that dimension of the clustering space is zero, while if both have high-confidence mastery levels, then the distance is simply the difference between the two levels.

Stacked Bar Charts

Stacked Bar Charts are more commonly presented like a normal bar chart, but in addition to showing bars for categories, each category has an internal breakdown of components. Each stacked bar can be thought of as a pie chart in linear form; this is actually a better way to display the information than pie charts, as research has shown that pie charts are actually harder to interpret than linear representations (Cleveland 1985).

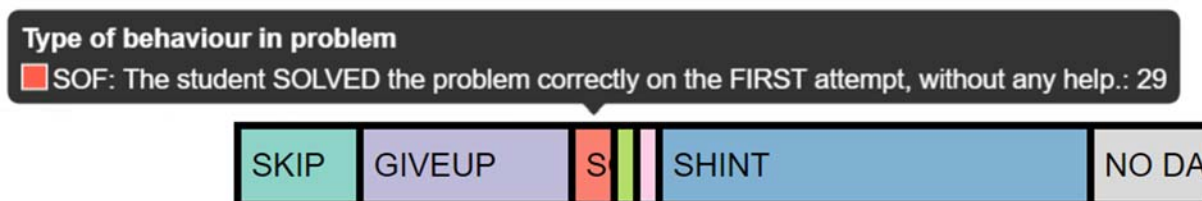


Figure 6: A horizontal stacked bar chart representing the proportion of problem solutions that are marked by particular effort events.



Figure 7: A horizontal stacked bar chart representing the relative frequency of various incorrect answers on a problem. Currently, only the letter C would be shown with this data.

In our case, effort values are an ideal scenario for using stacked bar charts. Effort consists of categories, which together cover all problem scenarios (when also accounting for the "NO DATA" case). By arranging the bar horizontally, this would allow a single table to show all of the bar charts at once, making it easy for a teacher to, at a glance, compare effort behavior between students or clusters (Figure 6). Stacked bars could also be used to represent the most common wrong answers to problems (Figure 7) - Susan Tarallo found the existing information helpful and wanted to know more; this would make it easy to represent the full spread of wrong answers, while still clearly indicating the most common one.

Like the existing bar chart, this would require a little bit of effort to understand initially, but by using the hover-over text hints like the chart does, a legend would not be necessary. Seeing all of the stacked bars together would also help with learning, as it would be easy to compare, hover over, and read what components are contributing to the visual differences.

The *Class Summary Per Student* report could use stacked bars to show effort; this would address use case 4, encouraging or congratulating students, by allowing the teacher to recognize exemplary or problematic behavior. The combined *Class Summary per Common Core Cluster and Problem* report could use effort bars aggregated over the class within a cluster or problem; this would address use case 1, adjusting problems for the class, by showing problems or clusters where behavior is substantially better or worse. For example, if a cluster is too difficult, the teacher may see that students are skipping its problems the majority of the time, and decide to review that cluster with the class.

Problem Timelines

While stacked bars can show the behavior of students on problems categorically, it does not display temporal data, which can help understand how students are approaching problems. For example, a scenario that the effort value would have marked "GUESS" may in fact be the student working through a difficult problem and making some mistakes; knowing how much time was spent between attempts is the key to separating those two scenarios. Similarly, "SKIP" may be the student simply flipping through problems without reading them, or reading them and deciding they don't know how to solve it; these two behaviors should be addressed differently by interventions.

A problem timeline consists of a line, on which the problem start, finish, hint, and attempt events are plotted (Figure 8). In order to address the issue where the scale of the time is very different between problems, a standard scale is set (for example, the full length of the line is 5 minutes), and then timelines that exceed this are shrunk to fit and marked with their total duration; I have included an example implementation of this in HTML in *Appendix C*. To help teachers explore timelines and get a sense for the scale of the time, hovering over the events and line segments between should show more detail, such as "asked for their first hint", or "3m 27s seconds before first attempt".



Figure 8: A problem timeline showing a significant problem-reading period, then an incorrect attempt, followed quickly by a hint, and later by a correct attempt.

In addition to being displayed for a specific problem solving attempt, it may be helpful to show a "typical" problem timeline for some category, such as a student's typical timeline, drawn from all of their problem solutions on a day. I have included a draft algorithm for averaging timelines in *Appendix A*.

Problem timelines would make an excellent supplement to effort stacked bar charts for the purpose of supporting use case 4 (encouraging or congratulating a student), as well as getting more context for a decision in use cases 1-3 (adjusting problems). They could be added to the *Class Summary Per Student* report, displaying an average timeline directly in the column and the exact timeline in each problem row when the per-problem summary is expanded. Similarly, average timelines could be included in the *Class Summary Per Common Core Cluster or Problem* combined report, although they may not be as useful there (averaging temporal behavior between students may not be very accurate).

Insights

While the previously discussed visualizations do a lot to support the four use cases, they still leave it entirely to the teacher to analyze the data and find meaningful scenarios. Insights, in the form of short snippets of text suggesting an action based on a trend, can be used to directly provide actions for teachers to carry out. This can both relieve some of the burden of analysis, and also help teachers to learn how to explore the data on their own.

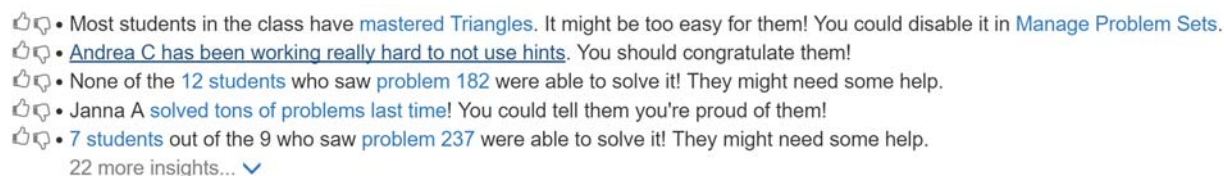


Figure 9: A hypothetical set of top 5 insights, containing links to relevant data or hover-over elements that show the relevant problem or students.

Insights could be displayed in a section above the existing reports, and provide links to parts of the report that support the suggestion (Figure 9). For example, insights could suggest:

- [student] has been working really hard to not use hints! You could congratulate them!
- Problems in [topic] have had lots of mistakes and skips today. Maybe the class could use a review?
- Most students in the class have already mastered [topic]. It might be too easy for them!
- None of the [number] students who saw [problem] were able to solve it! They might need some help.
- [student] has been solving tons of problems today! You could tell them you're proud of them!
- Students have been taking a long time to solve [problem]. Maybe you could teach them a trick for it?

The key challenge for Insights will be selecting which ones to display. In order to avoid overwhelming teachers, there should only be a few displayed initially. Each type of insight needs a way of ranking how useful it is; this could be calculated with a base level of interest (a bias term), multiplied by a normalized magnitude of effect seen. For example, for insights about changes in mastery, the magnitude could be the average change in mastery level; for a student working hard to avoid hints, the magnitude could be the difference in the number of SOF events divided by the total number of SOF and SHINT events. The difficulty lies in making sure that these rankings are consistent with each other, otherwise certain types of insights will dominate the rankings or never show up.

The weights used for this ranking could be adjusted with teacher feedback, by providing a simple "thumbs up" / "thumbs down" system. These votes could then be used to adjust the bias terms for insights of that type. In order to improve the personalization of this filtering, multiple bias terms could be used, such as a bias term for the type of insight, and in the case of insights where they can have positive or negative forms, separate bias terms for each of these, as well as global (for all teachers) and local (for just this teacher) versions of the bias term. Updating the bias terms with decay will prevent continued voting from causing them to scale too far; a simple exponential decay update rule for doing this is presented in *Appendix B*. Because different insights may have different numbers of bias terms, they should be combined by averaging or multiplying them together (assuming they range from 0 to 1) rather than being summed.

Alternatively, this could be viewed as a collaborative filtering problem, where insight types are items and teachers are users providing ratings of these items. There are many different approaches in this area, but item-based (also known as model-based) algorithms are likely to be more appropriate (Sarwar et al. 2001). However, recommender systems typically rely on a larger body of rating data than Mathspring is likely to accumulate.

Final Recommendations

Topics or Clusters

One of the major readability problems for teachers using the system is the alternate classification of problems as belonging to Common Core Clusters or Mathspring Topics. Based on teacher feedback, they seem more interested in viewing them through clusters. This is most likely because it aligns with the existing structure they must implement in their own classroom, as their students will be assessed according to Common Core Standards.

I believe that topics and clusters are essentially interchangeable from the system standpoint, and it would greatly benefit teachers to remove topics and organize problems by clusters instead. This could be relatively easily achieved by creating a new set of topics that are directly aligned with Common Core Clusters, and then reassigning problems to these new "topics".

However, if Mathspring keeps the existing topic system, then it becomes important to present many of the tables in both Topic and Cluster form, because teachers will want to use Clusters to view the data when deciding how to help the students in their own classroom, but need to view it through Topics when deciding how to adjust what questions are active within their Mathspring classroom. By removing topics, this switching capability or duplication of reports would be unnecessary, which would also make the reports easier to navigate and read.

Ideal set of reports

Accounting for both the existing reports and the new visualizations I have discussed, I believe that an ideal set of reports would consist of the following:

1. *Insights*: approximately five of them displayed at the top of the reports section
2. *Class Summary Per Student*: For each student, display a stacked bar chart for effort values, an averaged problem timeline, and several other statistics about overall usage, such as number of problems attempted, number of clusters encountered, and a stacked bar chart for the number of clusters with A, B, C, D, or unknown grades (as defined by mastery). The subtable for individual problem data should include problem timelines. Above the table should be a slider to be used for adjusting the clustering of students into groups based on their effort behavior.
3. *Class Summary Per Student Per Common Core Cluster*: A heatmap table adapted from the existing *Per Problemset* report. Above the table should be a slider to be used for adjusting the clustering of students into groups based on their mastery of various clusters.
4. *Class Summary Per Common Core Cluster and Problem*: The top level should be a table containing clusters, the number of problems encountered by the class, an effort value stacked bar chart, and an average problem timeline. A toggle to expand or contract all clusters should be available above the table, and expanded clusters should show the same data, but also show a stacked bar chart for most frequent incorrect responses.

For reports 2, 3, and 4, pagination should be used to allow flipping between a complete summary and per-day reports.

Future Work

The most potential for future work lies in the clustering methods and the design and ranking of insights. Different approaches to clustering and presentation may significantly improve its usefulness as a tool to teachers. Insights have great potential, but may take a large amount of work to produce valuable suggestions and rank them appropriately. Further research into parallels with recommender systems and collaborative filtering may uncover more similar work that could better inform approaches to ranking insights.

References

- Arroyo, I., Woolf, B.P., Burleson, W., Muldner, K., Rai, D., Tai, M. (2014) *A Multimedia Adaptive Tutoring System for Mathematics that addresses Cognition, Metacognition and Affect*. *International Journal on Artificial Intelligence in Education*.
- Bienkowski, M., Feng M., Means, B. (2012, Oct). *Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief*. U.S. Department of Education Office of Educational Technology.
- Donoho, David. (2015, Oct). *50 years of Data Science*.
- Heffernan, N., Heffernan C. (2014). *The ASSISTments Ecosystem: Building a Platform that Brings Scientists and Teachers Together for Minimally Invasive Research on Human Learning and Teaching*. *International Journal of Artificial Intelligence in Education*. 24(4), 470-497