Digitally-Delivered Diagnostic Assessments

Based on Learning Trajectories Supporting Personalization and Competency-Based Approach in 6–8 Mathematics

Jere Confrey, Project Director

Joseph D. Moore Distinguished University Professor

Science, Technology, Engineering and Mathematics (STEM) Department
College of Education, North Carolina State University

Garron Gianopulos, Psychometrician

Science, Technology, Engineering and Mathematics (STEM) Department
College of Education, North Carolina State University

Charlene Marchese, Mathematics Supervisor

Freehold District

Brian Gong, Center for Assessment

Wednesday, June 22, 2016 Session Time: 10:00 AM, Session Location: Franklin Hall 13

Project Support

The Bill and Melinda Gates Foundation

National Science Foundation

Scaling Up Digital Design Studies

Overview

- The SUDDS team
- How our DLS can help close the gap
- A demonstration of the Digital Learning System (DLS)
- A closer look at our assessments
- What we have learned

The SUDDS Team

Jere Confrey
Project Director

Will McGowan
Project Coordinator

Meetal Shah
Learning Sciences

Garron Gianopulos
Psychometrician

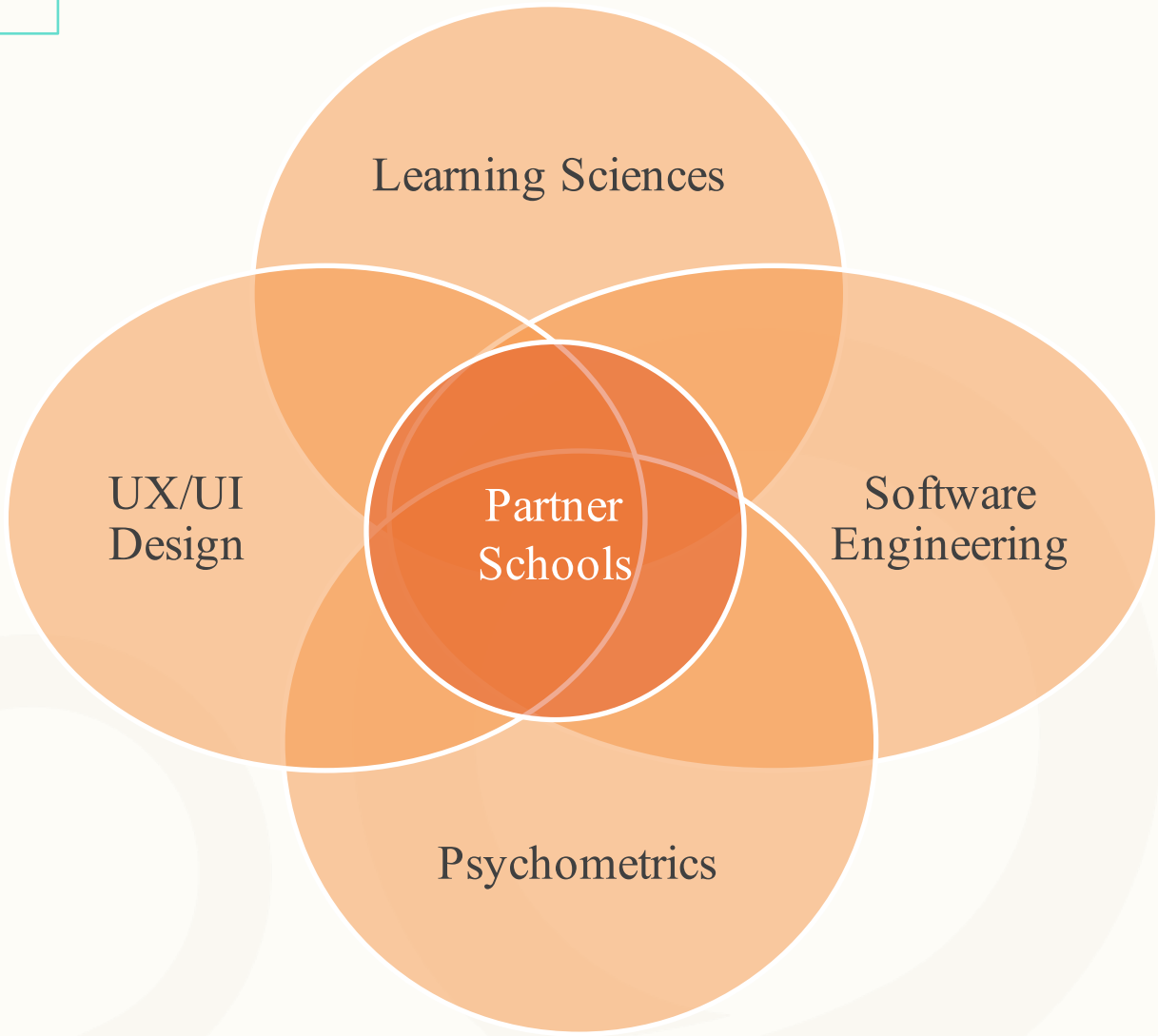
Yungjae Kim
Software Engineer

Douglas Ivers
Software Engineer

Basia Coulter
UX/UI Designer

Other contributors: Pedro Larios and Seth Jones

A Cross-Field Approach



Research-Derived Design

- Synthesis of the Literature
- Design Studies
- Usability Studies
- Cognitive Interviews
- Field Tests
- Agile Methodology

Foundational Beliefs

- Start with the end-user in mind
- Maximize functionality and aesthetics
- Design should be theory-driven
- Learning theories should be explicit and testable
- Improvements should be continuous
- Smart machines cannot replace teachers

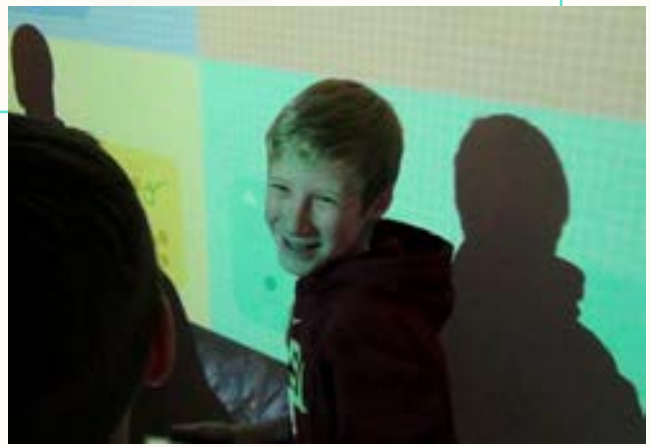
Overview

- The SUDDS team
- How our DLS can help close the gap
- A demonstration of the Digital Learning System (DLS)
- What we have learned
- New research directions and next steps

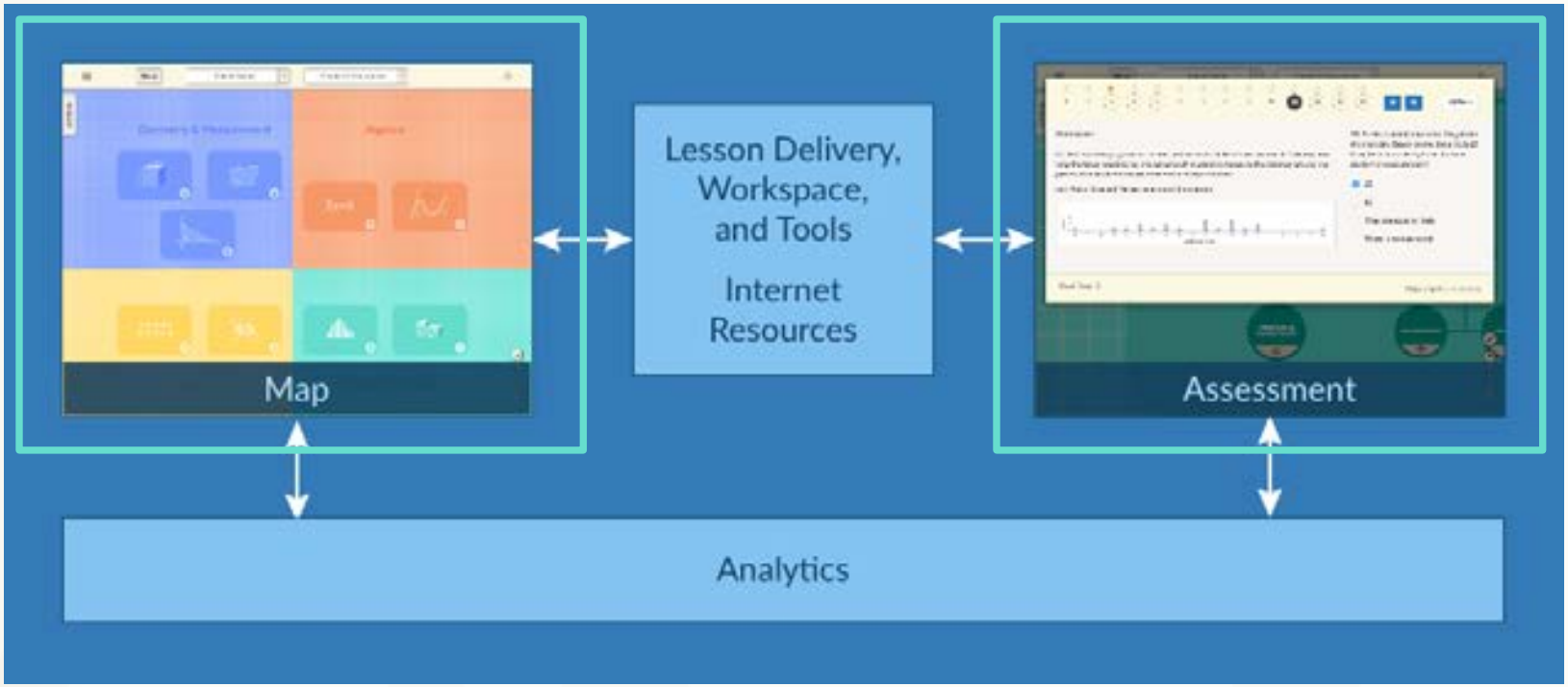
What do Middle Grade Math Teachers and Students Need to Close the Performance Gap?

The Goals of the SUDDS Digital Learning System (DLS)

- Increase Coherence of Digital Content
- Personalize Learning
- Provide Actionable Feedback
- Support Flexible Grouping



What is a Digital Learning System?



Confrey, November, 2015

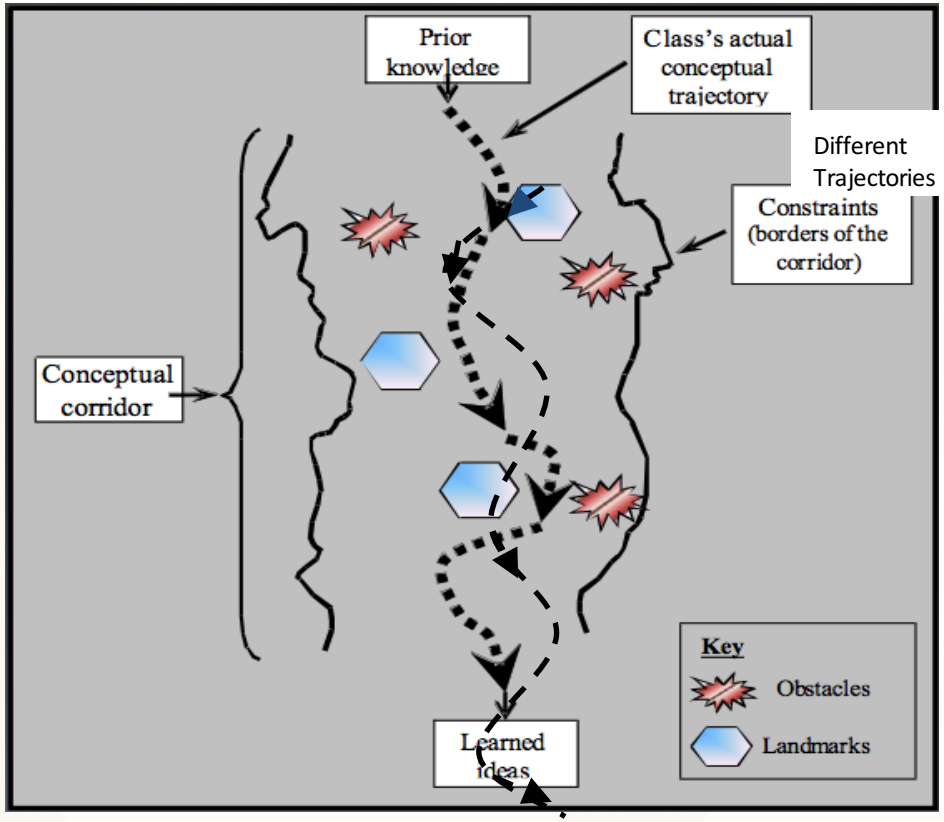
The SUDDS DLS

Our tool is an innovative digital container, where students and teachers can…

- Navigate the content of middle school mathematics around big ideas and research-based learning trajectories
- Select and sequence aligned open source curricular resources
- Assess, in real time, students' progress, identifying needs and next steps

What is a Learning Trajectory?

A learning trajectory connects what students bring to instruction, to a target concept, and delineates a set of landmarks and obstacles that students are likely to encounter as they move from naïve to sophisticated understandings



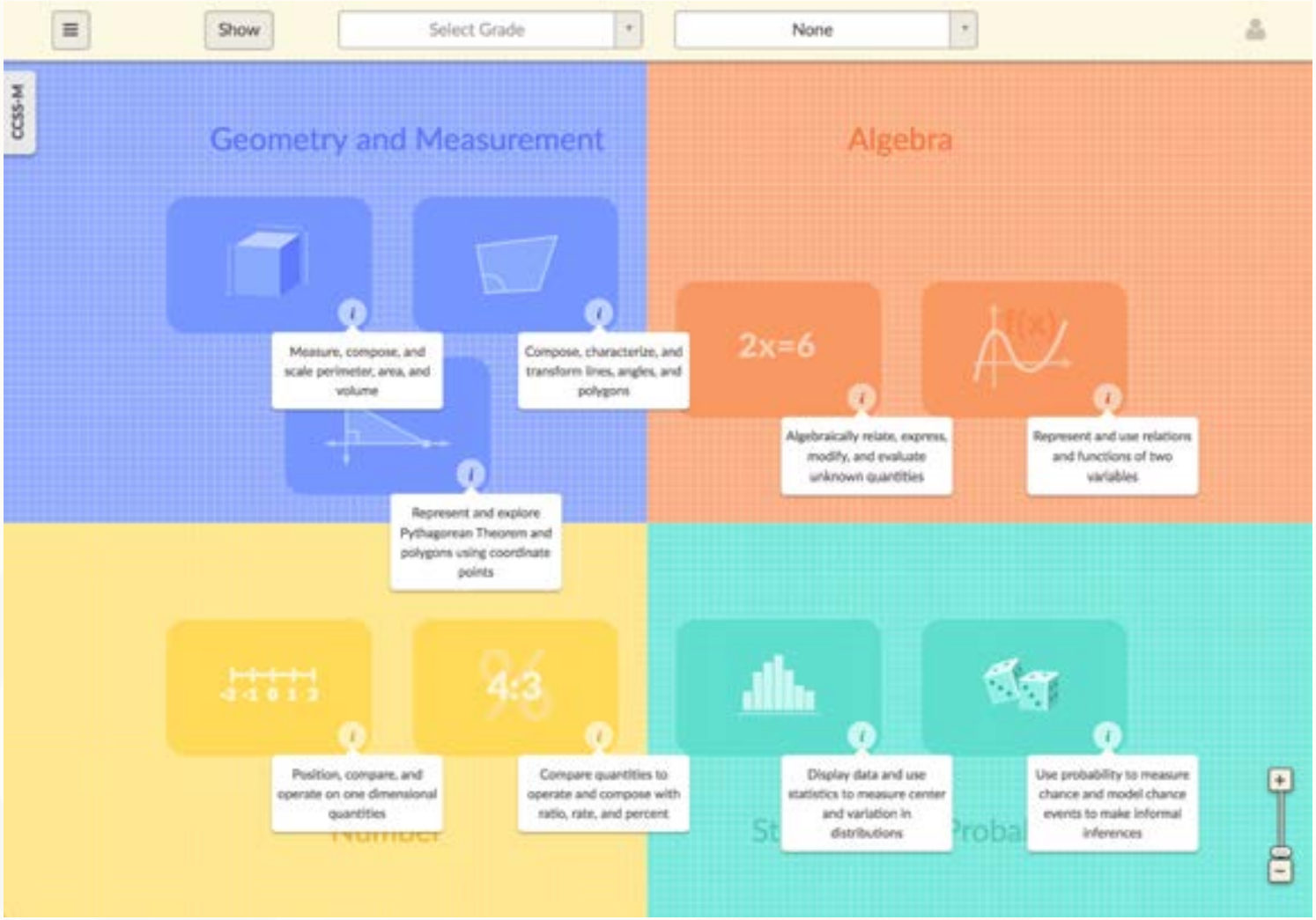
Overview

- The SUDDS team
- How our DLS can help close the gap
- A demonstration of the Digital Learning System (DLS)
- A closer look at our assessments
- What we have learned

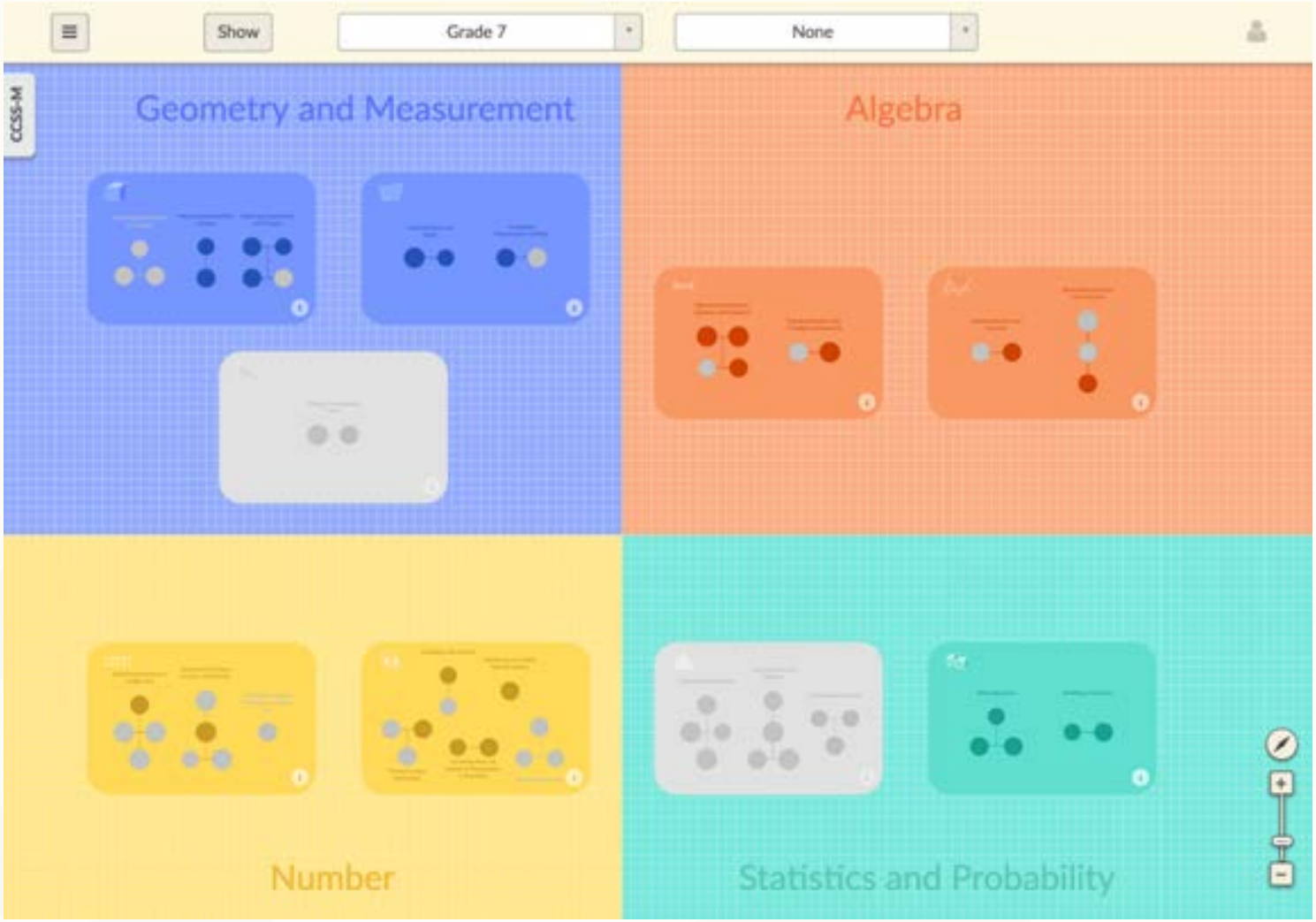
Demo of the SUDDS Learning Map

- Demonstration version of the map
- Production version of the map

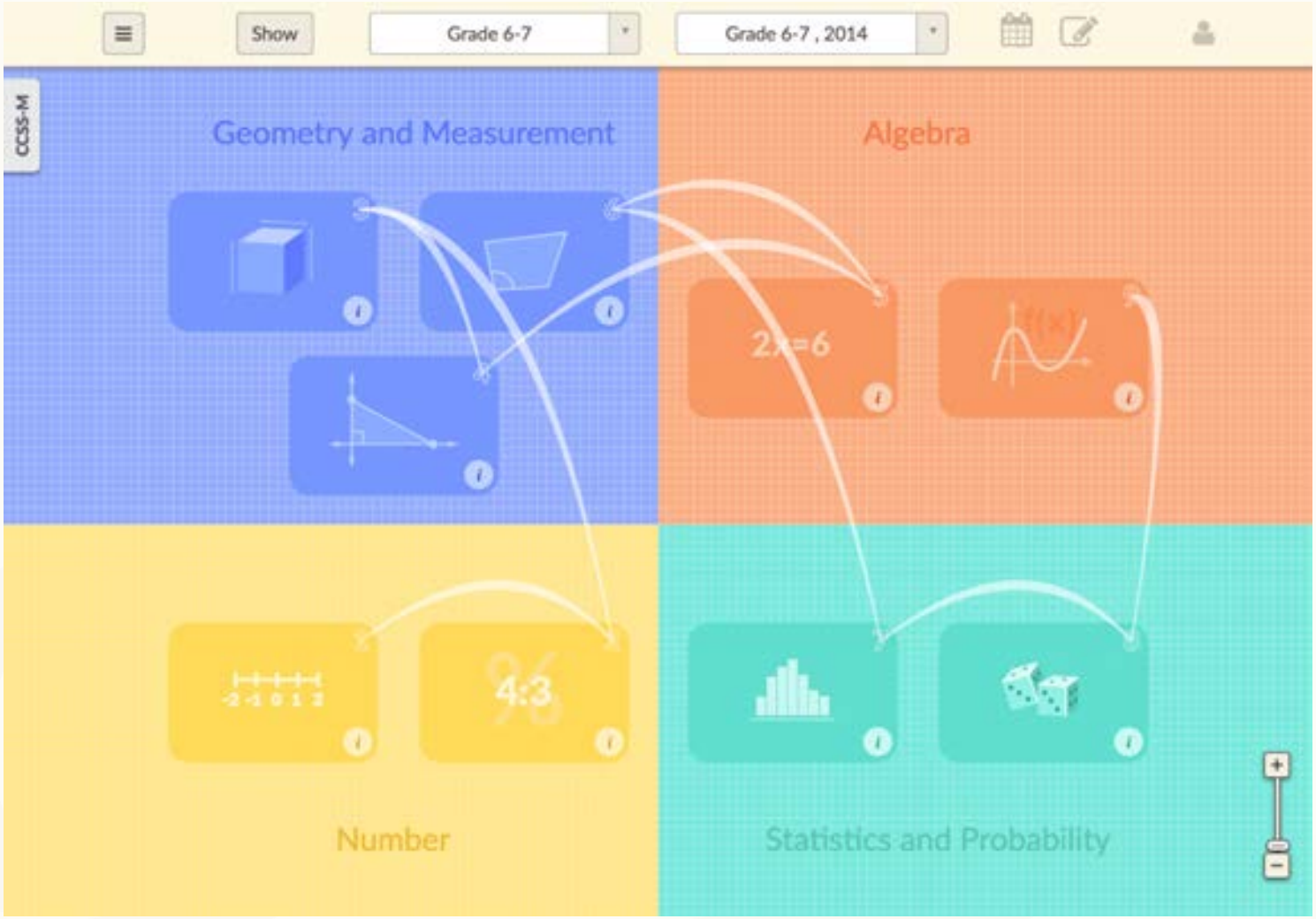
Middle School Mathematics in Nine Big Ideas



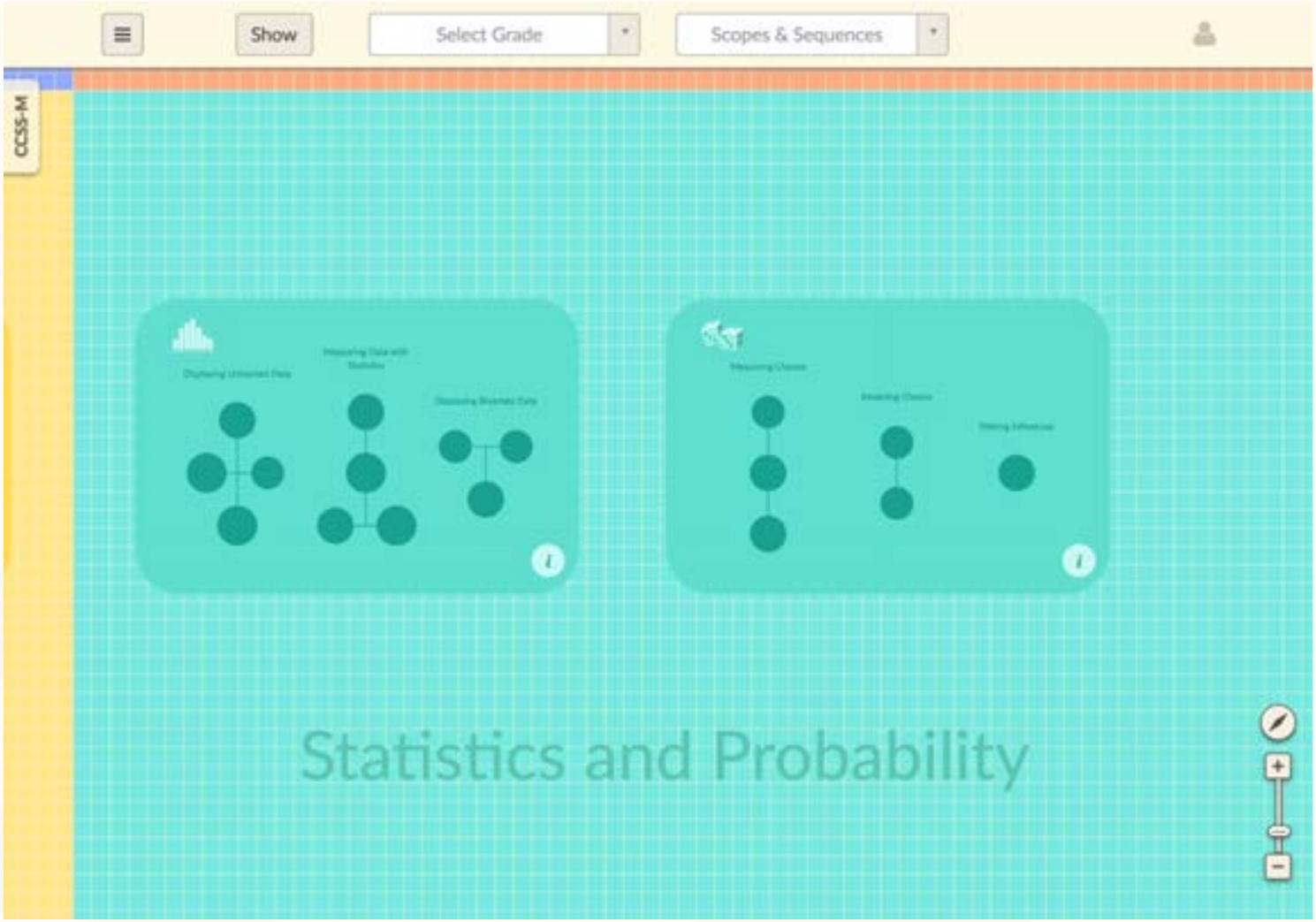
Apply Grade-Specific Filters



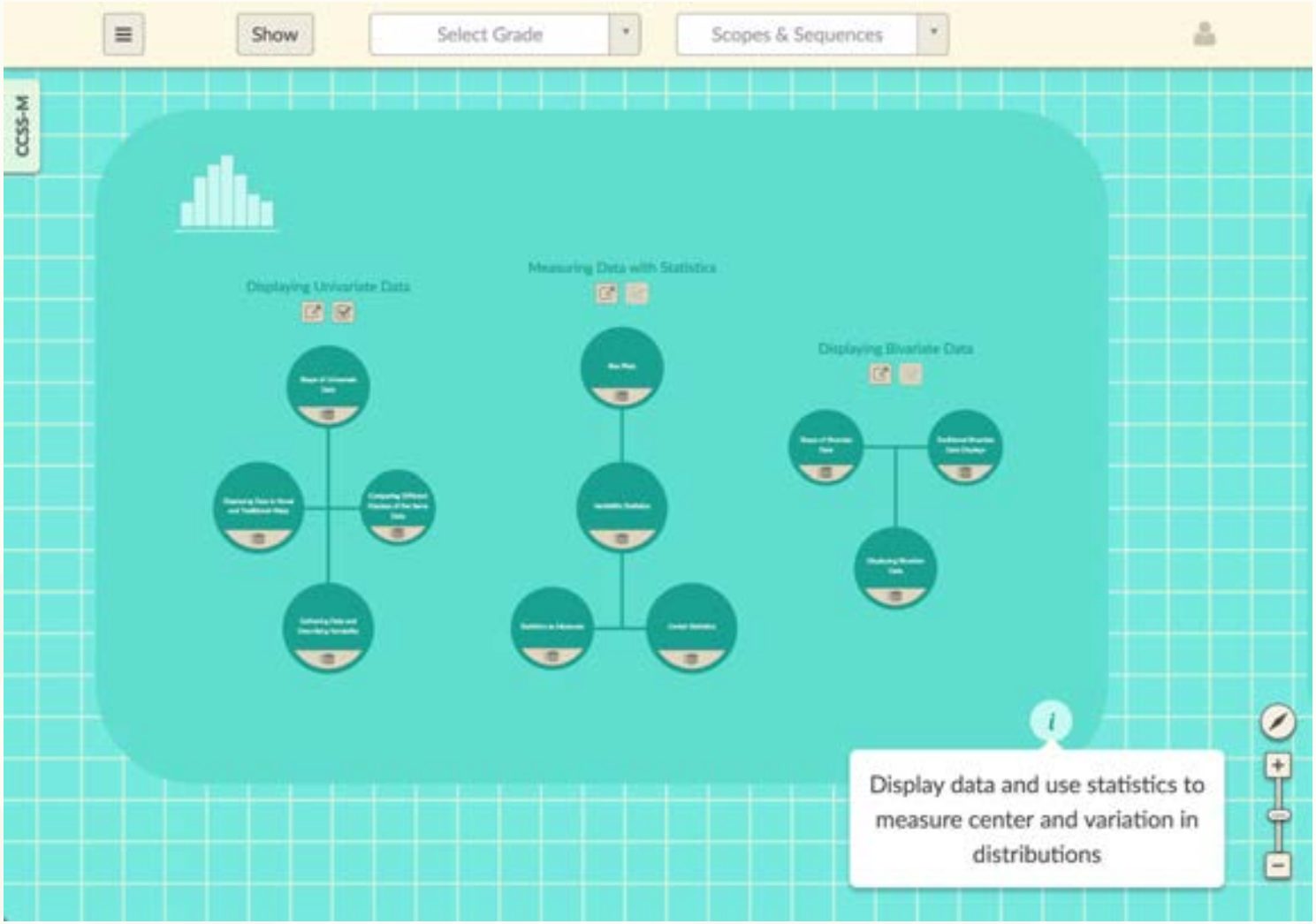
Apply Scopes and Sequences to Support Personalization



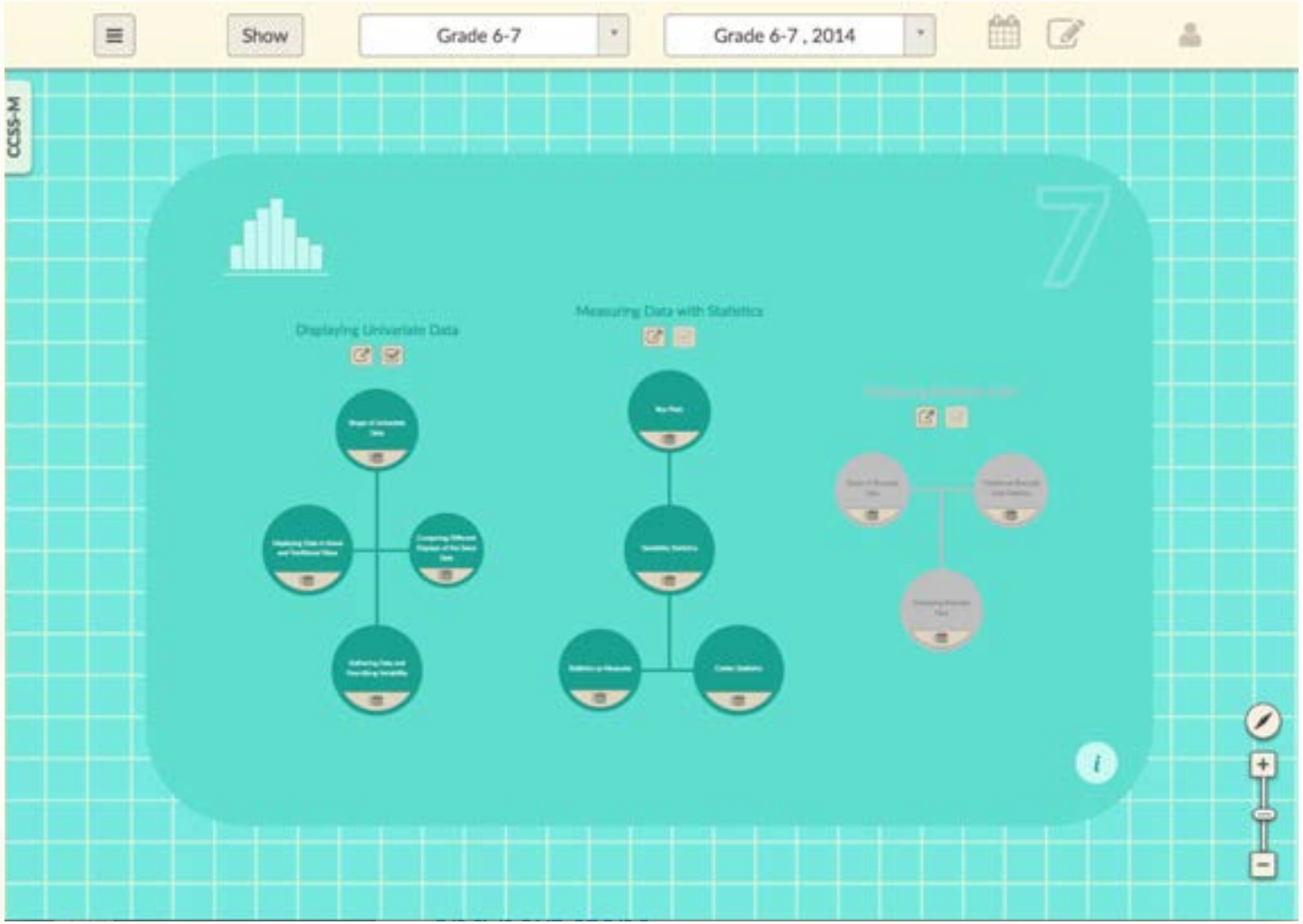
Zoom Into a Field



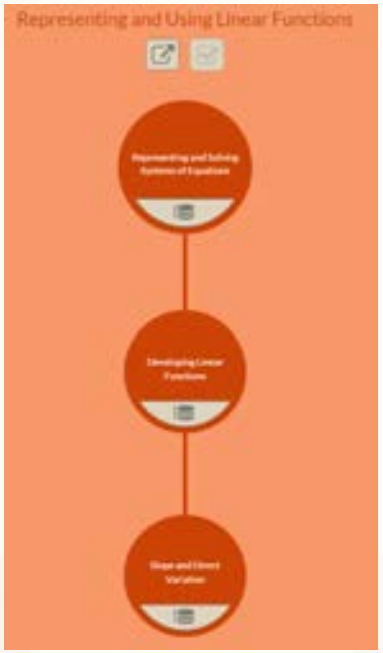
Zoom Into a Region



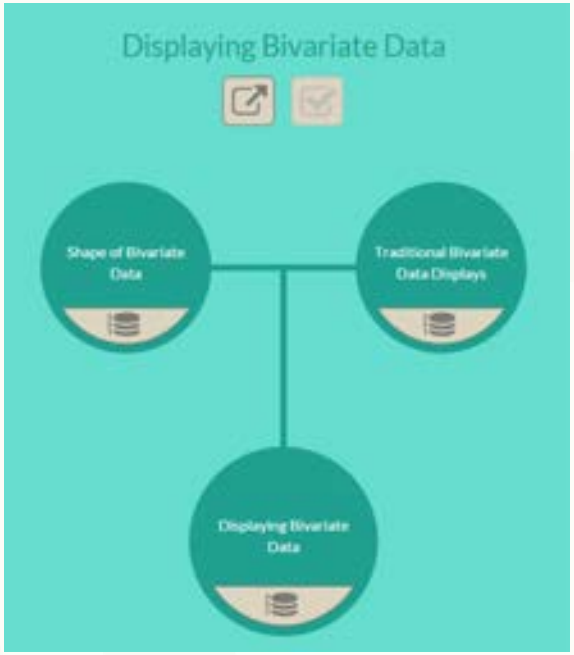
Filter Down to a Grade Range



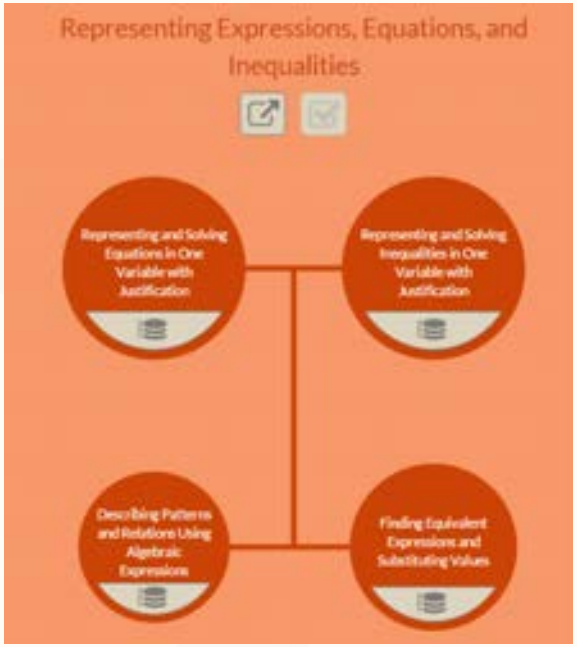
Shapes of Clusters Inform Possible Sequences



1 then 2 then 3
(linear path)

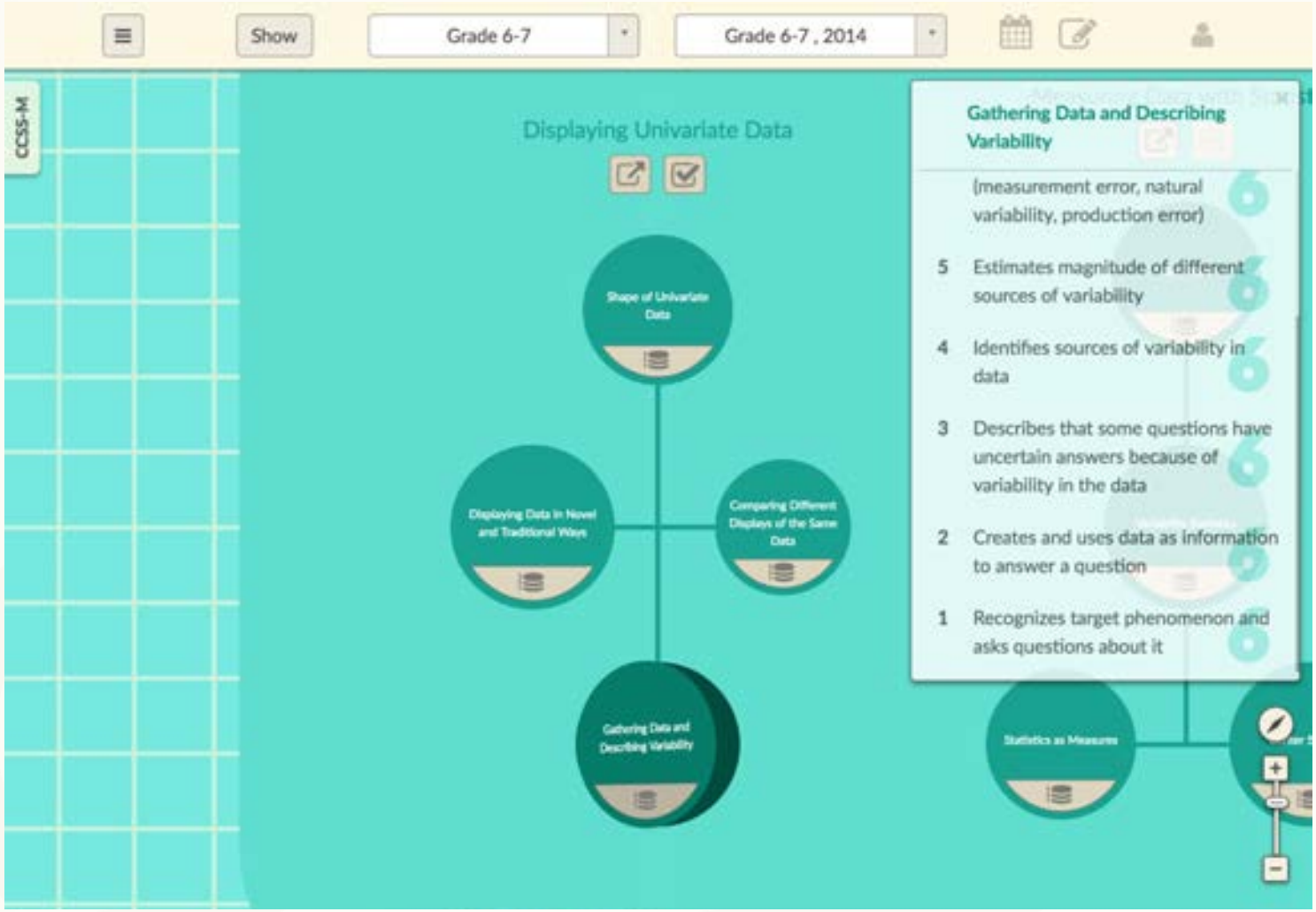


1 then (2,3) or (3,2)
(Divergent)

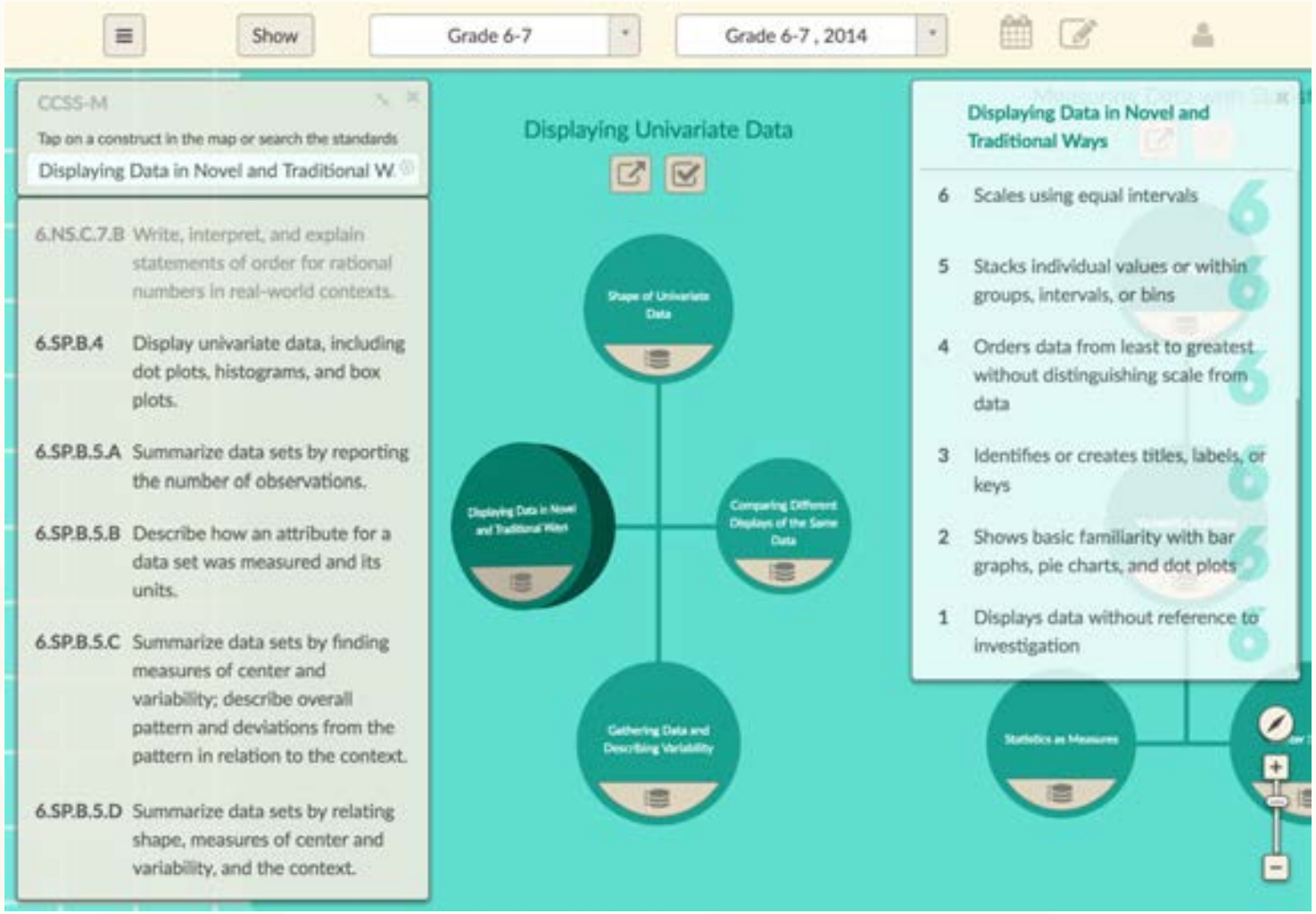


(1, 2) or (2,1) then (3,4) or (4,3)
(Convergent then Divergent)

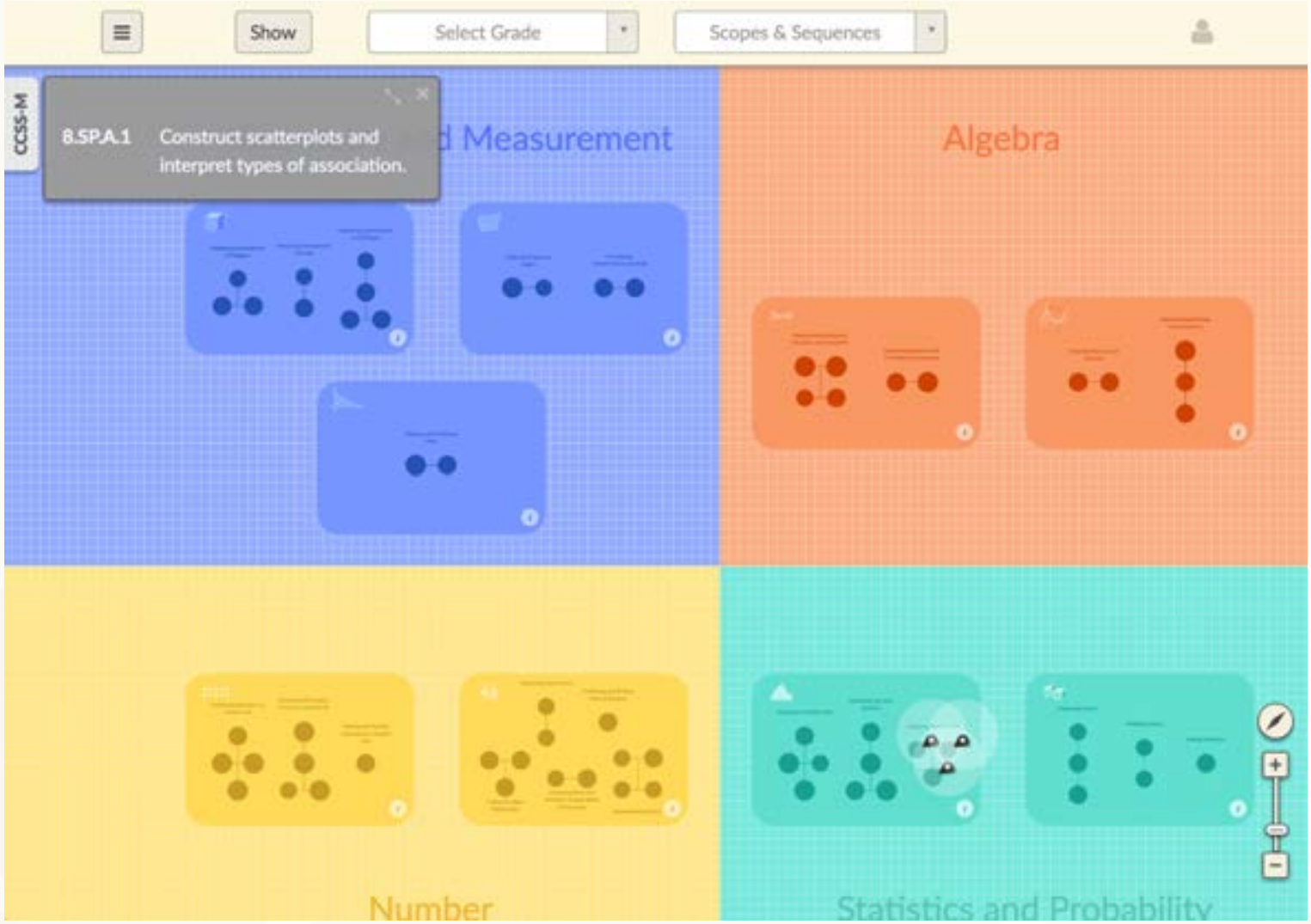
Open a Construct to Reveal its Learning Trajectory



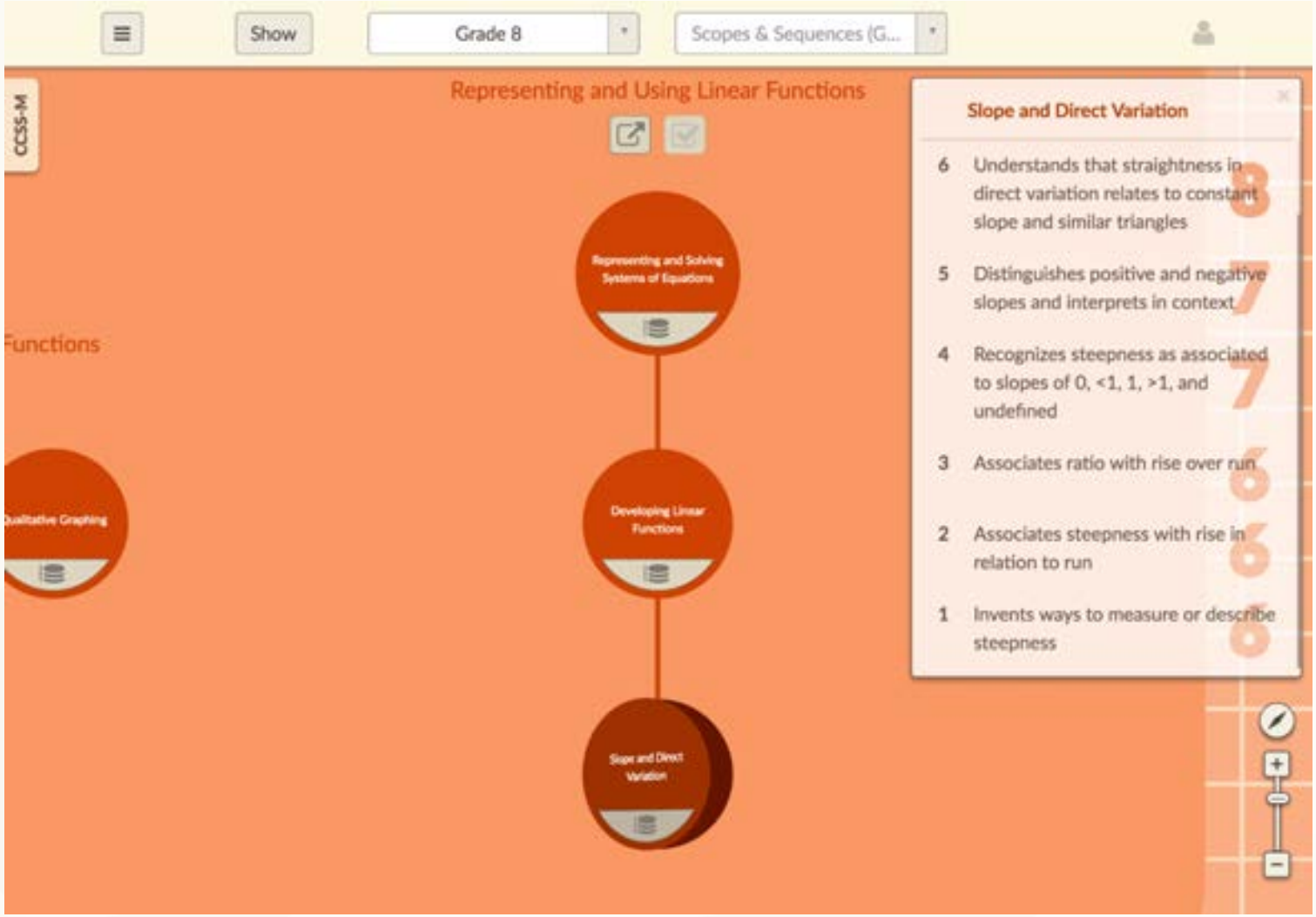
Find the Common Core Standards Linked to the Learning Trajectory



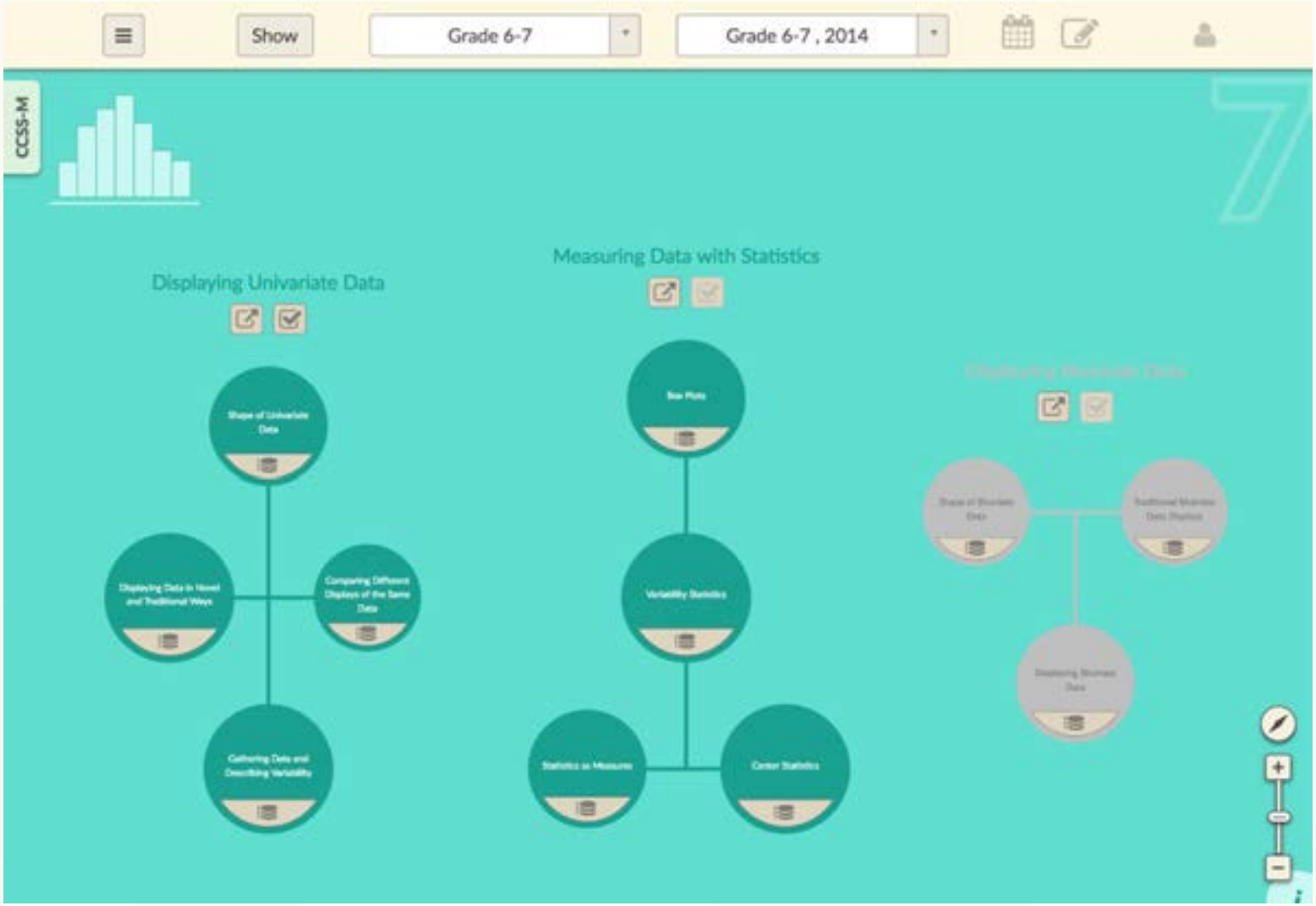
Find the Common Core Standards In the Big Ideas



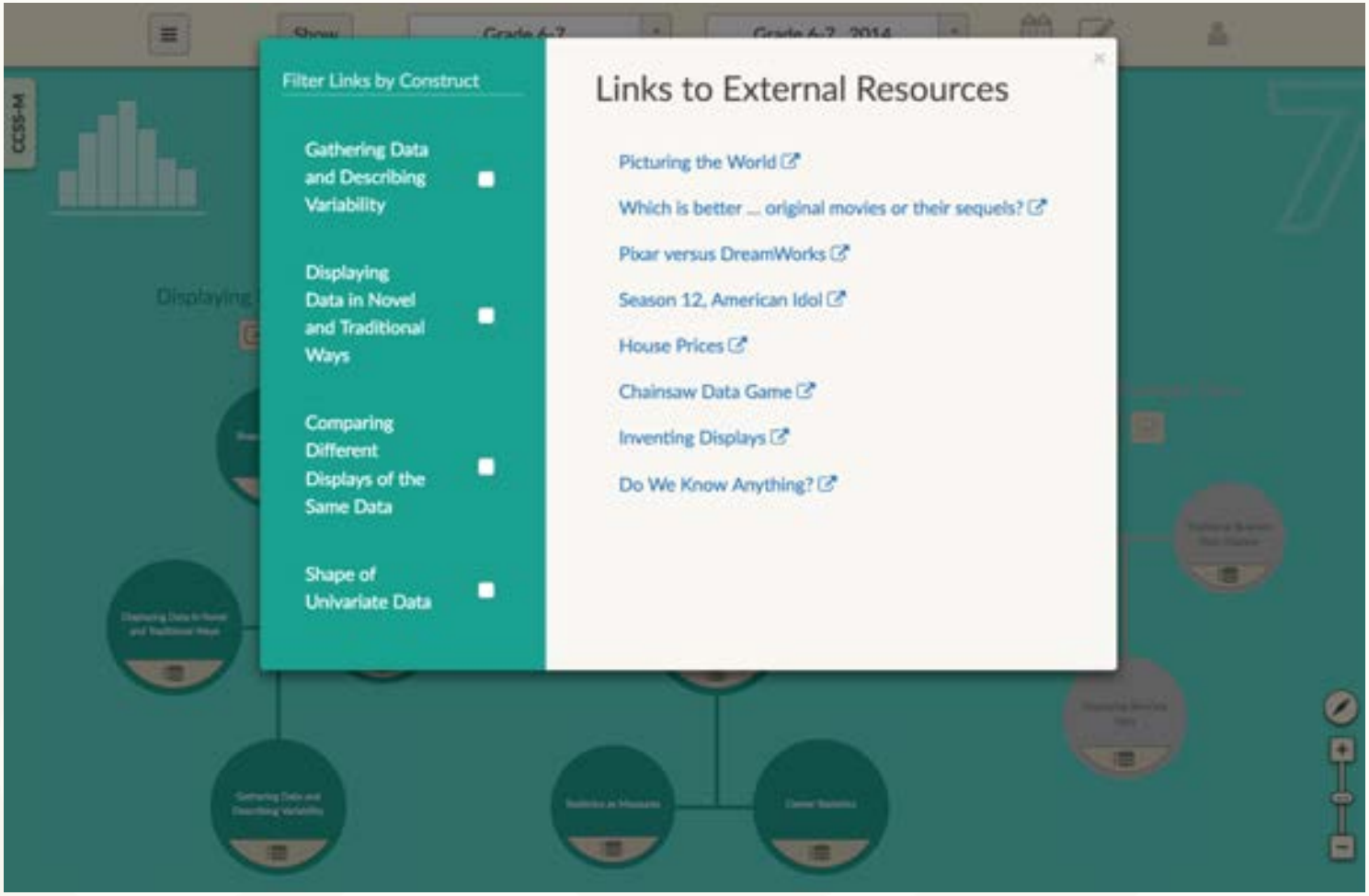
Move Below or Above Grade



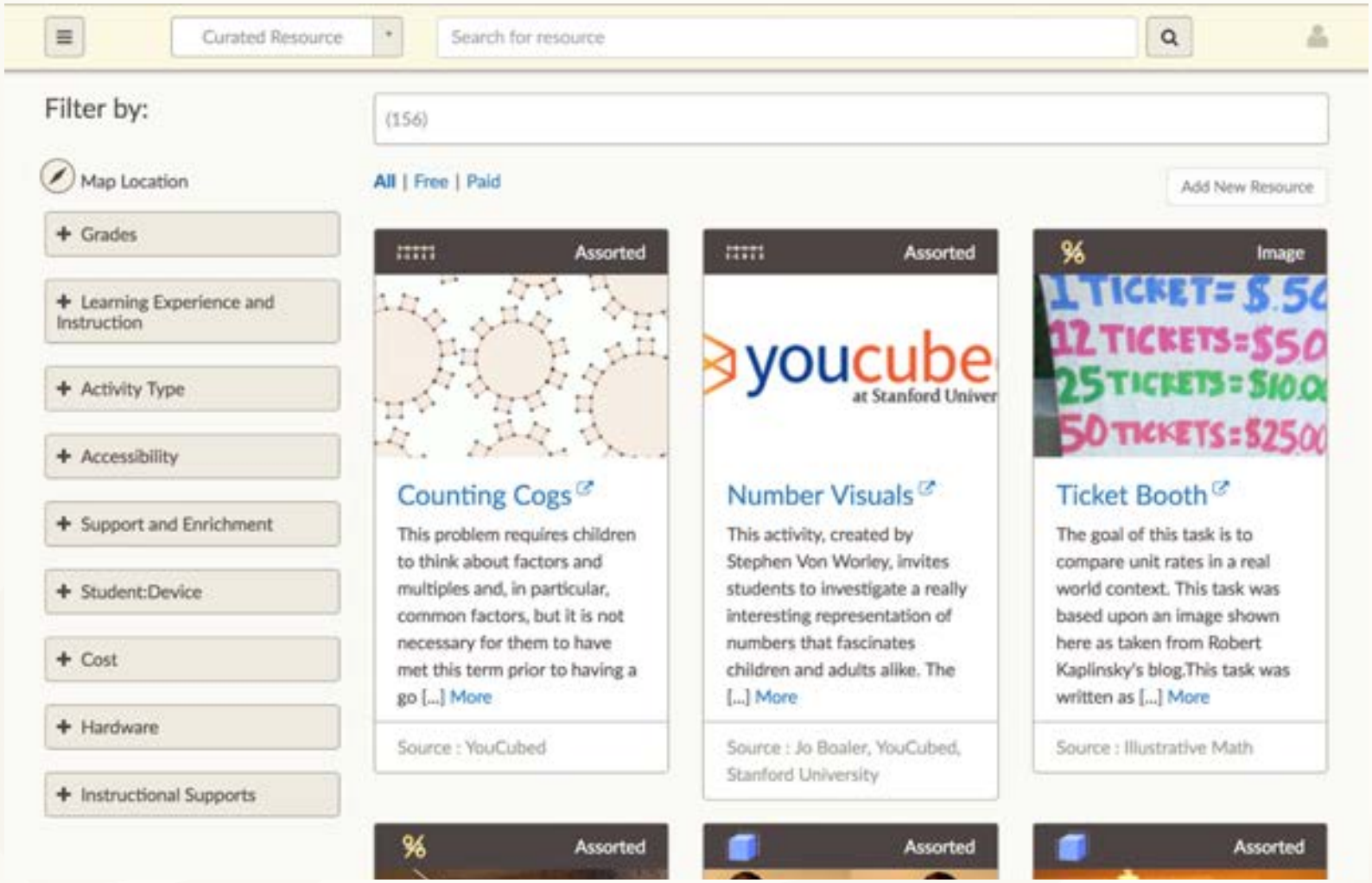
Find the Link to Resources



Select Aligned External Resources



Browse the Library of External Resources

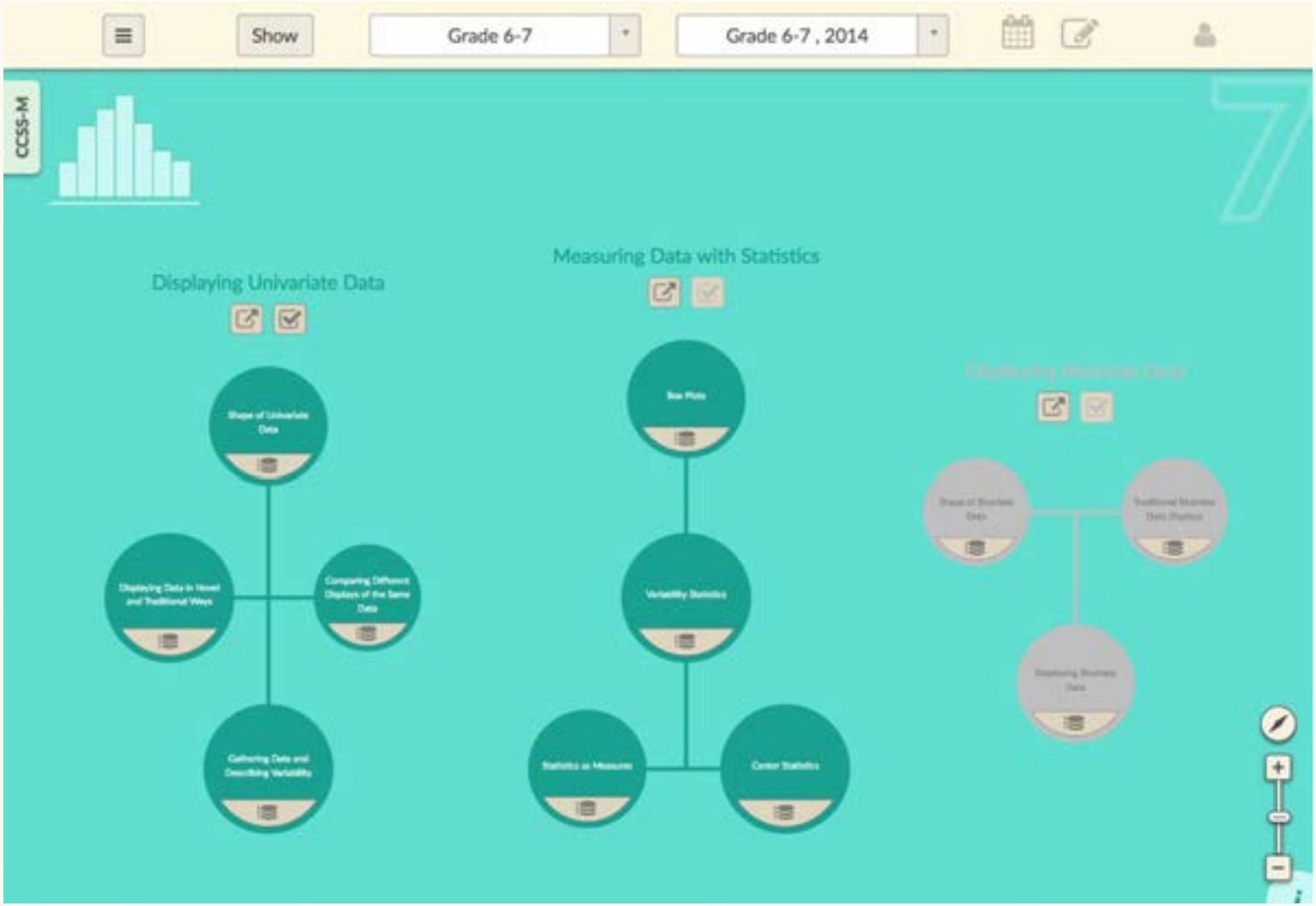


The SUDDS Learning Map

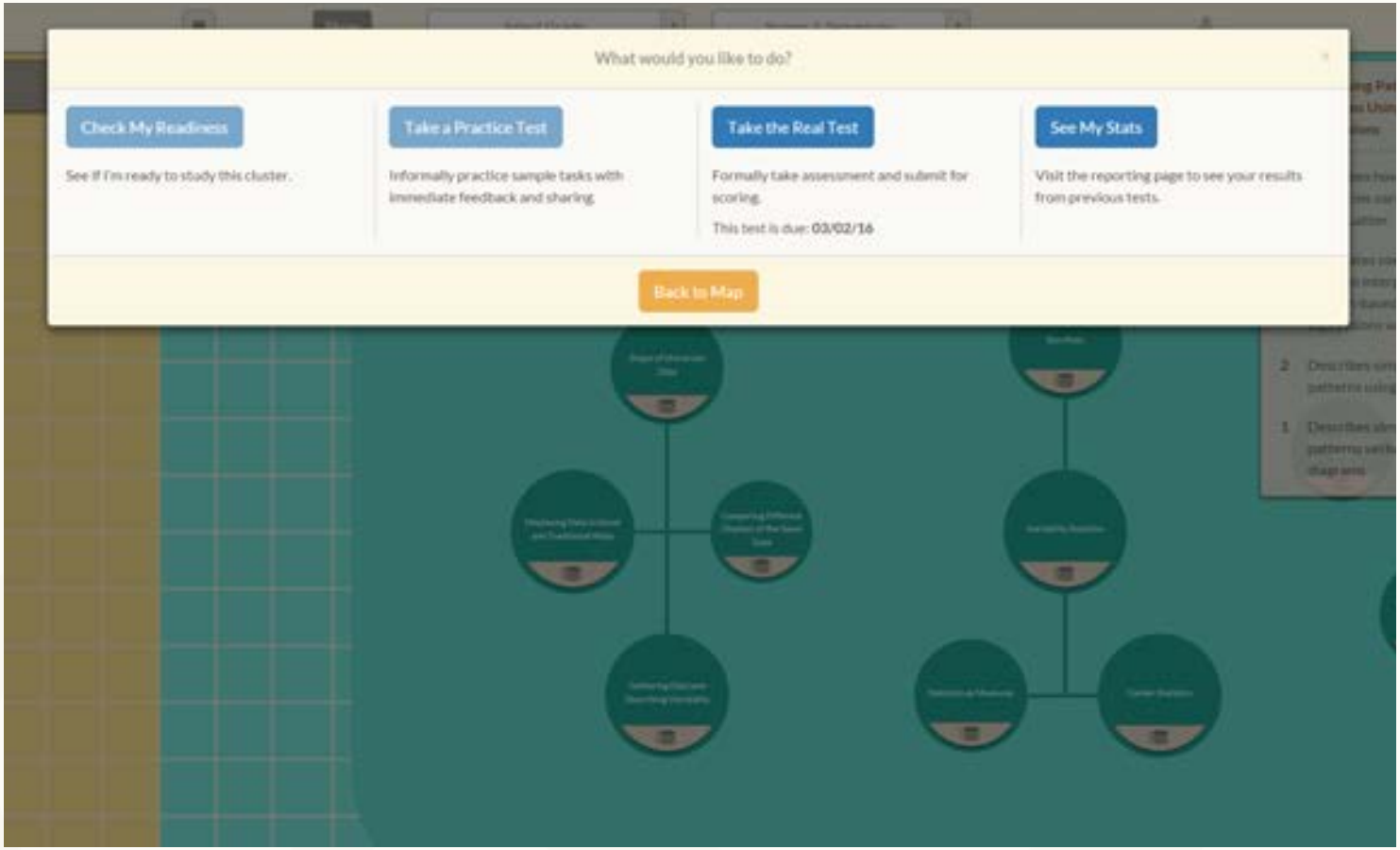
Our tool is an innovative digital container, where students and teachers can…

- Navigate the content of middle school mathematics around big ideas and research-based learning trajectories
- Select and sequence aligned open source curricular resources
- Assess, in real time, students' progress, identifying needs and next steps

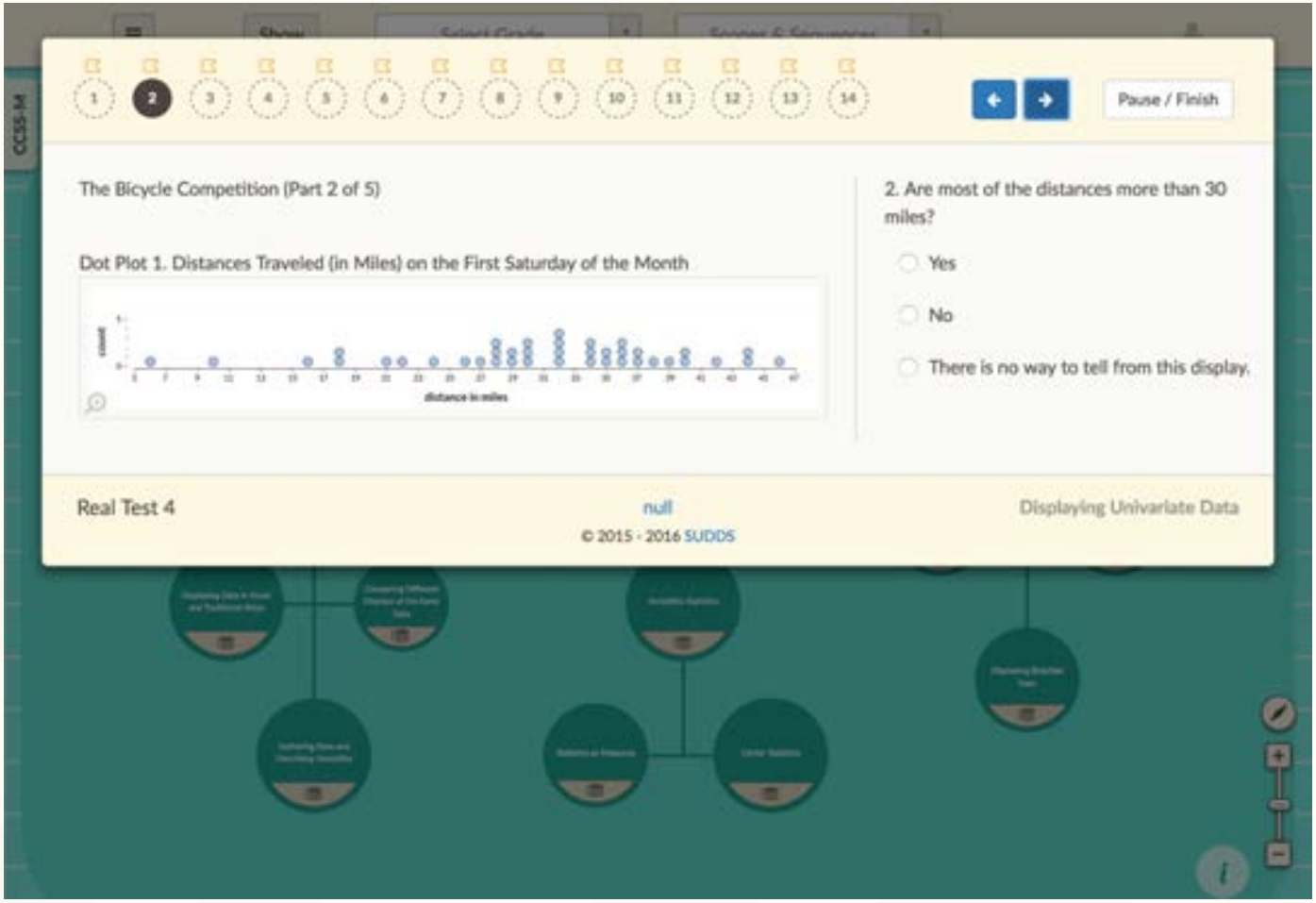
Find the Link to Assessments



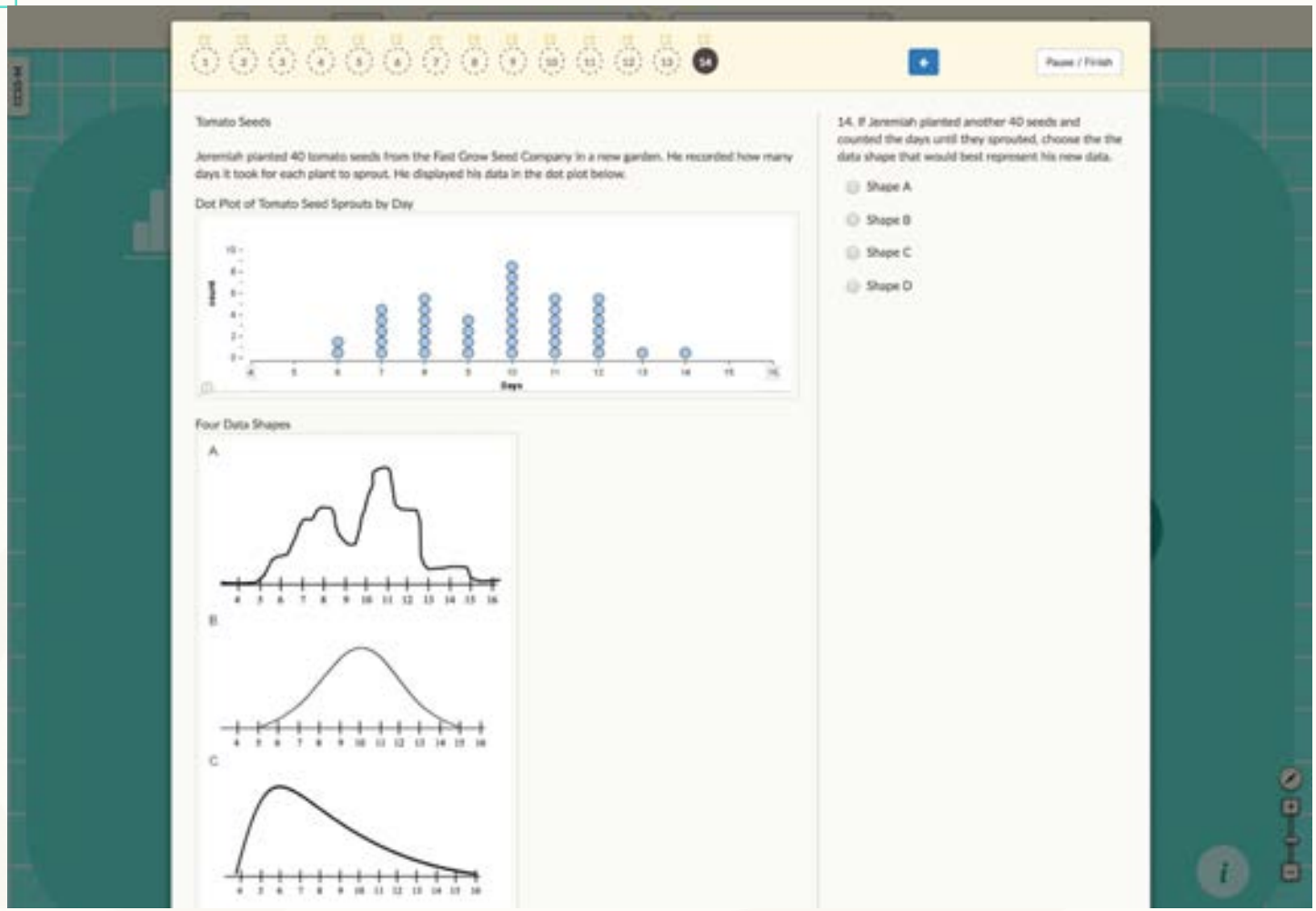
Take an Assessment



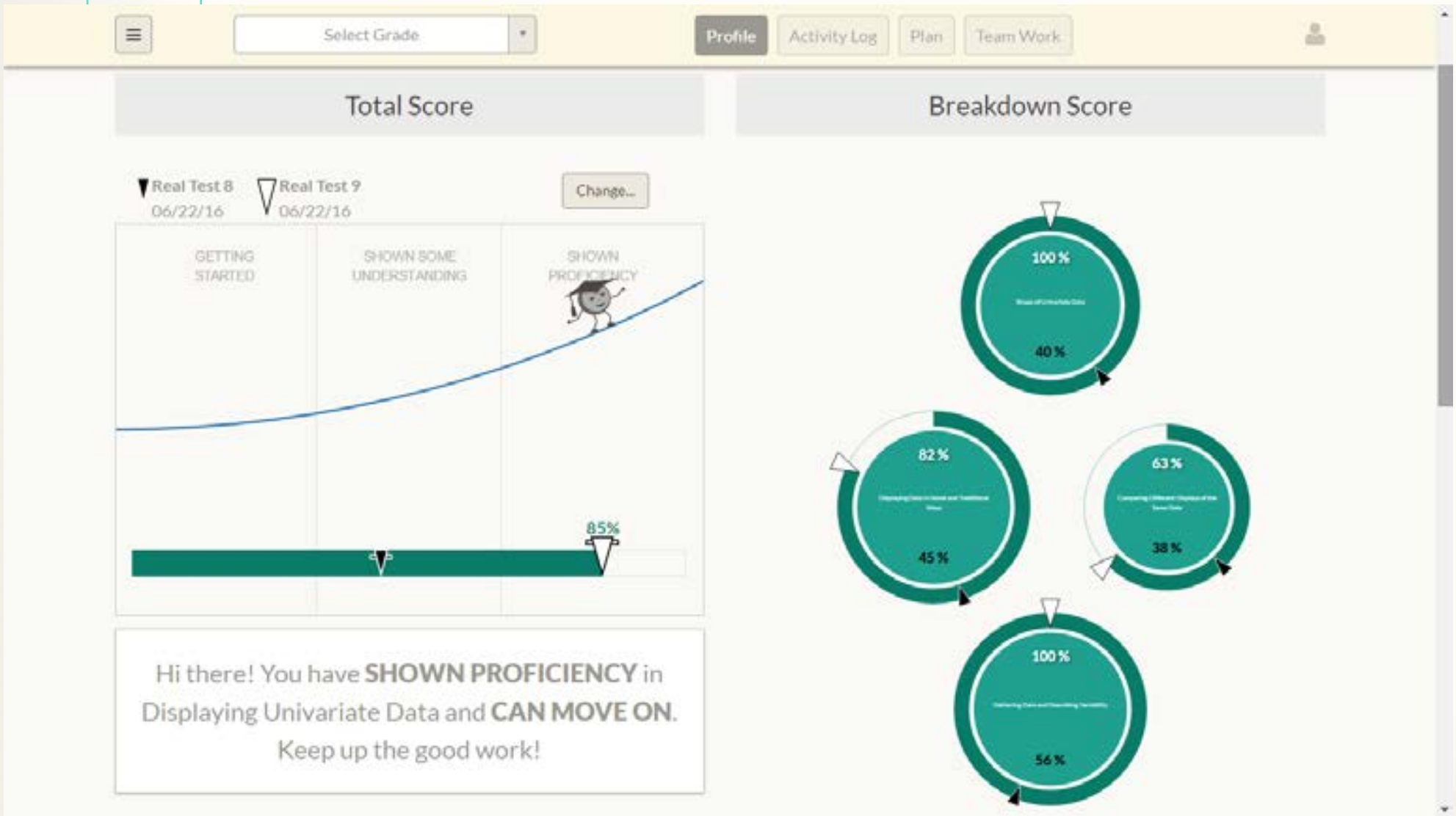
Interact with Items and Tasks in Fully Responsive Pages



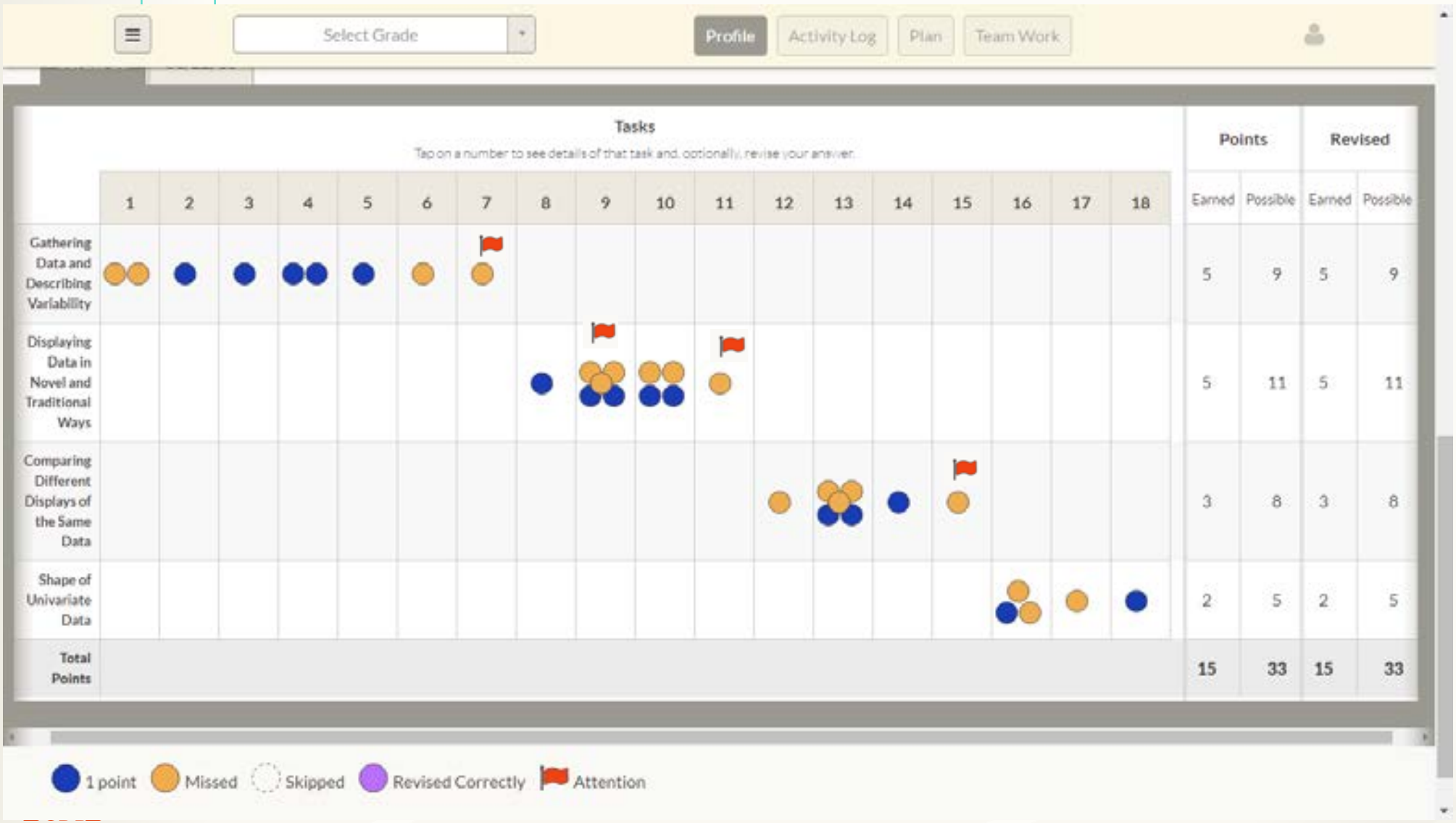
Tasks Designed to Maximize Learning, not Minimize Testing Time



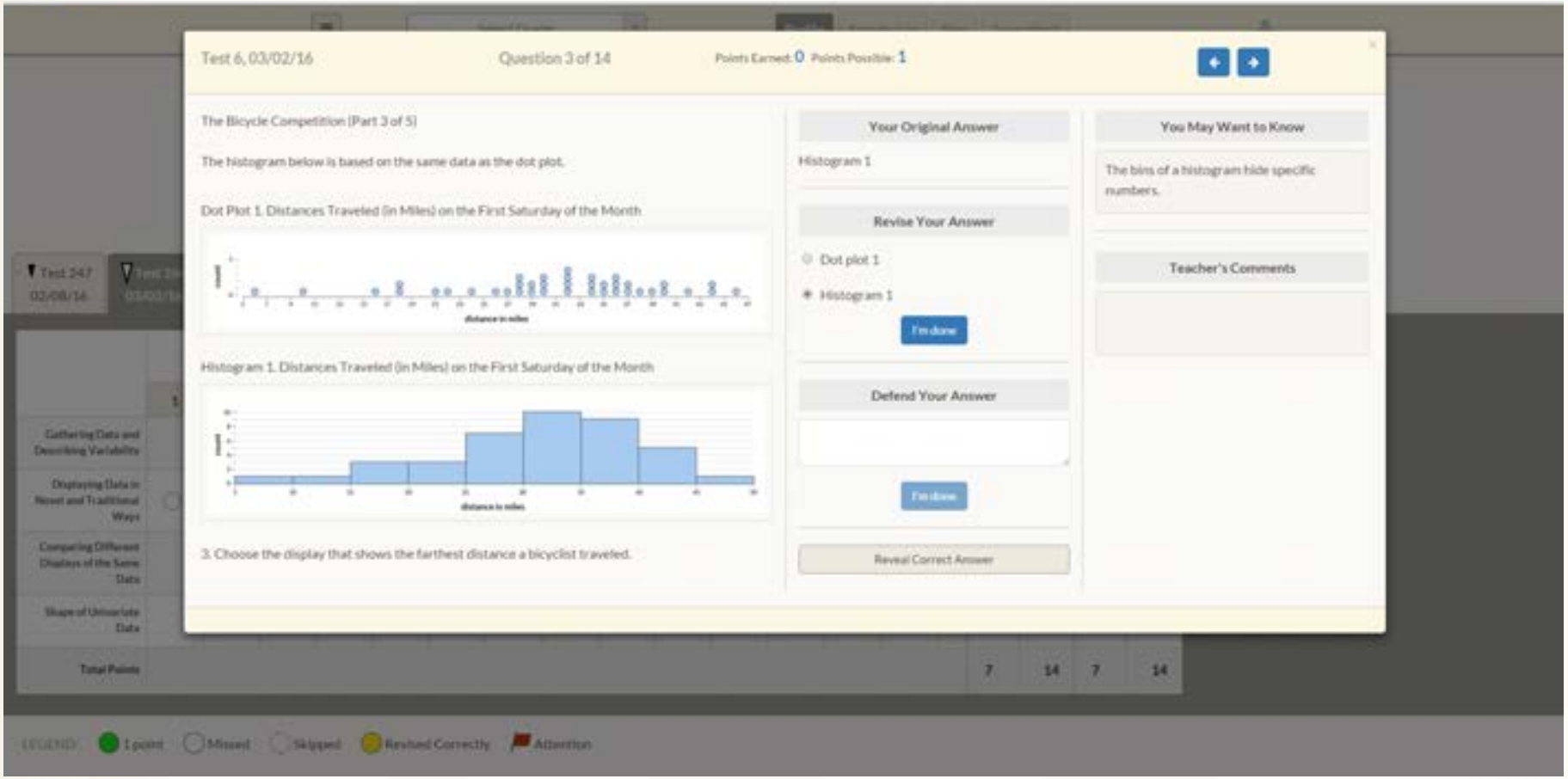
Messages Encourage a Growth-Mindset



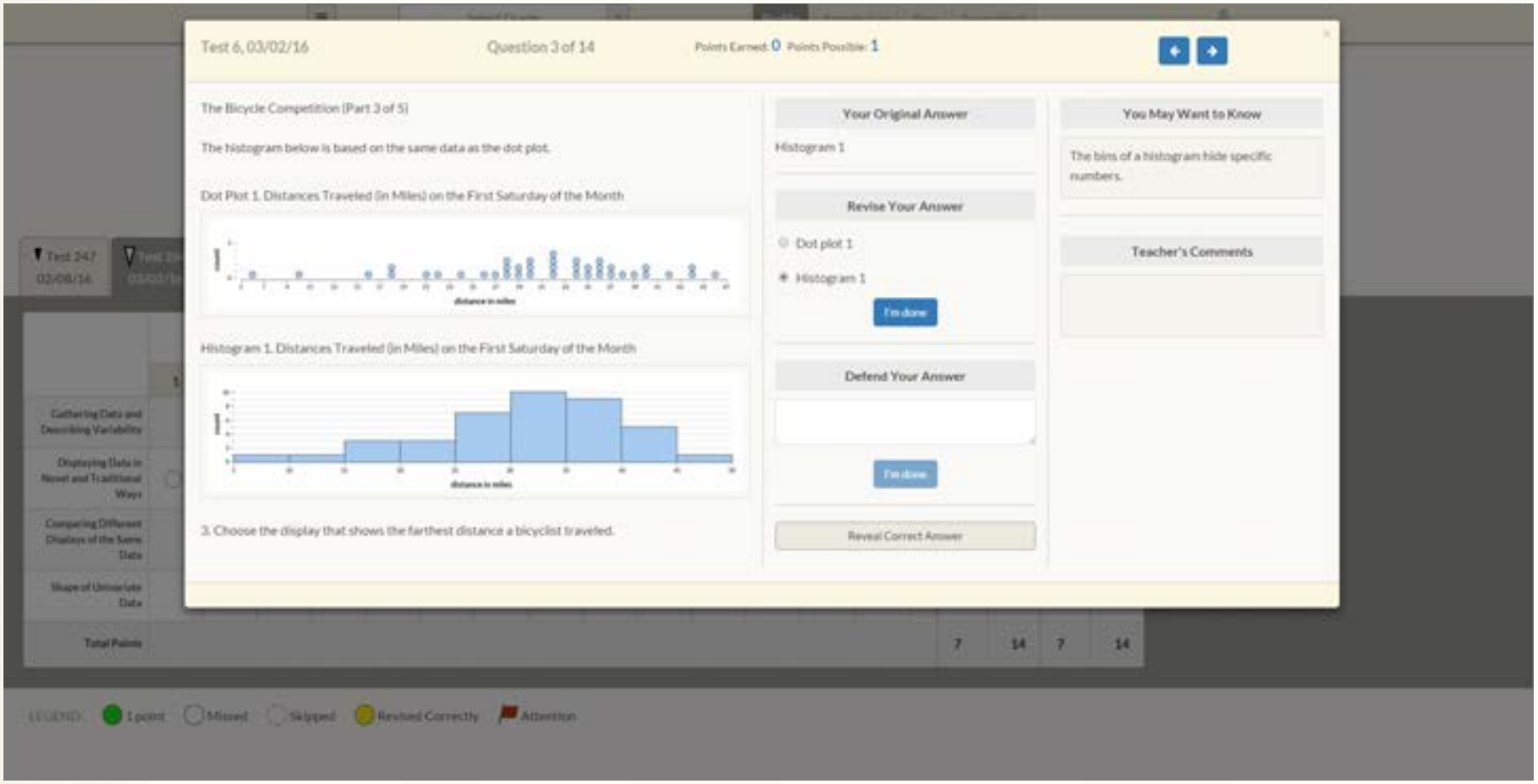
Visualize Score Patterns and See Misconceptions



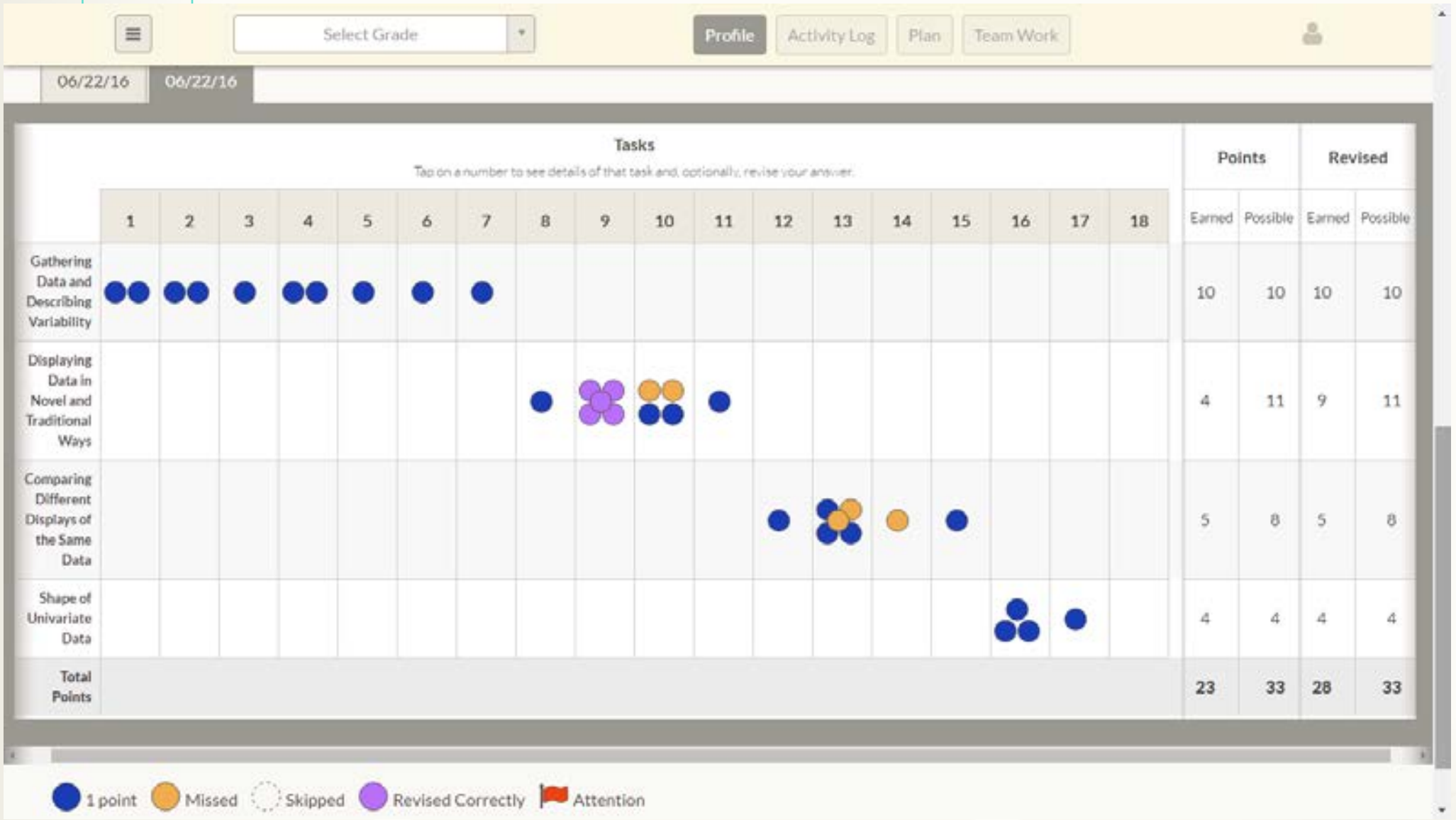
Reveal the Key, Revise, or Defend an Answer



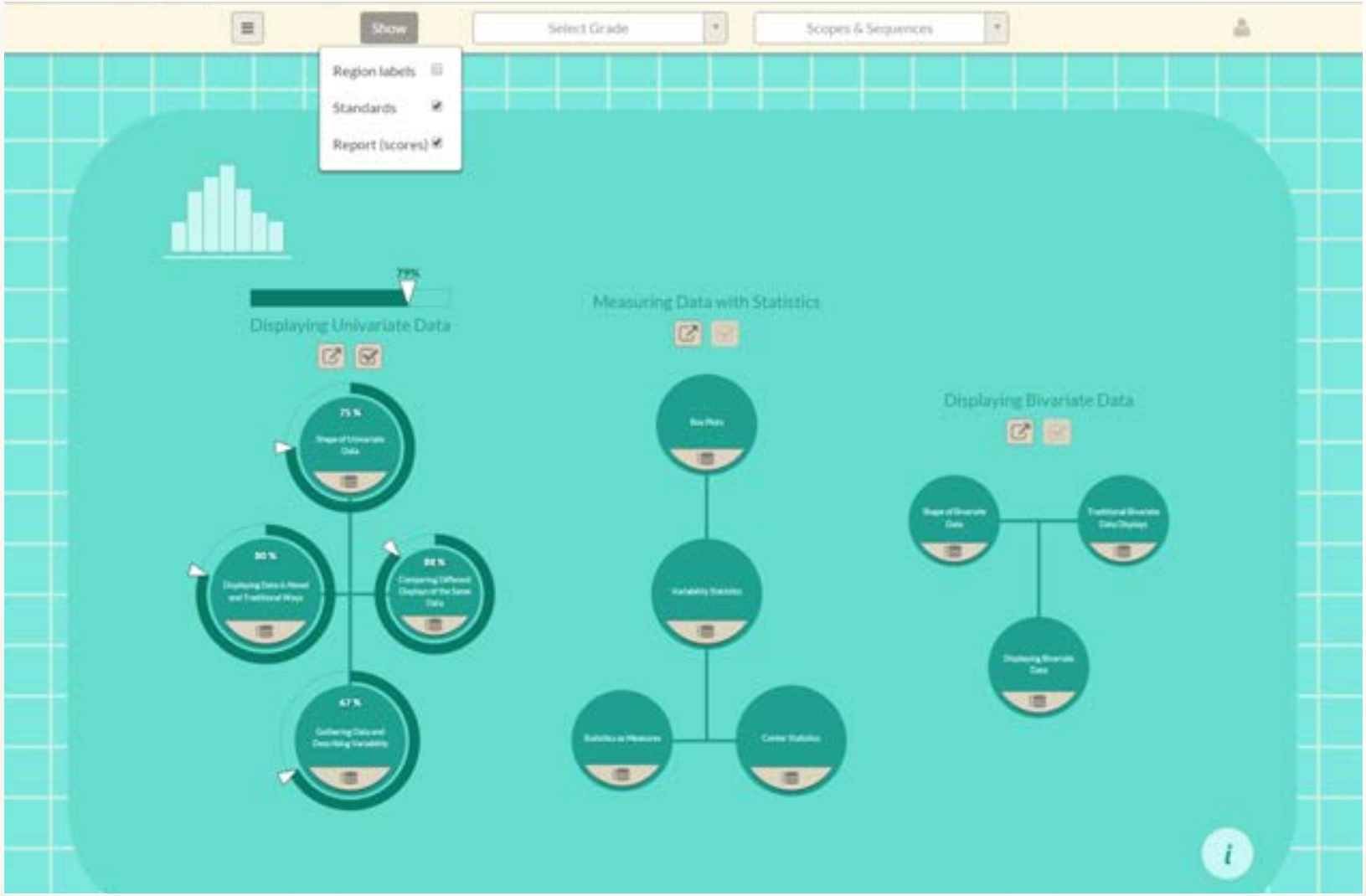
Give and Receive Commentary



See the Results of Your Revisions



Visualize Score Profiles Within the Map



Overview

- The SUDDS team
- How our DLS can help close the gap
- A demonstration of the Digital Learning System (DLS)
- A closer look at our assessments
- What we have learned
- New research directions and next steps

A Learning Map

Underlying learning trajectories

Elaboration document

Diagnostic Assessments

Pellegrino's Cognitive Assessment Framework*

- A model of student learning in the specific academic domain
  - Relational Learning Clusters
  - Learning Trajectories
- Defined expectations (hypotheses) about the kinds of observations that will provide evidence of competency
  - Elaboration documents
- A framework for interpreting the results of the assessment
  - Score reports built around Relational Learning Clusters
  - Placing students in the learning trajectories

*Page 22 of Cognitive Diagnostic Assessment for Education edited by Leighton and Gierl

Assessment Results for Instructional Guidance Needs to be:

- Timely
- Systematic for all students
- Accurate
- Relevant to what is being taught
- Informative on student progress
- Precise
- Can be taken multiple times

Diagnostic Assessments

- Practice Tests focus on each construct and its learning trajectory
- "Real Tests" focus on a Cluster to avoid over-testing
- Consist of 10 items and take about 30 minutes
- Are coordinated with curriculum and used formatively for instructional decision-making
- Are machine-generated and scored immediately
- Focus on conceptual issues of understanding

Diagnostic Assessments

- All items were field-tested at 3 middle schools in two districts
- District 1 is high SES and has two schools; District 2 has one school participating enrolling both medium-high and low SES students
- A class received the same problems for an RLC in field-testing, hence not all levels were given during one assessment

Tasks Were Designed to be Diagnostic

Tasks and items were designed to:

- differentiate between low, intermediate, and pro levels of performance as defined by the Learning Trajectory
- Measure progress along the LT
- Flag misconceptions and systematic errors
- Be consistent with an elaboration document

Elaboration Document

Cluster: Comparing Ratio and Solving for Missing Values In Proportions

Construct: Finding Missing Values in Proportions

Construct Summary

Students solve missing value problems using different methods. They build up using base or unit ratios. When the corresponding value for the missing value is a multiple of the base ratio, they multiply both quantities by the same factor. For large missing values, they use multiplication or division to find a nearby value and use addition and subtraction of base or unit ratios to interpolate. They apply multiplication and division to solve 2 × 2 ratio tables with whole numbers and fractional entries. They distinguish between proportional- and non-proportional relationships and identify the constant of proportionality in tables, graphs and equations.

Level 3

Builds to equivalent ratios by multiplying quantities by n and adjusts by adding or subtracting base and unit ratios

Grade: 6
Examples/Cases: Students build up to ratios with larger numbers (greater than 100)
Misconceptions: Uses multiplication to get close and then adds or subtracts the same amount to or from both quantities
Items at this level:
2.05.13.03.02 (aligned, flagged for some by some groups, may save time if we can get fill-ins on the table itself.)
2.05.13.03.01 (aligned, need to change context or numbers in the item)

Level 4

Finds missing value in a ratio box (2 × 2 ratio table) using whole number multiplication or division of both values horizontally and vertically

Grade: 6
Examples/Cases: where known values are whole numbers and unknown values can be whole numbers or fractions
2 × 2 Ratio tables with whole number multiplication or division both horizontally and vertically, 2 × 2 ratio tables with whole number multiplication or division in only one direction (only horizontally or only vertically)
Misconceptions:
Items at this level:
2.05.13.04.01 (aligned)
Need items with whole number unknown values

Level 1

Builds up to a specific value of one quantity by incrementally adding base ratios or unit ratios

Grade: 6
Examples/Cases: Enter examples of student thinking, or important cases this level is meant to include.
Misconceptions: Adding the same amount to both quantities, keeps the ratio the equivalent
Items at this level:
2.05.13.01.01 (aligned, flagged for time by one group)
2.05.13.01.02
Need another item that involves unit ratios

Level 2

Builds equivalent ratios from base ratios or unit ratios by multiplying by n

Grade: 6
Examples/Cases: Enter examples of student thinking, or important cases this level is meant to include.
Misconceptions: Adding the same amount to both quantities, keeps the ratio the equivalent
Items at this level:
2.05.13.02.01 (aligned)
2.05.13.02.02 (aligned)
Need another item that involves unit ratios

Level 1 Item

Builds up to a specific value from by incrementally adding base ratios or unit ratios

Rohan mixes a particular shade of orange paint by adding one gallon of yellow paint and three gallons of red paint. The job requires five gallons of yellow paint, so he made a table to help him determine how much red paint he needs to make the same shade of orange paint.

Gallons of yellow paint	Gallons of red paint
1	3
2	
3	
4	
5	

Use the table to the left to complete the following statements.

For two gallons of yellow paint, he needs ____ gallons of red paint.

Enter a number

For three gallons of yellow paint, he needs ____ gallons of red paint.

Enter a number

For four gallons of yellow paint, he needs ____ gallons of red paint.

Enter a number

For five gallons of yellow paint, he needs ____ gallons of red paint.

Enter a number

Sample Scoring of a Misconception

		Response Samples by Score Category			
Yellow paint	Red Paint	Full Credit			
1	***3***				
2	?	6			
3	?	9			
4	?	12			
5	?	15			
	% Full credit	***74%***			

Level 1
Builds up to a specific value of one quantity by incrementally adding base ratios or unit ratios

Grade: 6
Examples/Cases: Enter examples of student thinking, or important cases this level is meant to include.
Misconceptions: Adding the same amount to both quantities, keeps the ratio the equivalent

Sample Scoring of a Misconception

		Response Samples by Score Category (Errors in red)					
Yellow paint	Red Paint	Full Credit	Partial Credit				
1	***3***						
2	?	6	3,	6,	6,…		
3	?	9	9,	9,	12,…		
4	?	12	12,	4,	24,…		
5	?	15	15,	20,	48,…		
		74%	***5%***				

Level 1

Builds up to a specific value of one quantity by incrementally adding base ratios or unit ratios

Grade: 6

Examples/Cases: Enter examples of student thinking, or important cases this level is meant to include.

Misconceptions: Adding the same amount to both quantities, keeps the ratio the equivalent

Sample Scoring of a Misconception

		Response Samples by Score Category (Errors in red)						
Yellow paint	Red Paint	Full Credit	Partial Credit	Misconception				Other Incorrect
1	***3***	***74%***	***5%***	***12%***				***9%***
2	?	6	3, 6, 6,…	4,	5,	5 ,	9,…	
3	?	9	9, 9, 12,…	5,	7,	6 ,	15,…	
4	?	12	12,4, 24,…	6,	9,	7,	21,…	
5	?	15	15,20,48,…	7,	11,	8,	27,…	

Level 1

Builds up to a specific value of one quantity by incrementally adding base ratios or unit ratios

Grade: 6

Examples/Cases: Enter examples of student thinking, or important cases this level is meant to include.

Misconceptions: Adding the same amount to both quantities, keeps the ratio the equivalent

Contrast a Misconception with the Correct Conception

Misconception:

Misconception:

Adding the same amount to both quantities, keeps the ratio equivalent.

Conception (level 4):
When the corresponding value for the missing value is a multiple of the base ratio, multiply both quantities by the same factor.

Conception (level 1):

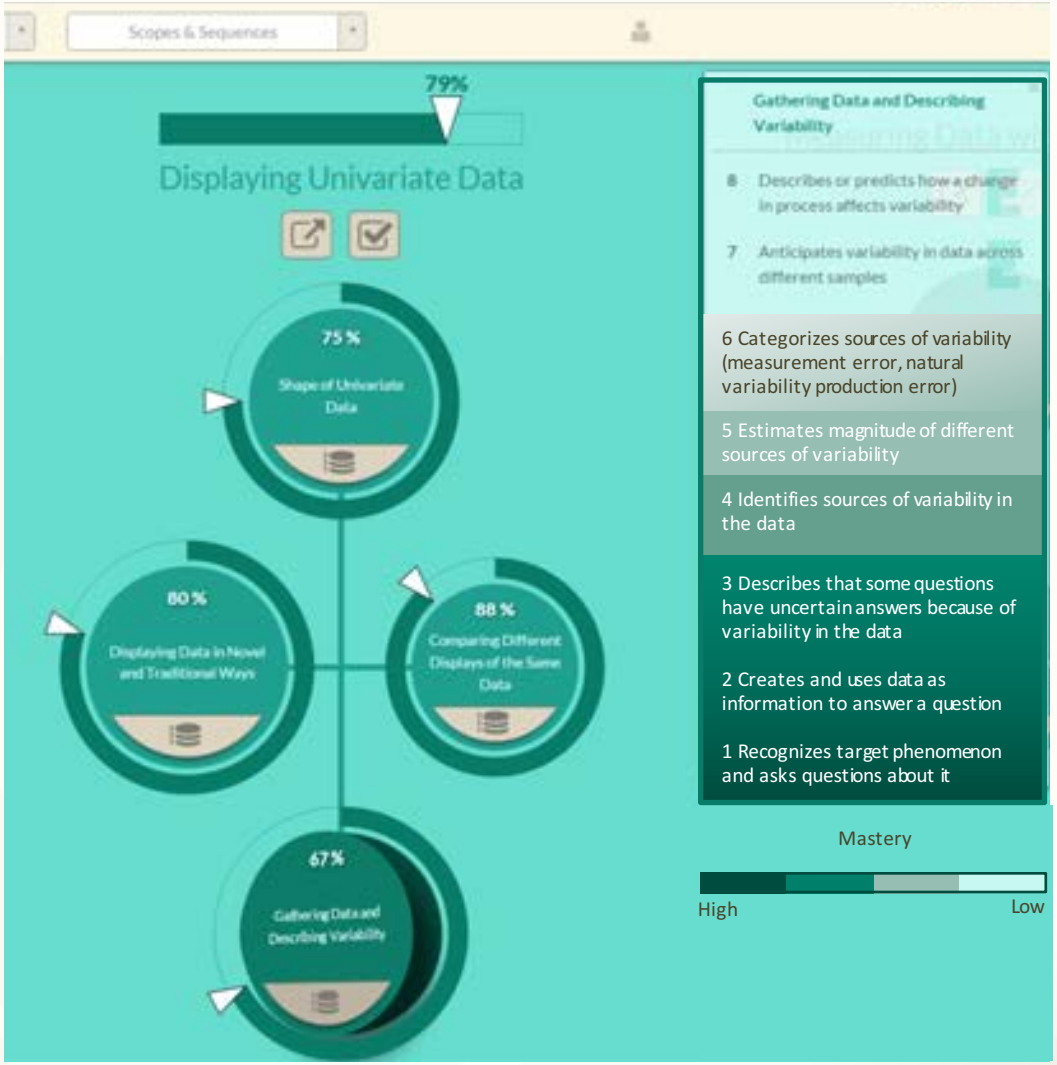
Build up to a specific value of one quantity by incrementally adding base ratios or unit ratios.

Assessment Design for Learning Trajectories

- A cluster of learning trajectories (LT) provides the test blueprint
- The LT replaces a cognitive taxonomy
- Some LTs span multiple grade levels
- LTs within a cluster are highly related.
- Constructs may relate to one another in multiple ways: linear, divergent, convergent, or a mixture

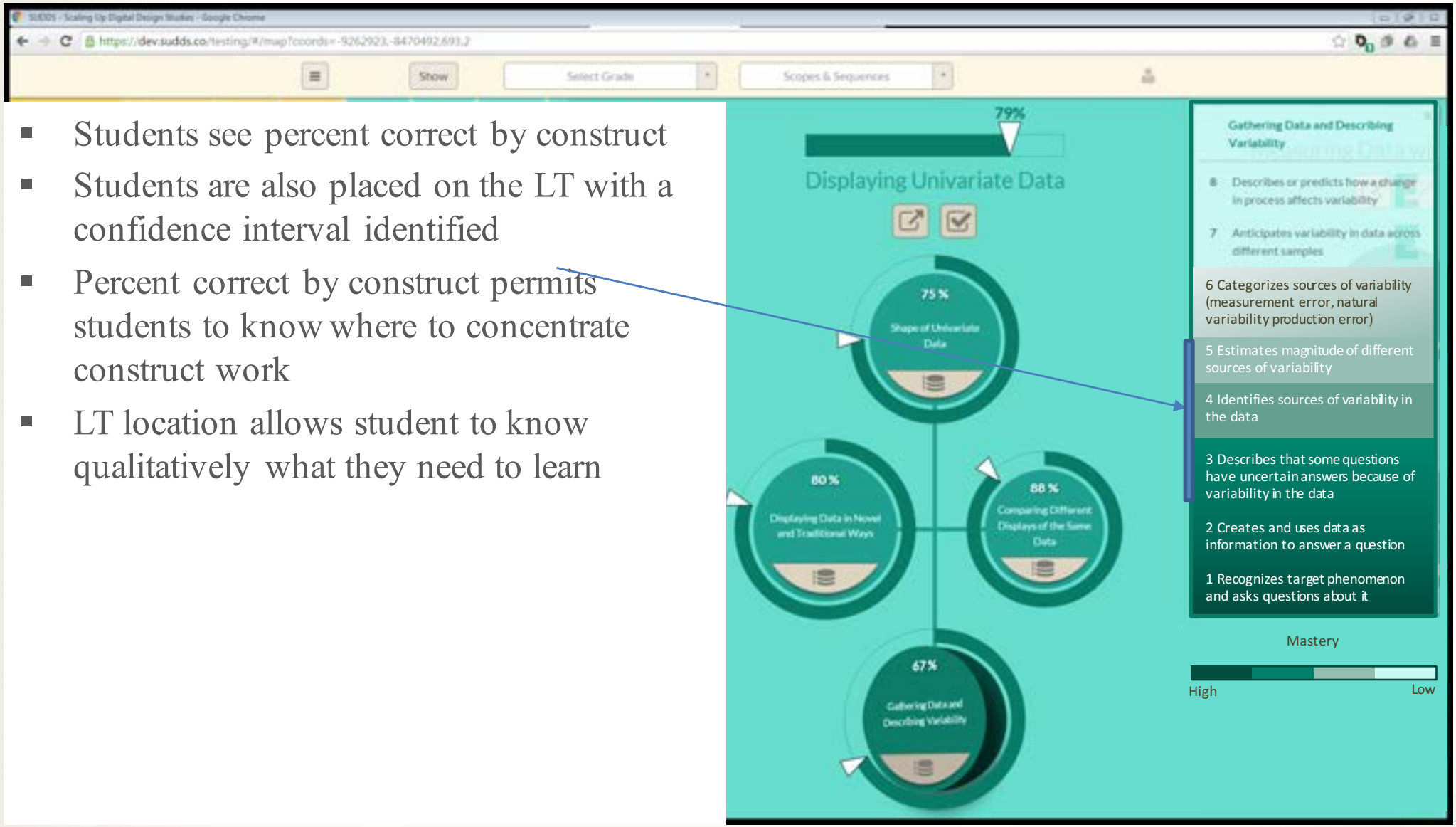
Locating Students in the Learning Trajectories

- LTs give a foundation on which we can place our IRT models
- Data modeling is confirmatory in nature
- Theories can be disconfirmed, theories can be improved, and retested through a research agenda
- The LT provides an interpretive framework for scores



Locating Students in the Learning Trajectories

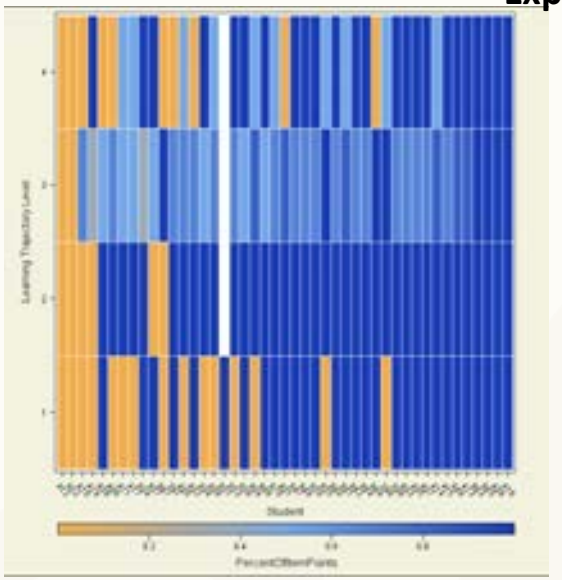
- Students see percent correct by construct
- Students are also placed on the LT with a confidence interval identified
- Percent correct by construct permits students to know where to concentrate construct work
- LT location allows student to know qualitatively what they need to learn

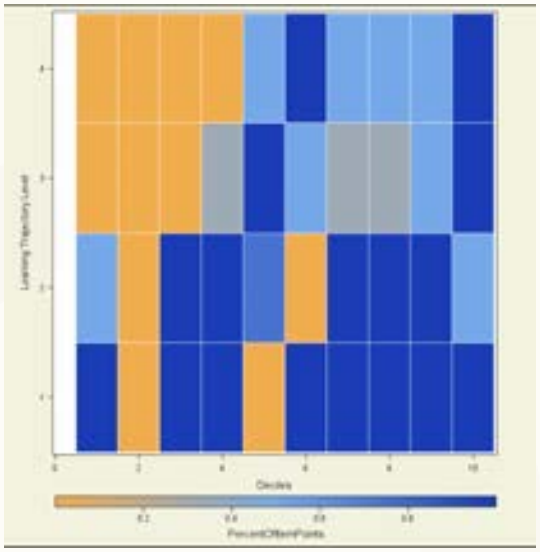


Locating Groups of Students in the Learning Trajectories

Construct: Describing Patterns and Relations Using Algebraic Expressions

Levels of Trajectories assessed ordered low to high





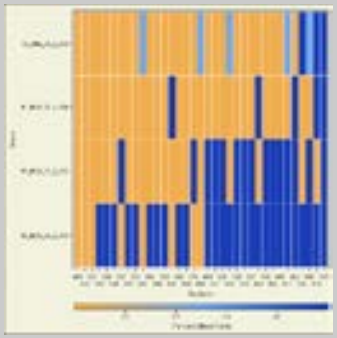
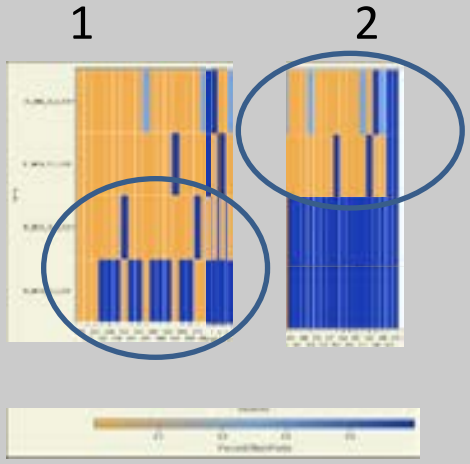
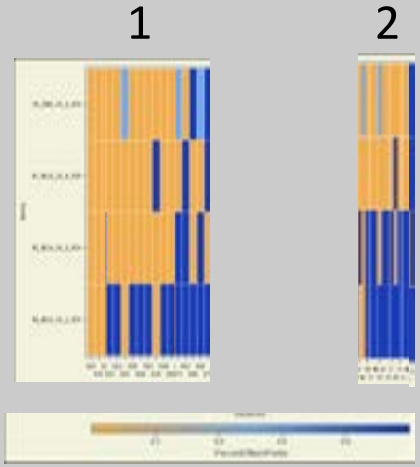
Blue is correct
Orange is incorrect

Students ordered on the X axis from low to high on each construct

Across students within a classroom

Across deciles within a grade

Group Students Based on Observed Score Patterns

	Homogeneous Groups	Heterogeneous Groups
	 	
Purpose	Each group receives a different activity chosen to fill shared gaps	Individuals learn from one-another
Score Patterns	Similar	Complementary

Overview

- The SUDDS team
- How our DLS can help close the gap
- A demonstration of the Digital Learning System (DLS)
- A closer look at our assessments
- What we have learned

Using LTs to Drive Professional Development

- Teacher training on learning map
- Students took diagnostic assessments
- Teachers reviewed results
- A two-week collaborative design study on the introductory clusters in statistics
  - Displaying Univariate Data
  - Measuring Data with Statistics

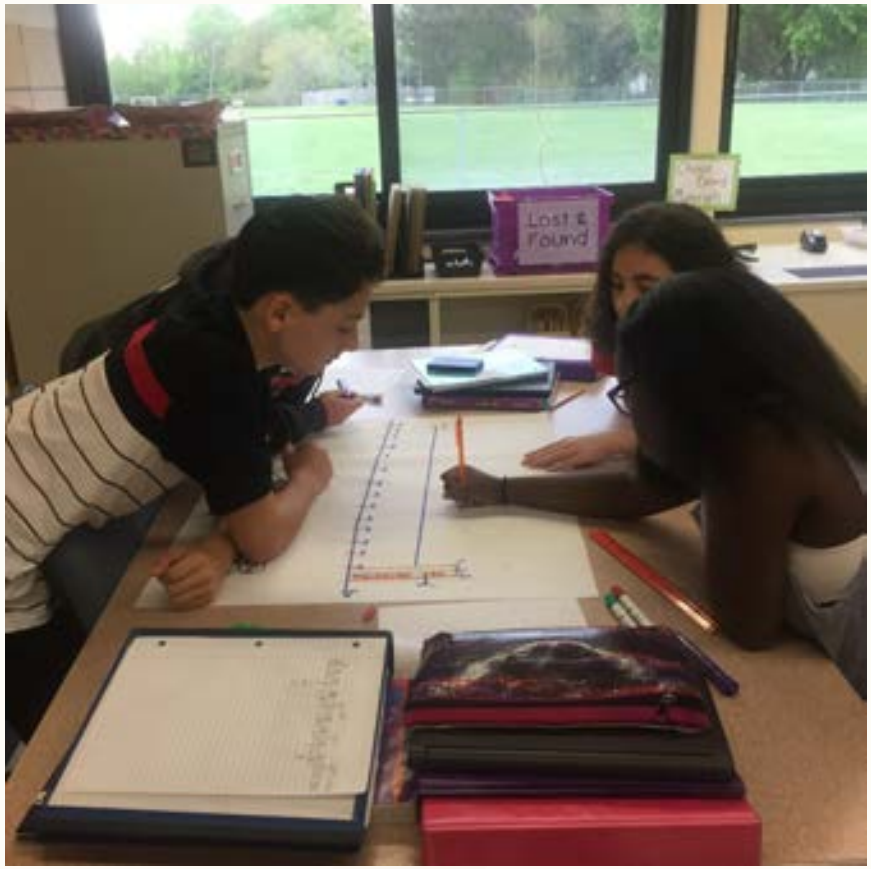
Impact on Student Engagement

- Provide them with open-ended tasks that elicit ideas
- Shift the classroom environment to allow students to express and explore their ideas
- Trust the students
- Include opportunity to learn for all students

Clip Five

Margaret Heritage's Positive Classroom Culture

- Mutual trust
- Intellectual rigor
- Expectation that ALL students learn
- Shared responsibility for learning
- Models of positive interactions
- Supportive, collaborative relationships



Impact of Immediate Feedback to Students

- Students take ownership of their results

Clip 4

Teacher Collaboration Around LTs

- Teachers discussed evidence of student learning through the LTs
- This impacted their planning the next day's instruction
- The LTs framed their conversations around student learning

Clip Three

Lab Sites

How Kids Create Their Own Knowledge

- The main topic of the debriefing session
- Questions
  - How you pose the question?
  - How do you support students without giving the answer?
  - How do you NOT say too much?
  - How do you help students hear each other's contributions?
  - How do you make sure your mathematical goal is being met?

Clip Six

How our Digital Learning System Can Help to Close Gaps

Summary

Learning trajectories

- ….coupled with professional development, can better prepare less-experienced teachers to plan, prepare for, and instruct students.
- …and tests span below grade giving teachers good leads on foundational gaps that need to be addressed if progress is stymied.
- … and tests span above grade giving teachers the freedom and support to move advanced students above grade.
- …are aligned with the common core state standards in Mathematics, but are meaningful without reference to the CCSS.
- Links provide previously vetted, high-quality learning materials that align with Learning Trajectories.

How our Digital Learning System Can Help to Close Gaps

Summary

- A variety of item types use interesting contexts, engaging visuals, and partial credit scoring that are sensitive to a wide range of student ability levels, keeping students motivated when grappling with challenging problems.
- Readability of items are at or below the targeted grade levels. Items will undergo a bias and sensitivity review in the near future.
- Growth-mindset is supported through-out our design.
- Heatmaps provide a means for teachers to group students according to need.

Next Steps

- Learn more about us at: https://sudds.ced.ncsu.edu/
- Visit the SUDDS map at: https://sudds.co/map
- Visit us on Teacher Connect at: www.suddsgroup.ning.com
- Email us at: sudds_group@ncsu.edu
- Contact us at:
- Jere_Confrey@ncsu.edu (919) 513-8523
- Garron_Gianopulos@ncsu.edu (919) 515-3890

New Research Directions

- An Integrated Testing System
- Matching Student Ability and Test Form Difficulty
- Task Modeling
- New Modeling Challenges
- Conventional versus LT-based Vertical Scaling

An Integrated System of Tests Each Designed for Different Diagnostic Purposes

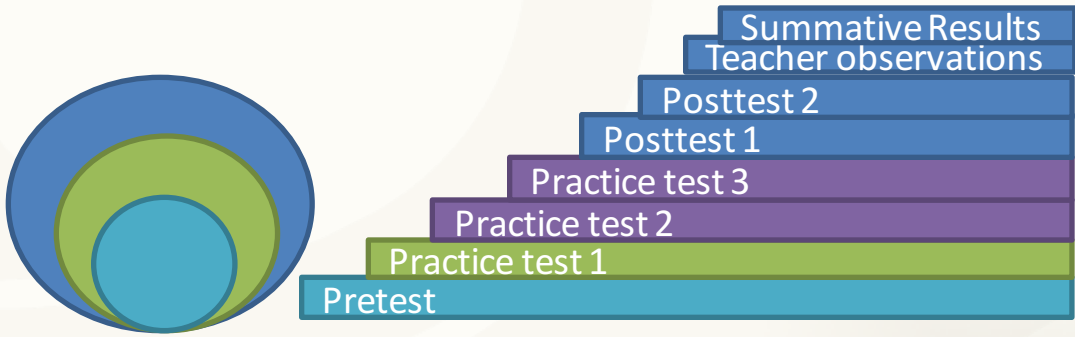
- The Problem: Domain under-representation is serious threat to the claims we wish to make with our testing system.
- In our original design concept, we would have had room for only 2 items from a learning trajectory, yet we want to claim that we can locate a student in each LT.
- A short test cannot possibly give us certain information about a large target domain.
- Confidence intervals will be just too large to support claims.
- We think it is too early to take an adaptive assessment approach.
- Our solution to this challenge is to divide the item pool into different tests, reducing the size of the target domain of inference for each test type: pretest, practice, and 'real' test.

Practice Tests
Prepare Students for Cluster Test

- Teachers or students pre-select range of LT levels for a practice form: Beginner, Intermediate, or Advanced.
- Practice tests are specific to the relevant section of a learning trajectory.
- Teachers give the competency test when cumulative evidence indicates students are ready.
- Eventually, the system will recommend tests based on predicted readiness.
- Competency tests focus on the section of the scale surrounding competency level.

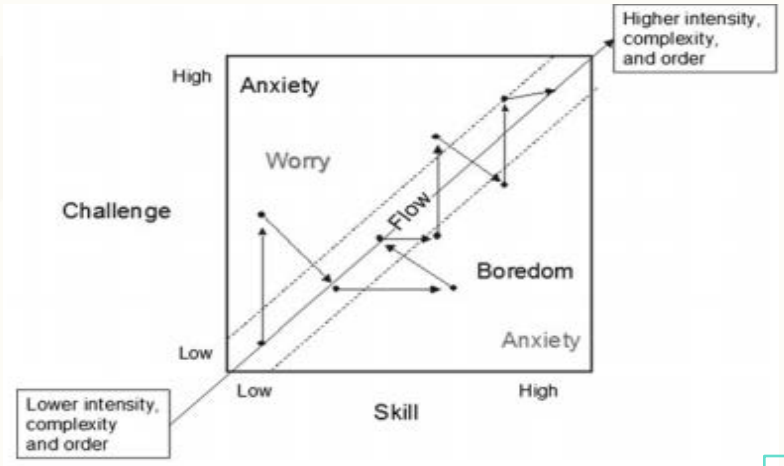
An Integrated System Designed to Collect Evidence Tied to Claims

- Evidence has a hierarchical structure (Post-test evidence(practice test evidence(pretest evidence))).
- The target domain of the pretest is the outcome space of practice tests.
- The target domain of the practice tests is restricted to the evidence observed in the post-tests.
- The target domain of the post-tests is restricted to evidence visible in the classroom, (and possibly) counter-balanced by external evidence from a summative measure valued by stakeholders.
- Total evidence accumulated across all tests can support final claim / interpretation of student readiness for next cluster.



Matching the Level of Challenge to the Student

- Flow research suggests that learners learn best when they are neither anxious nor bored by tasks.
- How will the testing system address the need to create an optimal learning environment for all students?

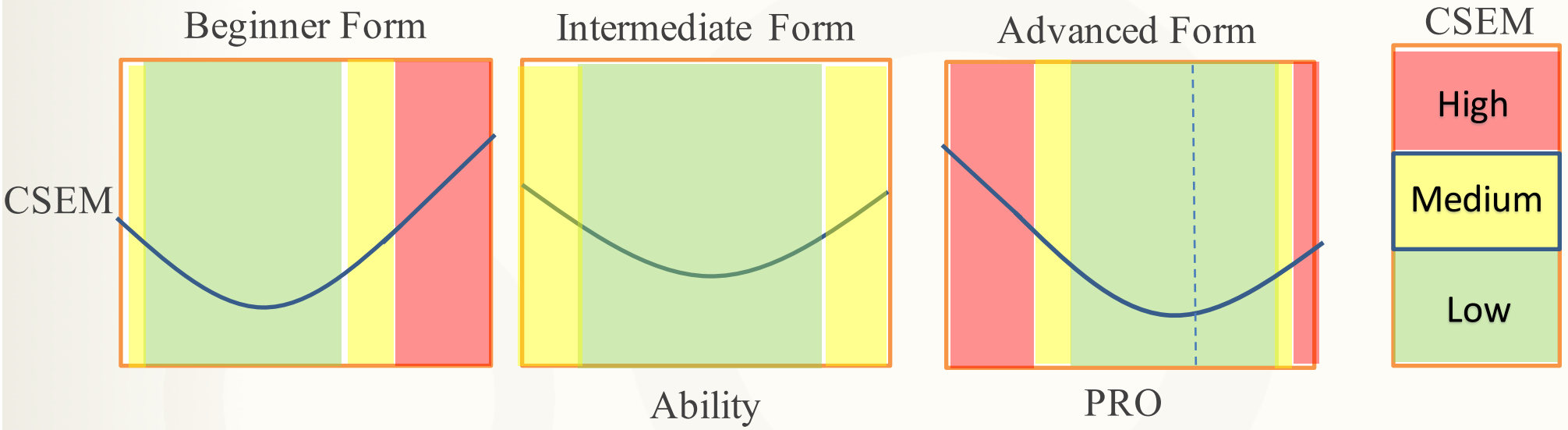


2.1 The first model of the flow state (Adapted from Csikszentmihalyi 1975/2000)

LT-based Vertical Scales Each with Different Tests each with their Purpose

Students or teacher select range of levels within a learning trajectory.

Forms are assembled on the fly to minimize measurement error (CSEM).



Purpose: Beginner practice Intermediate practice Advanced practice / competency test

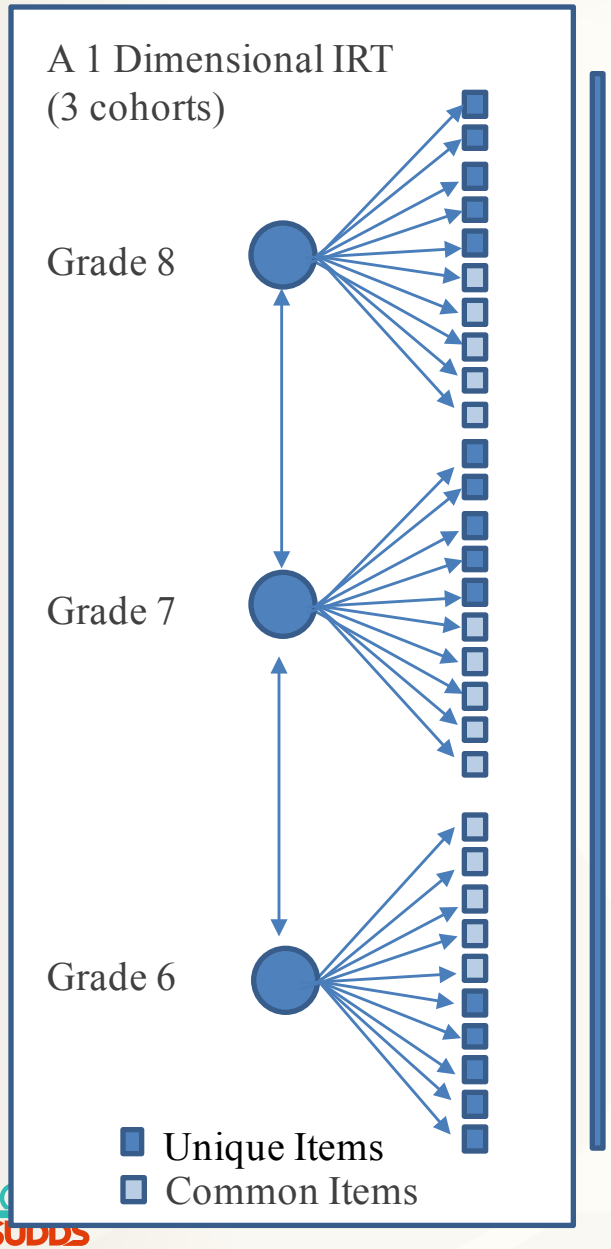
Quantitative Models

- Initially we are using the Rasch model to construct item maps (a.k.a Wright maps, construct maps)
- As our user-base increases, we will gain sufficient sample sizes to explore more complex models
- We plan to experiment with Rule Space Methodology to detect misconceptions, erroneous rules, and strategies of students
- We will explore Optimization Methods for Flexible Grouping
- We are considering Structural Equation Modeling and Bayes Nets for Learning Analytics

Modeling Challenges of Test Data Based on Learning Trajectories

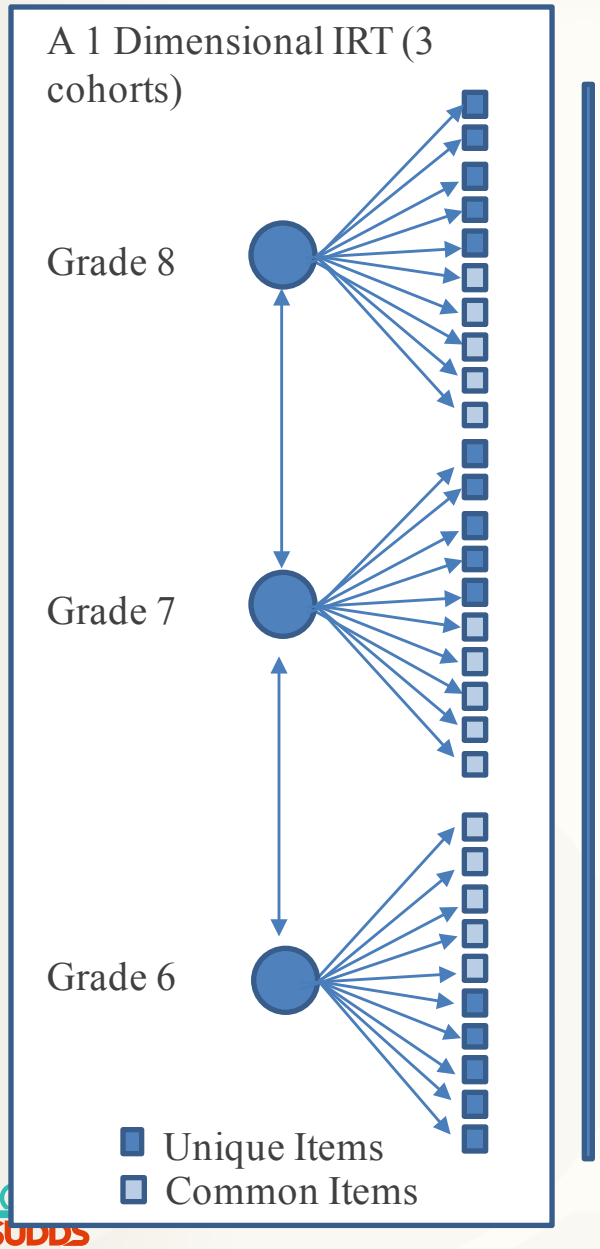
- Many scales to develop: 24 cluster scales and 64 LT scales
- Judging causes for item-data-misfit: failure to implement LTs in classroom, flawed item, or multidimensionality?
- Generating reliable and fine-grained diagnostic information concerning misconceptions and erroneous rules
- Modeling Student Growth through Learning Trajectories within and across grades
- Nearly ½ of all LTs span across grades

Conventional Vertical Scales



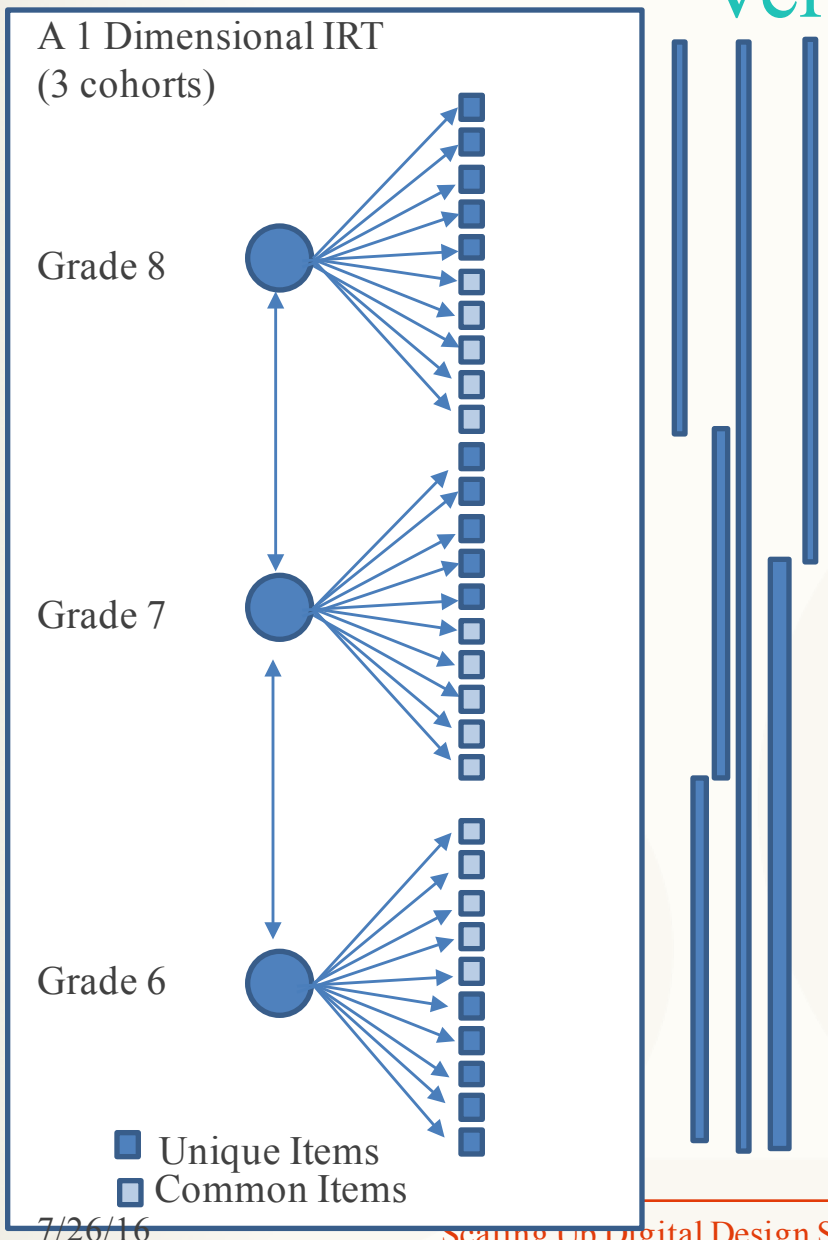
- A common approach to vertical scaling is via common item data collection design across grade cohorts & IRT linking/joint calibration, made possible via the strong assumptions of IRT: unidimensionality, local independence, nonspeededness, and model-data-fit.
- Within-grade unidimensionality can be achieved by selecting items with high item discrimination.
- However, the cost of selecting items in this manner is a reduction of subscale usefulness.

Conventional Vertical Scales



- Forcing many distinct scales into one can create a vertical scale that is less sensitive to change in one of the grade-specific scales.
- In practice, it is not unusual to see very small mean differences in vertical scale scores across grades, especially 7 to 8, and 8 to 9.
- Gains and losses cancel each other out.
- Such scales are not diagnostic as they typically cannot answer why the net change is small or negative.

A Learning Trajectory Approach to Vertical Scaling

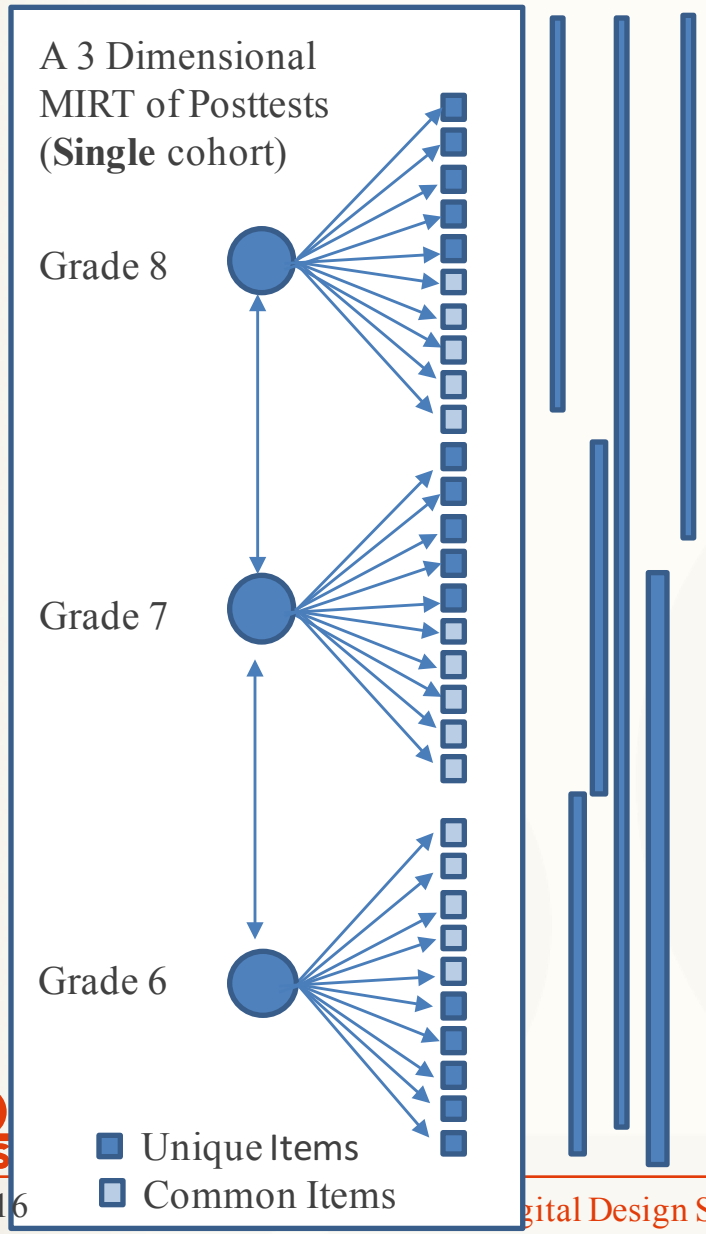


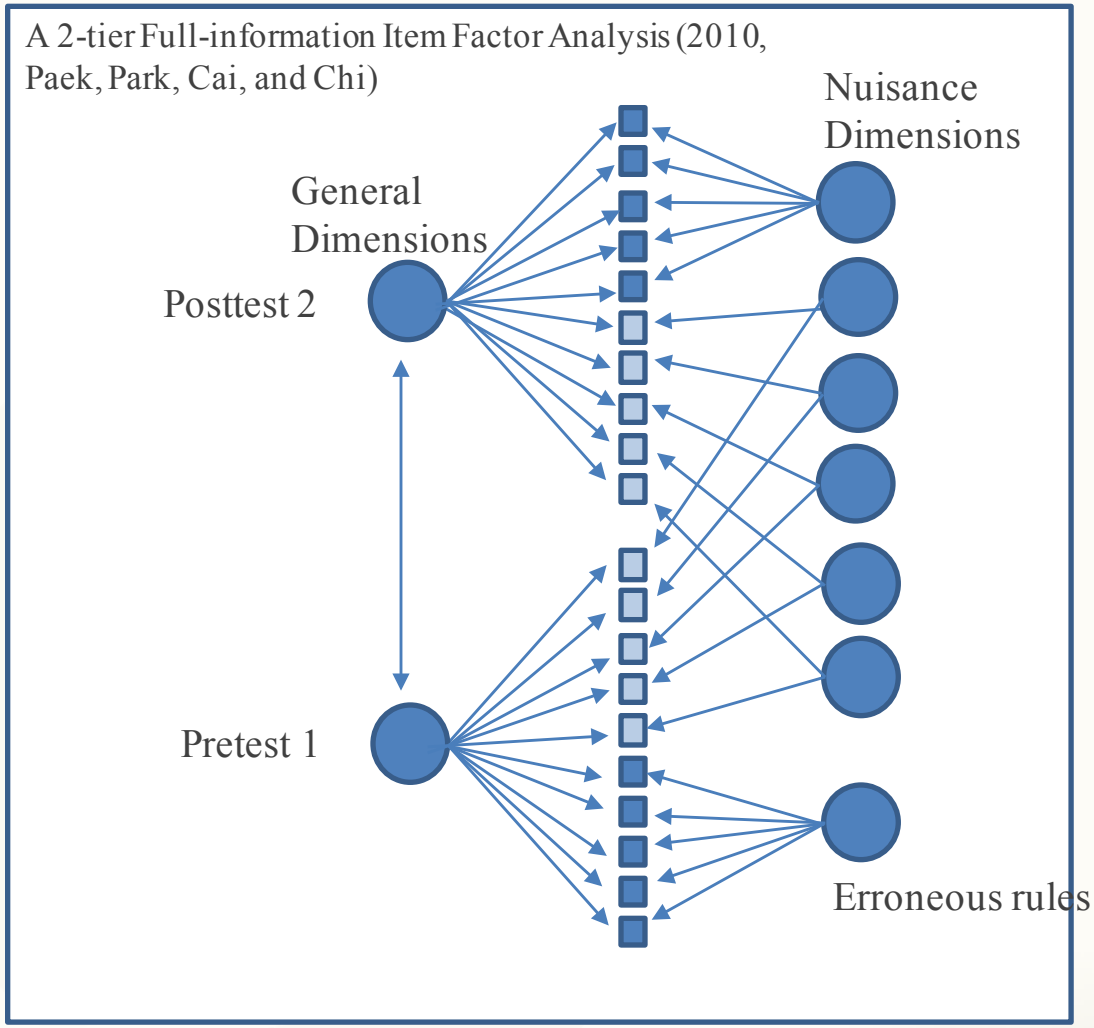
- An LT approach to vertical scaling may overcome the problem of across-grade multidimensionality by dividing the content into many distinct learning trajectory dimensions.
- We developed 64 LTs.
- ½ of our LT dimensions are within a particular grade, 1/2 span 2 or more grades.
- Items in each LT scale discriminate highly in the LT scale not with a general grade-specific dimension.

With-in Grade Multidimensionality

- Kikumi Tatsuoka demonstrated that novice math learners are more likely to use erroneous math rules than proficient students.
- In Tatsuoka's experience, once the response patterns of erroneous rules were removed from the data set, the factor structure became unidimensional.
- This suggests a unidimensional assumption may not hold across high-ability and low-ability subgroups of students who consistently use erroneous rules or misconceptions.
- How can such a factor structure be modeled?

Vertical Scaling as a Longitudinal Data Modeling Task





Next Steps

- Fit Rasch models to field test data and generate scales
- Set provisional standards to define grade level expectations
- Build forms for each test type by grade band
- Create reports for teachers at the class level

Criteria for High Quality Assessment

A well-articulated validity evaluation based on an interpretative argument (Kane, 2006) is provided that includes, at a minimum

1. Evidence of the validity of using results from the assessments for the three primary purposes, as well as any additional purposes required by the state (specify sources of data).
2. Evidence that scoring and reporting structures are consistent with structures of the state's standards (specify sources of data).
3. Evidence that total test and relevant sub-scores are related to external variables as expected (e.g. other measures of the construct). To the extent possible, include evidence that the items are "instructionally sensitive," that is, that item performance is more related to the quality of instruction than to out-of-school factors such as demographic variables.

Criteria for High Quality Assessment

4. Evidence that the assessments lead to the intended outcomes (i.e., met the intended purposes) and minimize unintended negative consequences. Consequential evidence should flow from a well-articulated theory of action about how the assessments are intended to work and be integrated with the larger accountability system.
5. The set of content standards against which the assessments are designed is provided. If these standards are the state's standards, evidence is provided that the content of the assessments reflects the standards, including the cognitive demand of the standards. If they are not the state's standards, evidence is provided of the extent of alignment with the state's standards.

Criteria for High Quality Assessment

6. Evidence is provided to ensure the content validity of test forms and the usefulness of score reports (e.g., test blueprints demonstrate the learning progressions reflected in the standards, and experts in the content and progression toward readiness are significantly involved in the development process).