RUTGERS

InqITS™

Real Time Science Assessment & Tutoring

Inq-ITS: AI-based technology for real time assessment, scaffolding, and instruction of science practices

Dr. Janice Gobert, Professor,

Rutgers Graduate School of Education

&

Founding CEO, Inq-ITS (inqits.com)

Overview

- The context of the problem
- Goal & Intro to Inq-ITS and Inq-Blotter
- Design & Development
- Fire hose Study 1: Identifying the Messy Middle
- Fire hose Study 2: Testing for Transfer of Rex
- Fire hose Study 3: Alerting & transfer
- Fire hose Study 4: TIPS
- Questions & discussion about formative assessment

Context: US & many others face a STEM crisis

1	China
2	Singapore
3	Estonia
4	Japan
5	Finland
6	Korea
7	Canada
8	Hong Kong
. . .	
18	U.S.A.

Int'l rank on science (PISA), 2018

RUTGERS

Science Inquiry Learning & Assessment

- NGSS and other state frameworks require that students learn these practices and teachers must provide evidence of students' competencies

- Asking questions
- Developing and using models
- Planning and carrying out investigations
- Analyzing and interpreting data
- Using mathematics and computational thinking
- Constructing explanations
- Engaging in argument from evidence
- Communicating findings

(NGSS Lead States, 2013; NRC, 2012)


BIRDS
SPIDERS


BILLY

Gobert, US News & World Report, May 2016

Our Goal

- We wanted to assess science practices *rigorously*, *at a fine-grained level (sub-components underlying each practice), and assess these at scale*

- We wanted to *better support teachers*:
  - *In their instruction of science practices at an actionable level*
  - *In their assessment of science practices*
    - *at a fine-grained level to report on a full range of competencies,*
    - *for Ss with longer learning trajectories, etc.*

- We wanted to *better support students*:
  - *In their learning of science practices so that they will transfer to new topics, as envisioned by the NGSS.*

Inq-ITS Overview & Components

- **Simulation-based assessments (157) for science (4-10 grade)**

  - to assess science practices designed with the assessment triangle (Pellegrino et.al.,2001) and Evidence-Centered Design (Mislevy et al, 2012)

  - include content variables based on students' misconceptions & difficulties with practices (Kuhn, 2005)

  - built, piloted and, refined with think alouds with individual students & teachers (Gobert & Sao Pedro, 2017)

- **Simulations, widgets, and tools capture students' interactions in Inq-ITS**

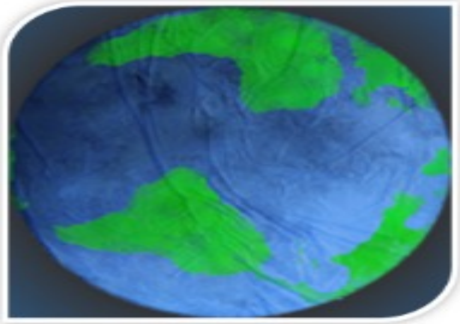
  - these elicit & capture student performances

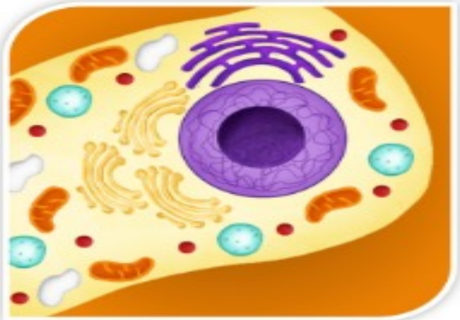
- **Performance assessments, reports, scaffolds, & alerts on students' inquiry practices/skills are generated**

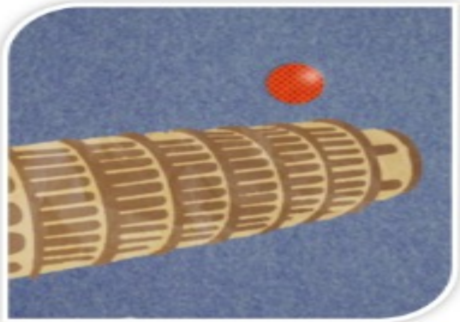
  - required operationalizing each inquiry practice into its respective sub-skills

- **ML, KE, and NLP are used**

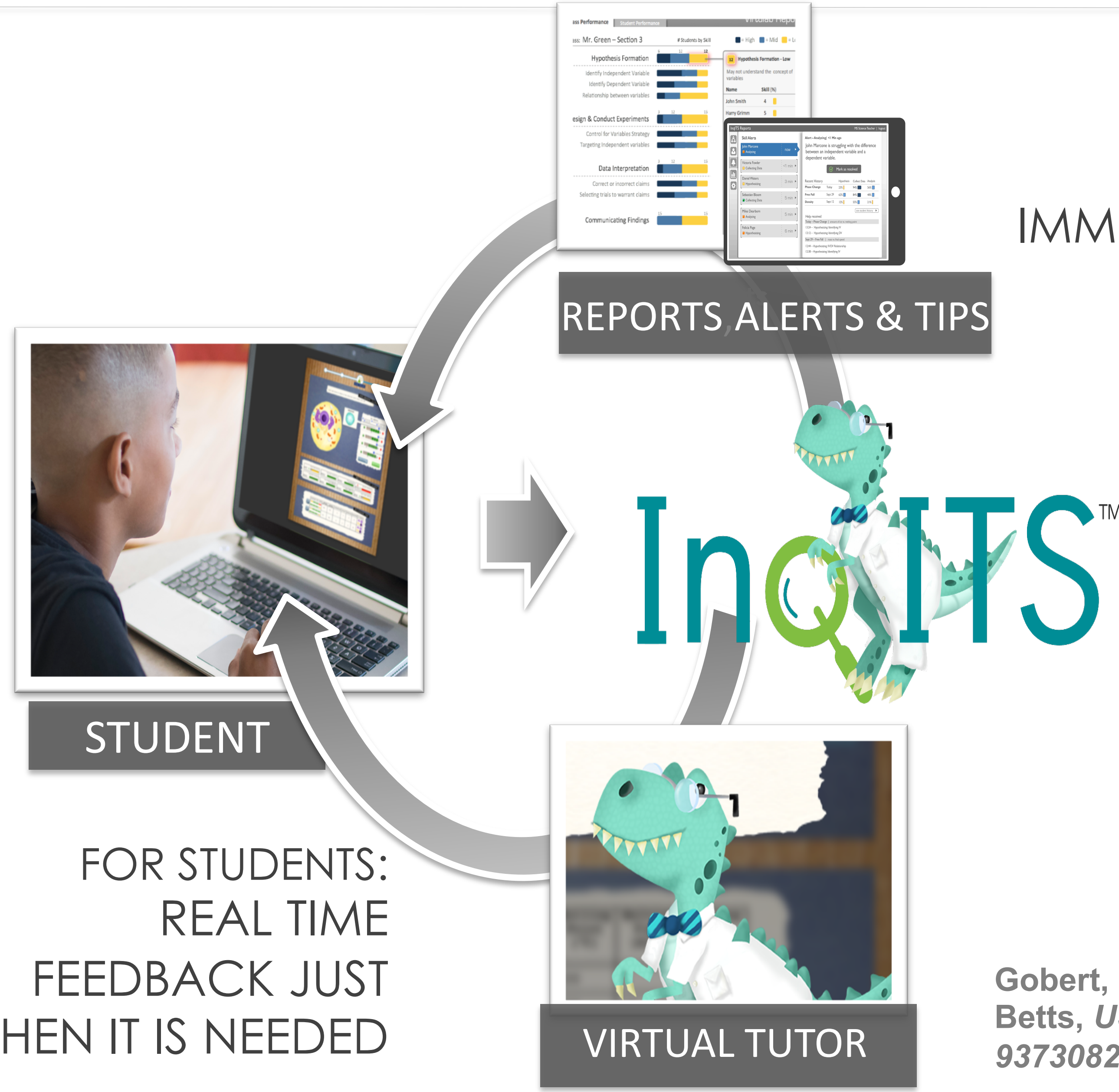
  - to assess & scaffold students' practices, and alert teachers on who needs help & how







Inq-ITS overview & components



FOR TEACHERS:
ACTIONABLE
REPORTS &
IMMEDIATE ALERTS
formative use

FOR STUDENTS:
REAL TIME
FEEDBACK JUST
WHEN IT IS NEEDED

Gobert, Sao Pedro, Baker, & Betts, *US Patent nos. 9373082, 9564057, 10186168*

Evidence-Centered Design

- Evidence-centered design (ECD, Mislevy, et al., 2003) builds off of Pellegrino’s, assessment triangle.

- ECD specifies the evidence to support the intended inference including data aggregation and data interpretation (Mislevy, et al., 2003; Mislevy, et al., 2009; Mislevy et al., 2020).

- Specifically, ECD provides fine-grained details guiding the design of the computational objects (key to eliciting observations) and the computational processes (key to interpretation) that are to be carried out in the assessment system in order to generate fine-grained, rigorous assessment data on students’ competencies.

- ECD has been used for a wide range of topics, including simulation-based assessments and ITSs (e.g., Clarke-Midura, Code, Dede, Mayrath, & Zap, 2012; Mislevy et al., 2014; Shute, 2011; Mislevy et al., 2020).

ECD, cont'd

- ECD is critically important to unpack and specify all aspects of the assessment process from start to finish, particularly for the assessment of science inquiry because of its ill-defined nature (cf., Kuhn, 2005).

- In brief, the assessment triangle & ECD provides a way to understand end-to-end assessment design that can inform assessments needed for teachers' needs for formative assessment and instruction of science practices reflected in reform documents (NGSS, 2013).

Gobert et al. (2012). *Journal of Educational Data Mining; Mislevy, Yan, Gobert, & Sao Pedro, 2020*

The Assessment Triangle & ECD

- First corner is Cognition
  - *making explicit the conceptions of how people learn and the knowledge and skills that are associated with the targeted knowledge/competencies.*

- Since all assessments are based in an underlying theoretical framework about how people learn (Gotwals & Songer, 2009; Scardamalia & Bereiter, 2008),
  - *it is necessary to specify or operationalize the knowledge and or cognitive processes underlying the targeted performance/skill/conceptual knowledge within its respective domain,*
  - *this includes both observable cognitive processes and those that are proposed as underlying cognitive activities associated with those processes.*

2nd corner: Observation

- The second corner is *observation*.
  - *The way(s) by which a student's knowledge or competencies are observed for a target conception or skill/practice.*
  - *The key to eliciting observations is designing an assessment item/task that will give students/learners opportunities to demonstrate their knowledge and/or competencies.*
  - *Note: the tasks must also elicit a broad range of competencies, which is important for assessment partial and developing competencies.*

3rd corner: Interpretation

- This refers to how people, i.e., researchers, *infer* students' internal knowledge, cognitive processes, and mental states (i.e., representations) from their observable behavior(s)/actions.
  - these were elicited by the tasks/items the student engages in
- *The knowledge, processes, and mental states to be interpreted are those that have been made explicit in the cognition corner of the assessment triangle.*
- *It is critical that researchers attend to how the data will be analyzed as part of the early part of the assessment design, otherwise, one runs the risk of not being able make strong claims about students' knowledge and competencies (Mislevy et al., 2020).*

Inq-ITS targets students' known difficulties with inquiry

- Some specific student difficulties are…
  - Identifying variables to target and observe
  - Conducting controlled experiments
  - Using data to determine relationships between variables
  - Doing mathematics associated with science inquiry
  - Using and explaining data (evidence and reasoning) while communicating results
  - Linking data with explanations

- Teachers spend considerable time scaffolding students' difficulties (Aulls, 2002), and if not done, inquiry can lead to students' alternative conceptions (Brown & Campione, 1994)
  - Real time assessment and monitoring can help this, and improve students' learning

Inq-ITS assesses, scaffolds, and alerts on students' question formation competencies

Students:

- **have difficulties forming testable hypotheses & choosing which variables to work with** (Chinn & Brewer, 1993; Klahr & Dunbar, 1988; Kuhn et al., 1995)– *e.g. Students sometimes use two Ivs or 2 DVs instead of an iv and a dv.*
  - *we scaffold on this and alert teachers on this!*

- **may not know what a hypothesis should look like** (Njoo & de Jong, 1993);
  - *we do not typically see this problem b/c our widget helps them wit the epistemic structure of a hypothesis, important to lower skilled students and ELLs*

- **avoid stating hypotheses that could be rejected** (Van Joolingen & de Jong, 1993; Klayman & Ha, 1987; Klahr, Fay & Dunbar, 1993)
  - *our widget allows students to select "Does not support my hypothesis" though Ss don't like to select this and some will try to go back to change their hypothesis. Teachers handle this instructionally re nature of science.*

… Assesses, scaffolds, and alerts on students' competencies at designing & conducting experiments

Students:

- **may not test their articulated hypotheses (i.e., their experiment does not match their hypothesis)**
  (Van Joolingen & de Jong, 1991b, 1993; Kuhn, Schauble, Garcia-Mila, 1992; Schauble, Klopfer, Raghavan, 1991)

- **may not gather sufficient evidence to test hypotheses (i.e., they do not do enough trials)**
  - *Often run only one trial (Kuhn, Schauble, Garcia-Mila, 1992)*
  - *running the same trial repeatedly (Kuhn, Schauble & Garcia-Mila, 1992; Buckley, Gobert & Horwitz, 2006)*

- **may change too many variables within the same trial** ***(Glaser et al., 1992; Shute & Glaser, 1990; Kuhn, 2005; McElhaney & Linn, 2008, 2010)***

- **may run experiments that are enjoyable to watch as opposed to hypothesis testing** ***(White, 1993; Schauble, Klopfer & Raghavan, 1991; Njoo & de Jong, 1993a).***

… Assesses, scaffolds, and alerts on students' data interpretation competencies

Students:

- **may draw conclusions based on confounded data** (Klahr & Dunbar, 1988; Kuhn, Schauble & Garcia-Mila, 1992; Schauble, Glaser, Duschl, Schulze & John, 1995).
  - *In their trials, they have changed too many variables within the same trial*
- **may change ideas about causality** (Kuhn, Schauble & Garcia-Mila, 1992)
- **do not relate outcomes of experiments to theories being tested** (Schunn & Anderson, 1999).
- **Make claims about data they did not collect (Li et al., 2019)**
- **may engage in confirmation bias during inquiry** (Klayman & Ha, 1987; Dunbar, 1993; Quinn & Alessi, 1994; Klahr & Dunbar, 1988; Dunbar, 1993).
  - *State conclusions that do not align with their experimental findings, despite having correct.*

… Assesses, scaffolds, and alerts on students' argumentation competencies

Students:

- **Use inappropriate and insufficient data and providing reasoning for their claims** (McNeill & Krajcik, 2011; Sadler, 2004).
    - *They choose trials that are confounded, or they do not choose enough trials to justify their claims, evidence, and reasoning.*

- **Have trouble linking their data to their hypotheses** (Chinn & Brewer, 1993; Klahr & Dunbar, 1988) or to their claims (Schunn & Anderson, 1999)

- **Rely on theoretical arguments rather than on experimental evidence** (Kuhn, 1991; Schunn & Anderson, 1999).

In progress: Assess, scaffold, and alert on students' mathematical competencies (IES Gobert & Sao Pedro)

- related to constructing graphs in science:
  - determining axes of graphs (Lai et al., 2016; Nixon et al., 2016)
  - determining what data to include in graphs (Lai et al., 2016; Tairab & Khalaf Al-Naqbi, 2004)
  - understanding the relationship between variables in data tables and graphs (LópezLeiva et al., 2016; Strobel et al., 2018)

- related to applying equations in science:
  - determining the type of functional relationships that exist between data (i.e., linear, curved, etc.); (De Bock et al., 2017; Lai et al., 2016)
  - understanding the components of equations (i.e., slope) (Lai et al., 2016; Nixon et al., 2016; Planinic et al., 2012)
  - applying best-fit lines (Casey, 2015; Nixon et al., 2016)

Inq-ITS will assess, scaffold, and alert on students' mathematical competencies

- choosing appropriate axes,
- plotting data that are unconfounded,
- plotting sufficient data to see a trend,
- creating a best fit curve through data by choosing an appropriate functional form (e.g. linear, nonlinear....),
- constant and coefficient, and recognizing when a best fit curve fits the data well.
- forming claims and warranting claims with mathematical data as evidence

Inq-ITS' virtual simulations (cont'd)

- Simulations permit authentic science inquiry learning and assessment

  - because they share many features with real apparatus, leveraging perceptual affordances (Norman, 1983)

- With simulations students:
  - *develop a hypothesis/ask questions*
  - *use models to plan and carry out investigations*
  - *analyze data (w or w/o mathematics)*
  - *warrant claims*
  - *construct explanations*
  - *argue from evidence*

Benefits of virtual simulations for cognition, observation, interpretation, & instruction

- Affordances of authenticity
  - *offer greater validity than m/c tests*
  - *currently not used fully to assess the skills/knowledge they were designed to foster!*
- Generate rich, high fidelity log files
  - *used performance assessment*
- Work products & inquiry processes can be used for assessment *(Rupp et al, 2010)*
- Can give immediate feedback; blending learning and assessment
  - *extra assessment time is not used*
  - *could replace summative tests*

But there are challenges to inquiry assessment in interactive environments …

- Complex tasks take longer => if fewer measures of "one type", could have *reduced reliability (Shavelson et al, 1999)*
- More than one way to conduct inquiry => variability in student responses
- Sub-tasks are not independent from each other => assumptions of conditional independence do not hold *(i.e., Classical test theory; Mislevy et al, 2012)*
- Traditional measurement methods tough to apply due to changing skill level as students learn in real time *(cf. Levy, 2012)*
- Theory needed to both distill/aggregate data *(Gobert et al., 2013)*, and to design categories a priori (critical to interpretation)
    - *so that results are pedagogically meaningful to key stakeholders, i.e., teachers, parents, students, & policy-makers.*

The opportunity for real time performance assessment, scaffolding, & alerting

- **Methodological advances in computational techniques,** i.e., data mining offer analytical leverage on students' learning *processes (not just products),*
  - *done in real time,*
  - *done at scale,*
  - None of which were possible before.

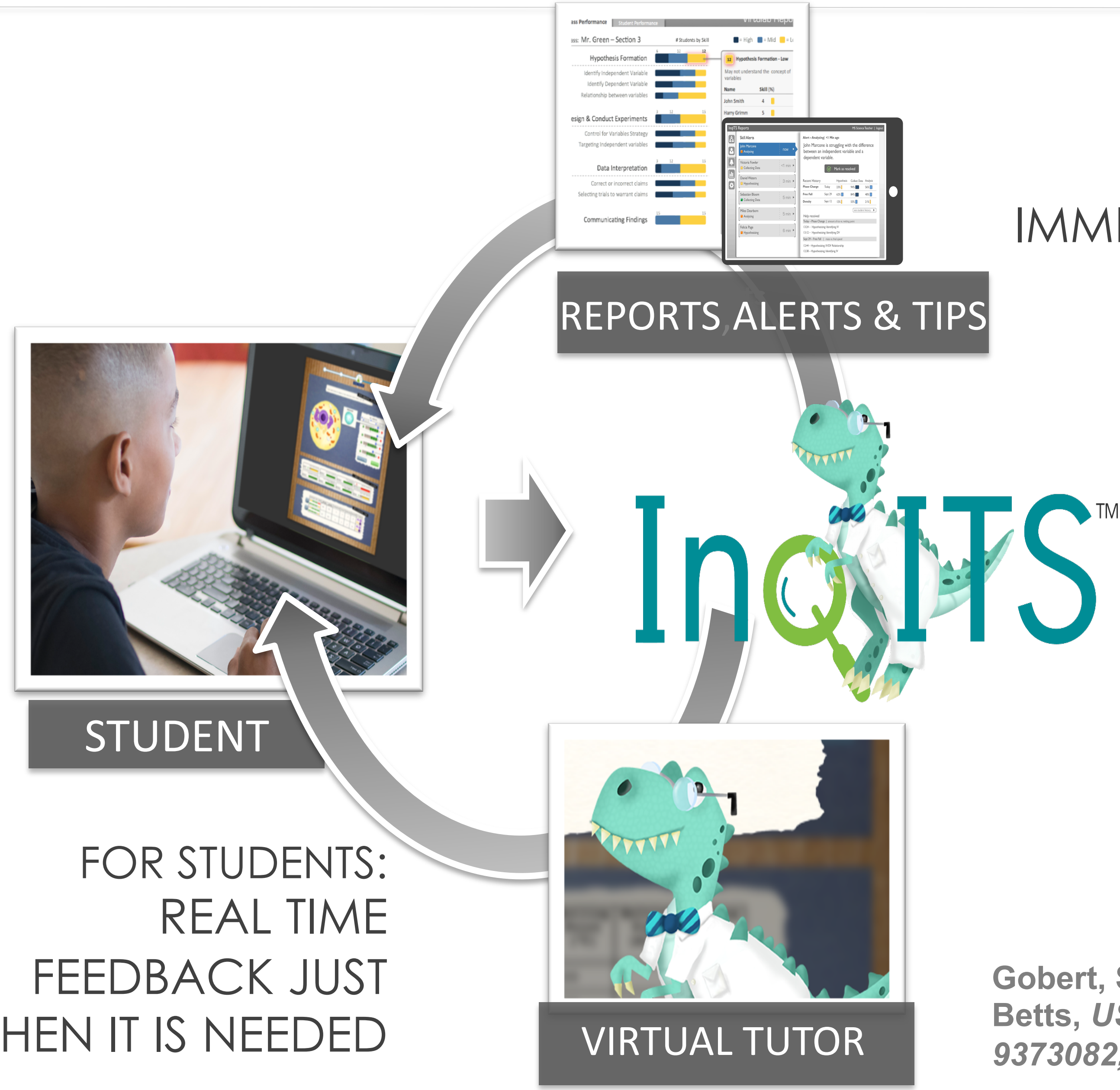
- Also computational techniques:
  - *can handle the 4 V's of log data (volume, veracity, velocity, and variability in data)*
  - *are desirable for ill-defined problems, like science inquiry*
  - *are transformative for both stealth assessment and real time scaffolding*
    - *Important to democratizing learning*
  - *offer scalability, important to Education reform*

J. Gobert, US News & World Report, May 13, 2016

RUTGERS

Inq-ITS overview & components

FOR TEACHERS:
ACTIONABLE
REPORTS &
IMMEDIATE ALERTS

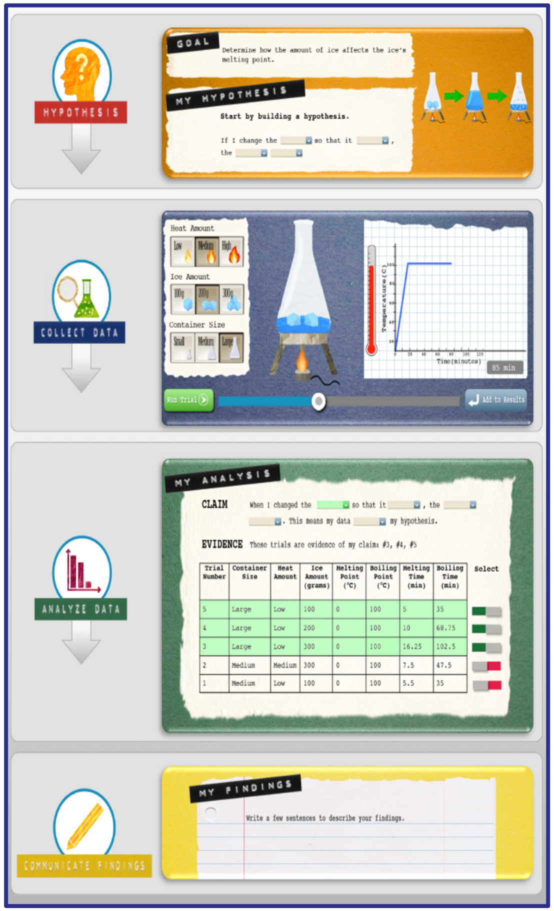


FOR STUDENTS:
REAL TIME
FEEDBACK JUST
WHEN IT IS NEEDED

Gobert, Sao Pedro, Baker, & Betts, *US Patent nos. 9373082, 9564057, 10186168*

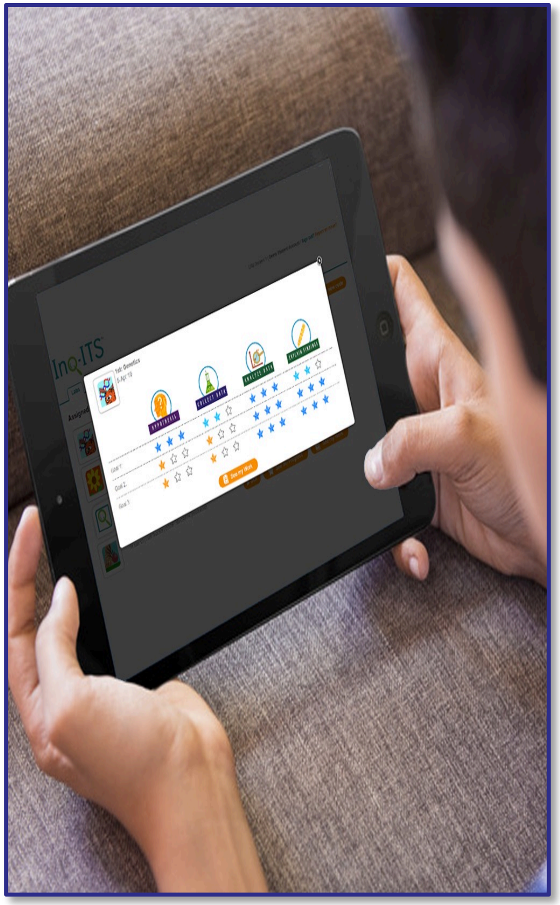
Components for Students

Authentic Inquiry



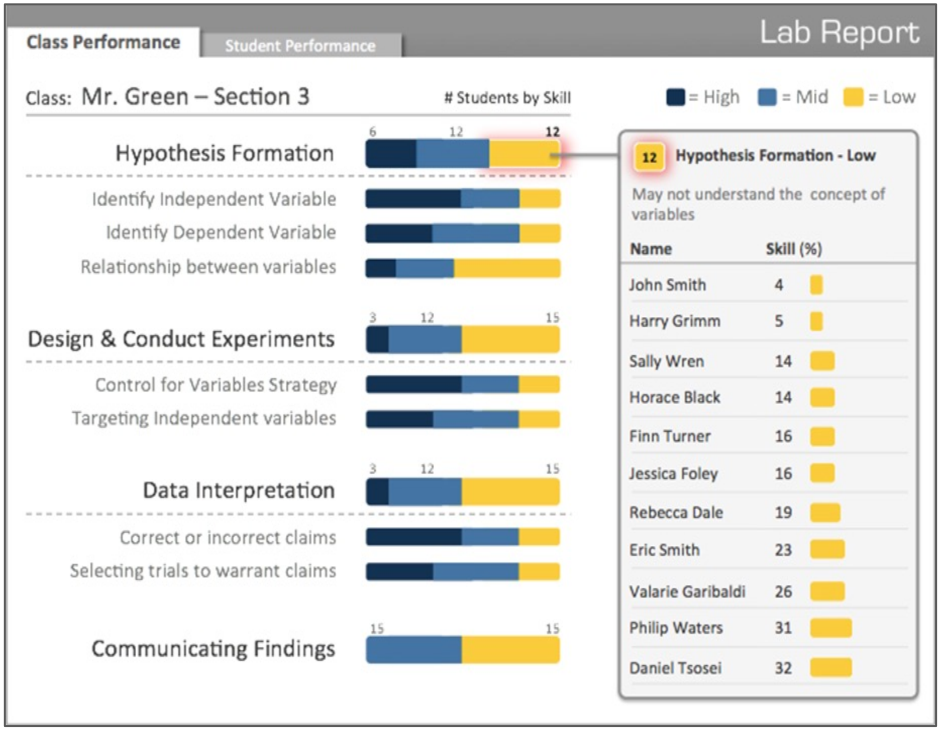
Rex, Personal Tutor

Student Reports

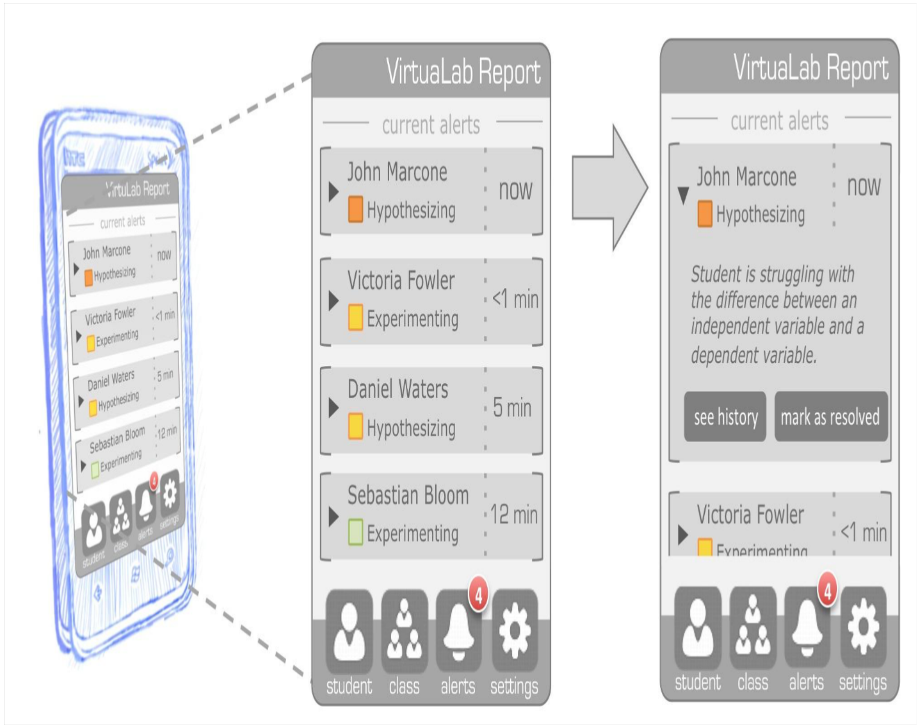


Components for Teachers

Real-Time *Reports* for Assessment

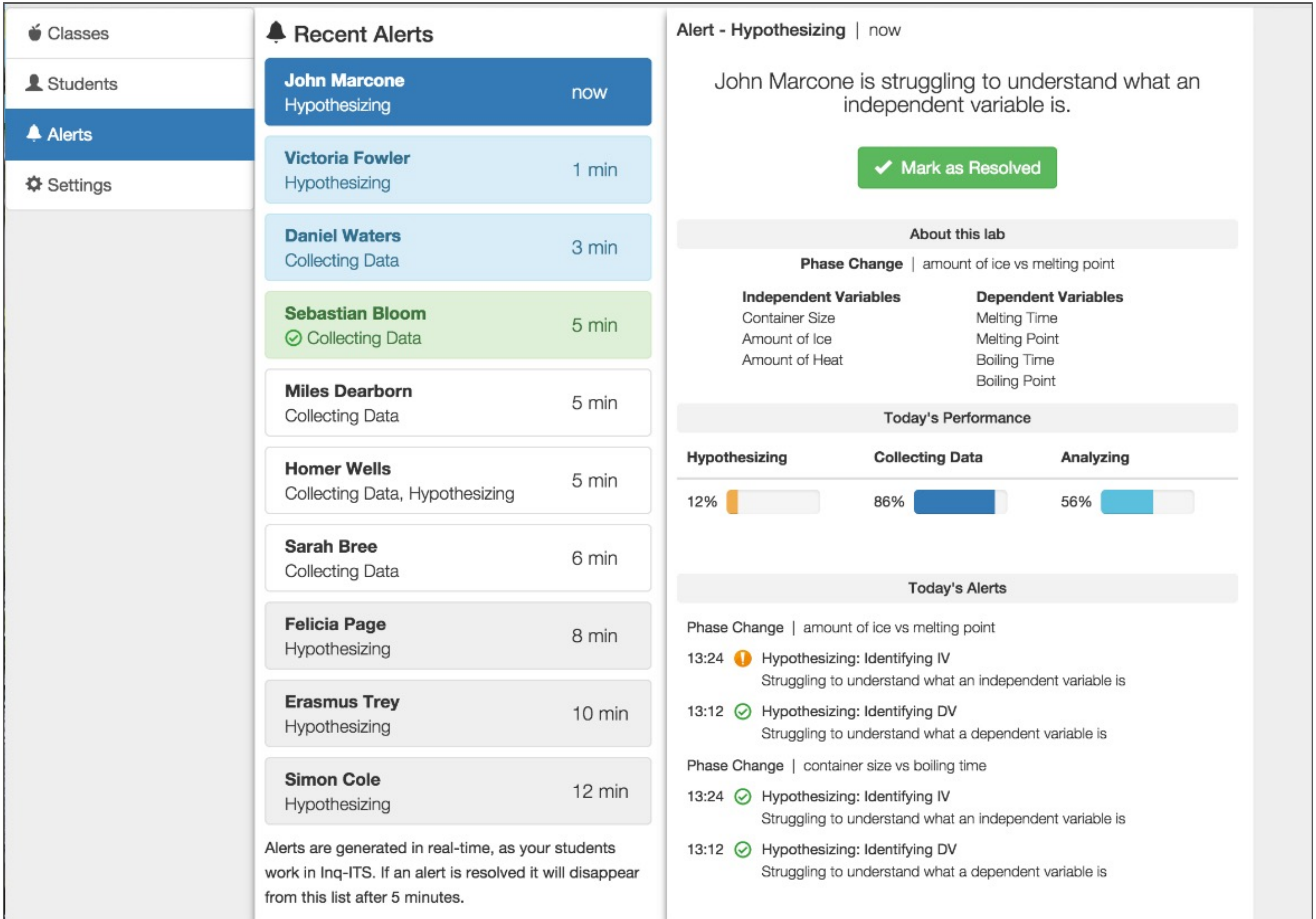


Real-Time *Alerts* for Instruction



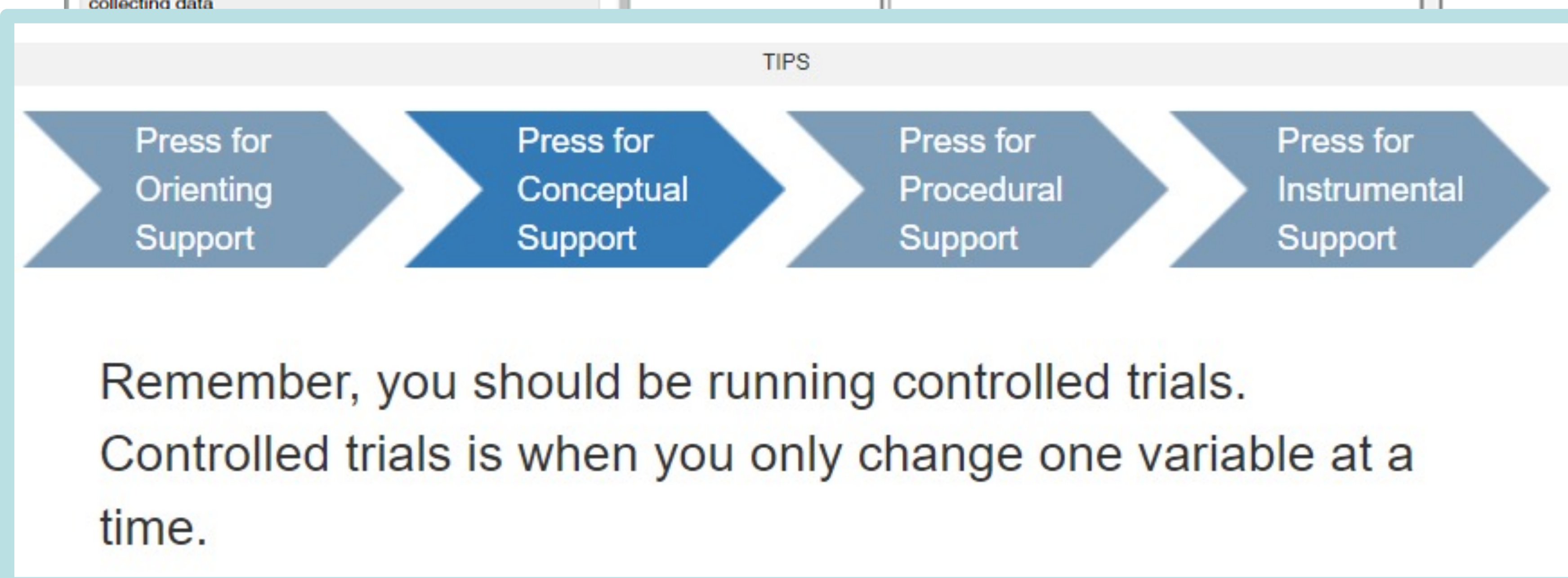
RUTGERS

Components for Teachers: Alerts



RUTGERS

Components for Teachers: TIPS (Teacher Inquiry Practice Supports)



Machine learning/EDM "under the hood" is used for real time assessment, scaffolding, and alerting

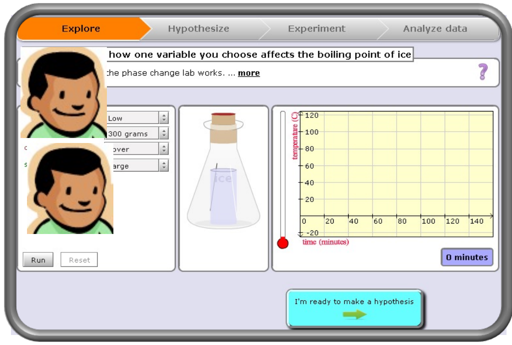
- Data mining & text replay tagging (Baker et al., 2006) are used to develop canonical models of what it means to demonstrate skill, and in turn, develop algorithms,
  - *These can handle variability of students' processes used in inquiry*
  - *We have done this for all inquiry practices; very close match to human scorers (90%+)*

  (Gobert & Sao Pedro, 2016).

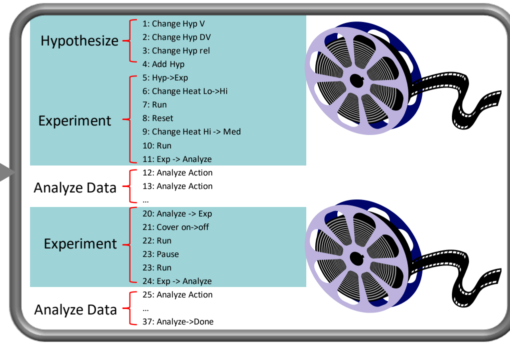
- Our algorithms *built on diverse student data* & validated over multiple topics *(Gobert et al., 2013, 2015)*
  - *Generalizability tested on new students not used to build models* (Paquette et al., 2014)

Summary of EDM development & validation

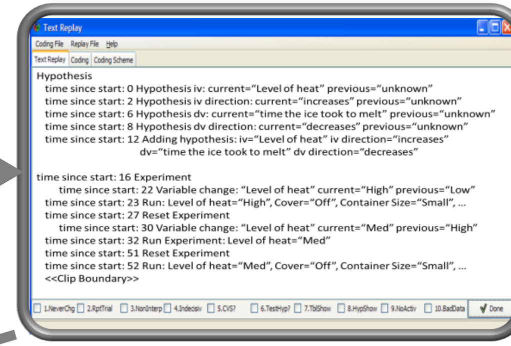
Collect Data



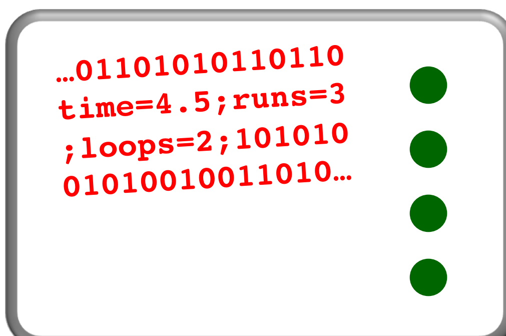
Build Text Replay Clips



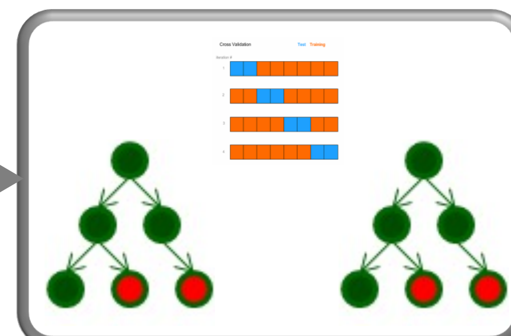
Tag Clips for Skills



Define Features w Rapid Miner



Build & Validate Detectors



Gobert, J., Baker, R., & Sao Pedro, M. & Betts, C. *Inquiry Skills Tutoring System*. US Patents 9373082, 9564057, 10186168

RUTGERS

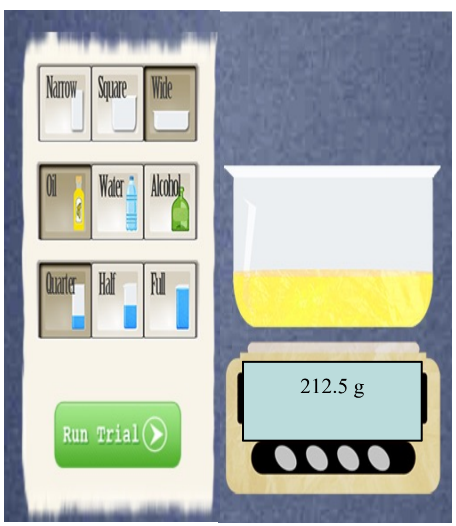
Study 1: Identifying the messy middle

"Doing" Science vs Explaining Science

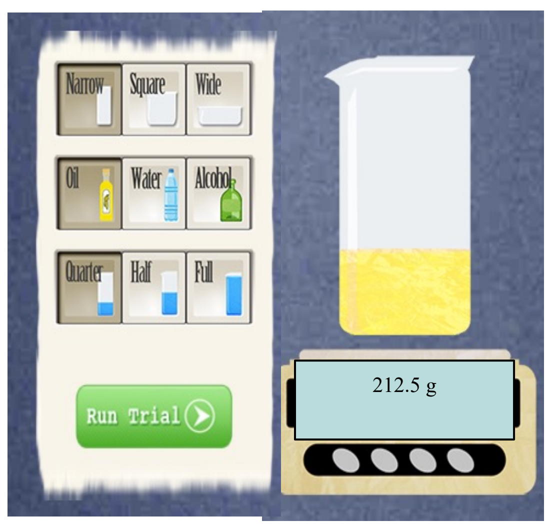
Doing		Writing
• Generating a research question • Forming a hypothesis • Collecting, analyzing, & interpreting data • Selecting data to warrant a claim • Etc.	vs	• Claim • Evidence • Reasoning • etc.
NGSS, 2013		Toulmin, 1958; McNeill et al., 2006

Participants and Materials

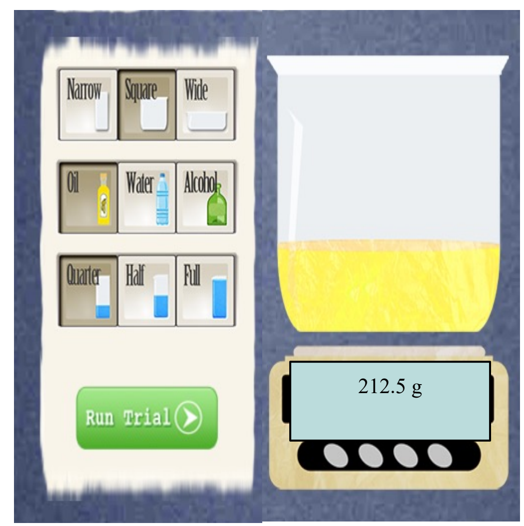
- 293 middle school students, 6 public middle schools
- Materials:
  - Inq-ITS: Density Virtual Lab
    - Shape of the Container Activity



Wide



Narrow



Square

Measures

1. Experimental-Doing in Inq-ITS
    a. All student actions (clickstream inquiry data) are stored in log files
    b. Automated scoring using Educational Data Mining and Knowledge Engineering for hypothesizing, collecting data, & analyzing data
    -13 sub-practices scored as present (1 point) or not (0 points)*

(Gobert et al., 2013, 2014, 2018; Moussavi et al., 2016; Moussavi, 2018; Sao Pedro et al., 2013; Sao Pedro, 2013)

Measures

2. Explanatory Writing in Inq-ITS
   b. Students' CER explanations were extracted from database
   c. Used rubrics developed and validated by Li et al. (2017b)

      -- Two raters hand scored student explanations from final Inq-ITS' stage according to sub-components of claim, evidence, and reasoning (CER)
   d. Inter-rater reliabilities were: claim (kappa = 0.964), evidence (kappa = 0.973), and reasoning (kappa = 0.774)
   e. *now autoscoring CER based on NLP (with very high correlations b/w humans and algorithms, .90, .94, .86 for C-E-R, respectively).

(Gobert et al., 2018; Li et al., 2017a, 2017b, 2018)

RUTGERS

Analyses

- K-means clustering was used to group students
- 144 of 293 students (49%) were in the messy middle:

False Positives ("Johnnys")

Low Doing – **High** Writing **(N =131)**	**High** Doing - **High** Writing
Low Doing – **Low** Writing	**High** Doing – **Low** Writing **(N = 13)**

False Negatives ("Billys")

Gobert US News & World Report, 2016; Li, et al 2017, 2018

Billys (False Negatives) cont.

- Ran a successful investigation and analyzed the data correctly, but did not know what to write:

CER	Writing
Claim	I already did this part
Evidence	the answer to this is in my claim. I think
Reasoning	I don't know

RUTGERS

Johnnys (False Positives)

- Wrote about an experiment that he did not conduct:

Doing	Writing
(table 1 below)	Claim: The shape of the container the liquid is in does not affect the overall density of the liquid.
(table 2 below)	Evidence: if you change the shape of the container to wide to square the

Doing (1):

Select	Trial #	Container Shape	Liquid Type	Container Filled To	Liq Ma (
✓	1	narrow	oil	quarter	21
✓	2	narrow	water	quarter	2
✓	3	narrow	alcohol	quarter	1

Writing (1):

Claim:

The shape of the container the liquid is in does not affect the overall density of the liquid.

Doing (2):

Select	Trial #	Container Shape	Liquid Type	Container Filled To	Liqu Mas (g
✓	1	wide	oil	quarter	212
✓	2	wide	oil	quarter	212
✓	3	wide	oil	quarter	212

Writing (2):

Evidence:

if you change the shape of the container to wide to square the

RUTGERS

Johnnys (False Positives)

- Miscalculated data, but reported correct data:

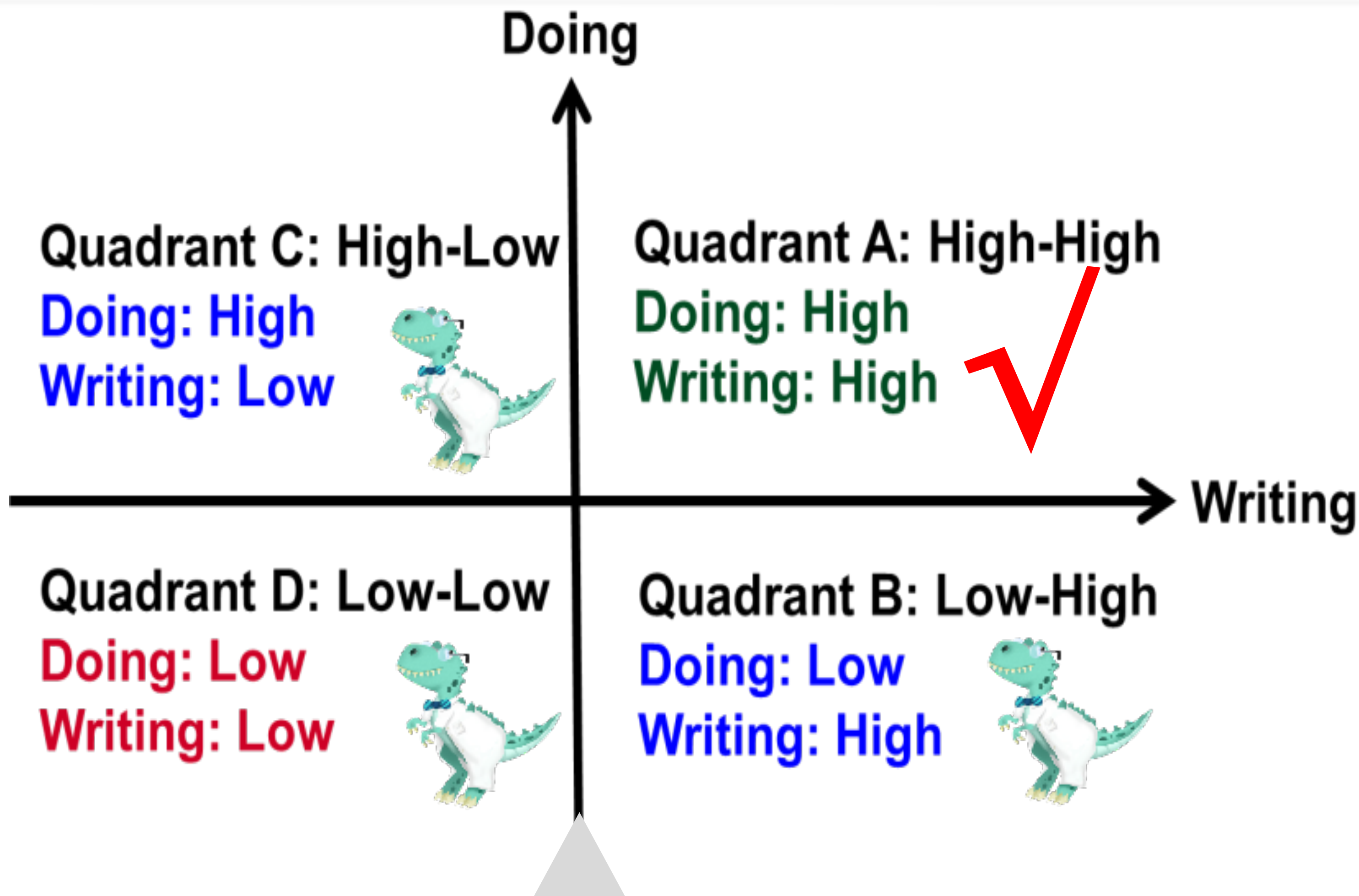
Doing

Trial #	Container Shape	Liquid Type	Container Filled To	Liquid Mass (g)	Liquid Volume (ml)	Liquid Density (g/ml)
1	narrow	oil	quarter	212.5	250	62500
2	narrow	oil	quarter	212.5	250	53125

Written Claim

When the container was both narrow and wide the density was 0.85.

RUTGERS



Discrepancy (30-60%) can exist between skills for doing inquiry and skills for explaining their inquiry.

Gobert, 2016; Li, Gobert, Dickler, 2017, 2018

Discussion: Study 1

- Our findings demonstrated that a messy middle (Gotwals & Songer, 2010) exists in students' inquiry performances,
  - *these can result in false negatives and false positives if there is an over-reliance on written assessments.*
- When replication, a misalignment in competencies was found for between 30-60% of students
  - *writing alone may not be sufficient for a number of reasons (poor communication skills or parroting)*
  - *what people say or write is not necessarily ground truth about what they know or can do, especially in STEM!*
- Assessments need to assess the full complement of inquiry practices expected in science policy documents such as the Next Generation Science Standards (2013)

RUTGERS

Study 2: Scaffolding in Inq-ITS

- Rex supports students on common difficulties with practices, triggered by algorithms in real time, when student needs it (not on-demand)
  - *Because students don't know they need help*
- Delivered via a pedagogical agent named Rex, a cartoon dinosaur
- Provide students with multiple levels of support

(Gobert et al., 2013; Moussavi et al., 2016; Sao Pedro et al., 2013; Gobert et al., in press)

Participants

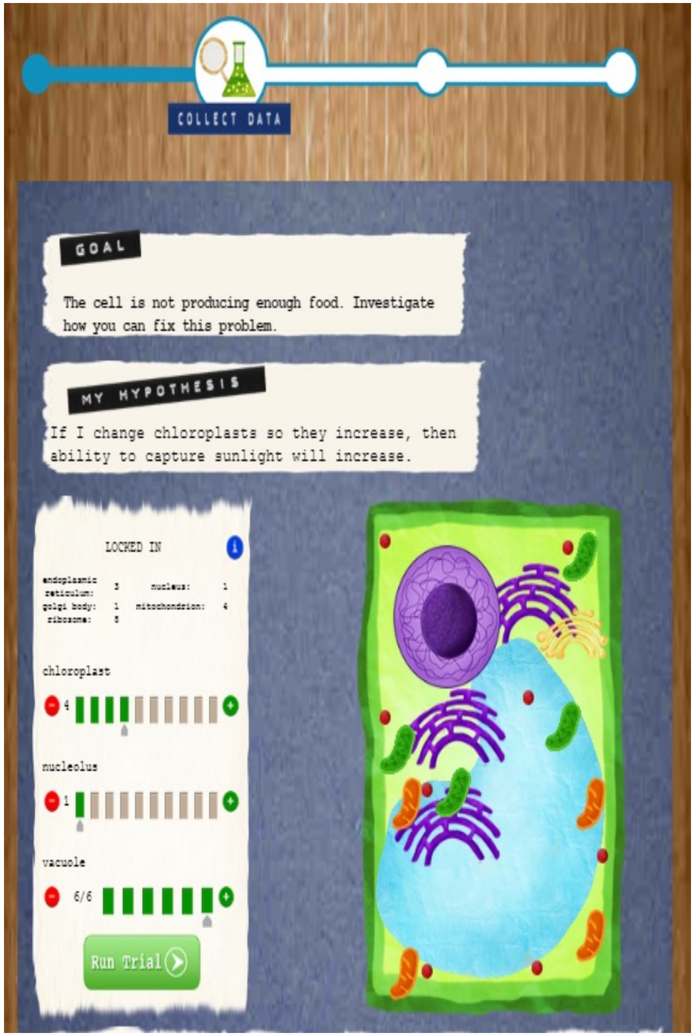
- 107 6th grade students completed four Inq-ITS virtual labs over the course of 170 days:

Materials: Inq-ITS

- Students' complete four inquiry stages in each virtual lab investigation:
  - Asking Questions/Hypothesizing
  - Carrying Out Investigations/Collecting Data
  - Analyzing and Interpreting Data
  - Explaining findings

- Each Inq-ITS lab (Animal Cell, Plant Cell, Genetics, Natural Selection) has 3-4 driving question activities

- Students received scaffolding **only** in the Animal Cell activity



(Gobert et al., 2013, 2014, 2018; Moussavi et al., 2016; Moussavi, 2018; Sao Pedro et al., 2013; Sao Pedro, 2013)

Materials: Inq-ITS Scaffolding in Animal Cell

The scaffolds become increasingly specific:

- Orient students to the current task
- Remind students of the steps to engage in the practice
- Give students necessary conceptual information
- Provide direct instructions on how to complete steps

"Check your claim. Remember the point of your experiment is to evaluate your hypothesis"

↓

"Make sure the variables in your hypothesis match those in your claim."

↓

"Your claim independent variable should be [IV]. Your claim dependent variable should be [DV]. This makes it so your claim matches your hypothesis."

(Anderson et al., 1995; Corbett & Anderson, 1995; Koedinger & Corbett, 2006)

Measures

- Students' inquiry performance is captured and assessed at the practice level (and sub-practice level) using KE- and EDM-based algorithms:
  - *Forming Questions/Hypothesizing*
  - *Carrying Out Investigations/Collecting Data*
  - *Analyzing and Interpreting Data*
  - *Warranting Claims*

Gobert et al., 2013, 2014, 2018; Moussavi et al., 2016; Moussavi, 2018; Sao Pedro et al., 2013; Sao Pedro, 2013

Analyses

- Repeated Measures ANOVAs were performed to investigate whether students' performance on each of the four inquiry practices was robust over time and across topics after scaffolding was removed

Animal Cell (Scaffolding)
- Time 1

→ 40 days →

Plant Cell
- Time 2

→ 40 days →

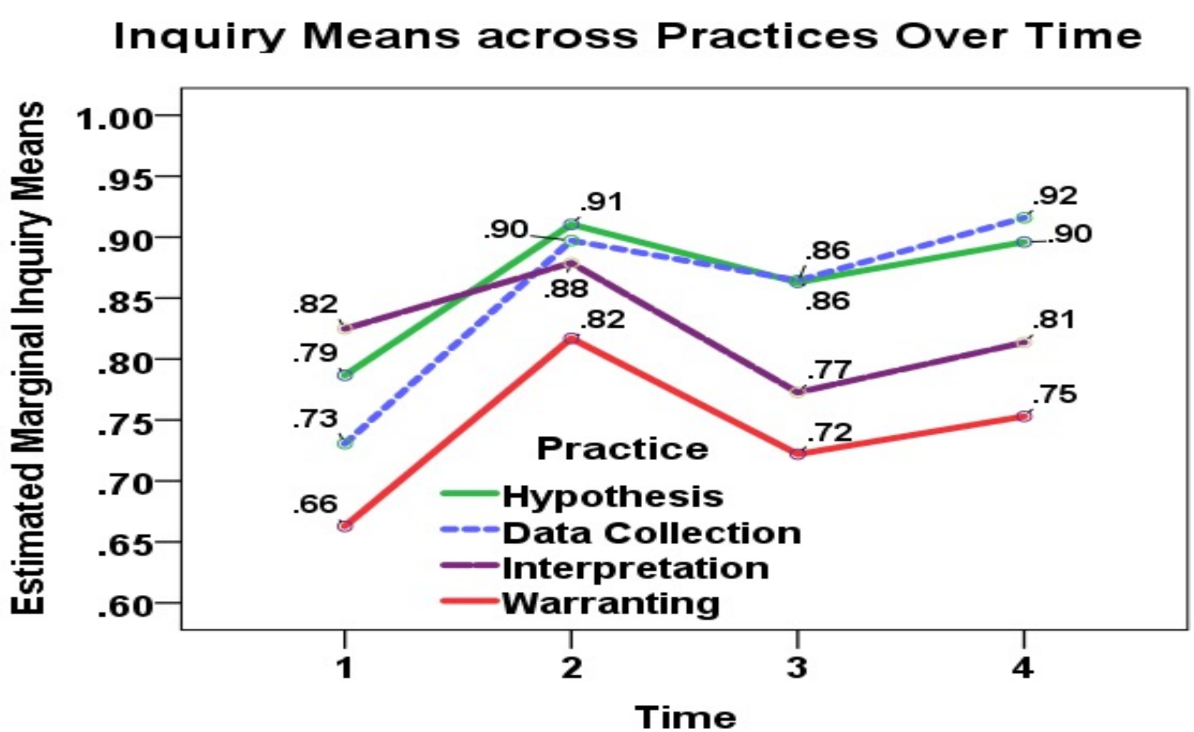
Genetics
- Time 3

→ 90 days →

Natural Selection
- Time 4

Results: Time x Practices

- Repeated measures multivariate analyses showed a significant two-way interaction between time and inquiry practice: $F(9, 98) = 11.00$, $p < .001$, $\eta 2 = .503$
- Tests of within-subjects effects were also significant for this interaction: $F(9, 954) = 9.28$, $p < .001$, $\eta 2 = .080$



Discussion: Study 2

- Task design, infrastructure, & algorithms enable real-time assessment & scaffolding by Rex
  - Targeted scaffolding in one topic *can* benefit student inquiry practices even after scaffolding is removed
  - Robust across topics and time (tested 170 days later)
  - Scaffolding for inquiry practices by Rex greatly supports the acquisition and refinement of competencies, which undergirds students' inquiry performance

- Sao Pedro, 2013; Moussavi, 2016; Gobert et al., 2018; Li et al, 2018; Gobert et al., in press

Study 3: Using the Inq-Blotter Dashboard to Support Teachers and Students on Science Practices

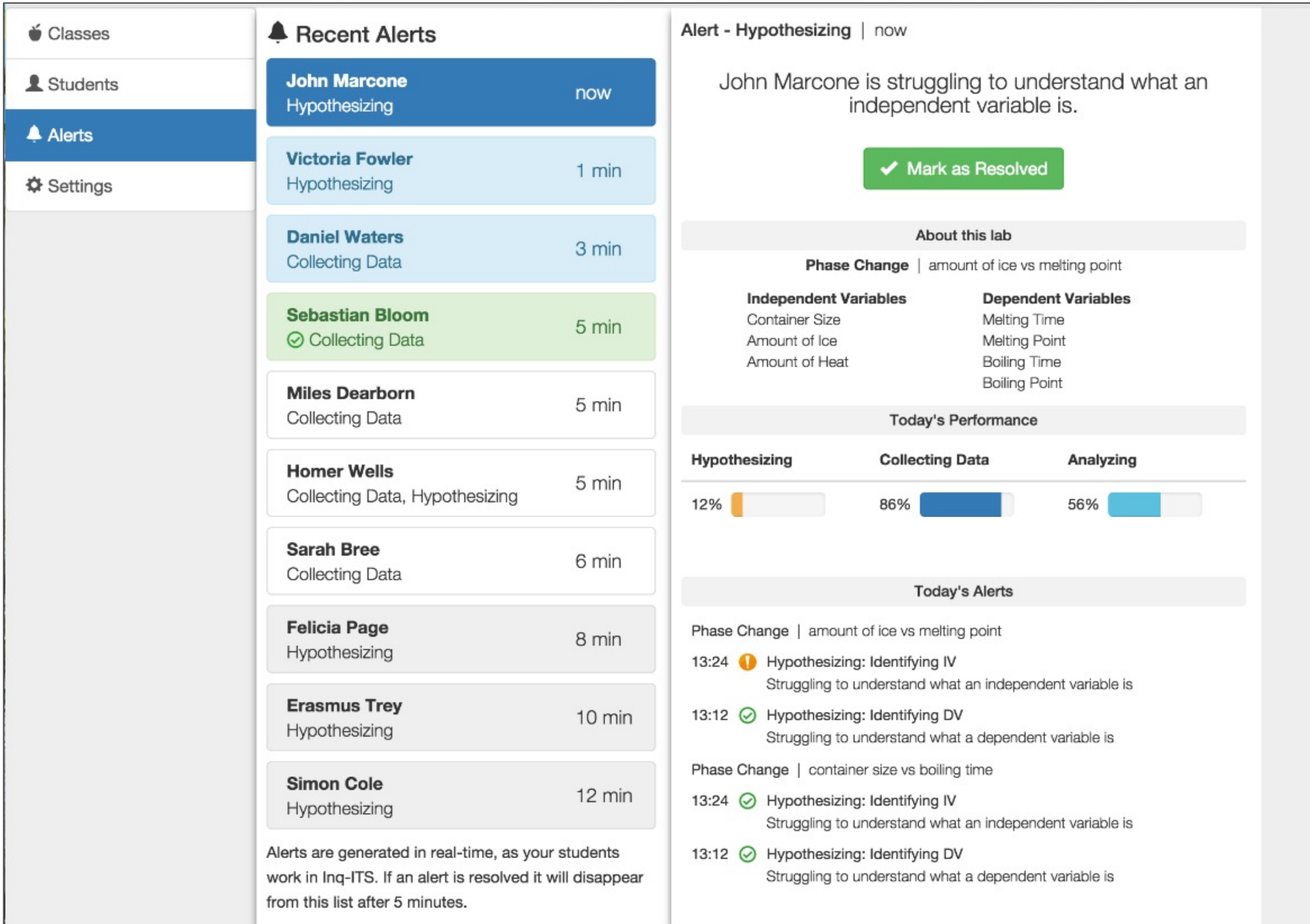
- There are many dashboards that use coarse-grained multiple choice items to help teachers…
- Few dashboards have the capacity to fully assess students' inquiry competencies at a fine-grained level
  - *(Lajoie et al., 2020; Martinez-Maldonado et al., 2015; Matuk et al., 2016; Tissenbaum & Slotta 2019; VanLehn et al., 2019).*

Inq-ITS provides fine-grained alerts & TIPS based on AI-based assessments, important for real time instruction of NGSS practices because….

- Teachers need formative data to inform instruction (Lehrer & Schauble; Berland & McNeill, 2010; Gotwals & Anderson, 2015) on-the-fly
- Teachers need guidance regarding how to use formative assessments (Hammer et al., 2012; Furtak et al., 2014)
- Teachers need information and data at the right grain-size in order for it to be pedagogically useful
  - *Finer-grained data has been found to better support teachers' instruction (Alonso, 2011; Alonso & Elby, 2009, 2011; Furtak,2012; Furtak et al, 2014; Harris, Krajcik et al., 2016; Corcoran et al., 2019).*

RUTGERS

Components for Teachers: Alerts



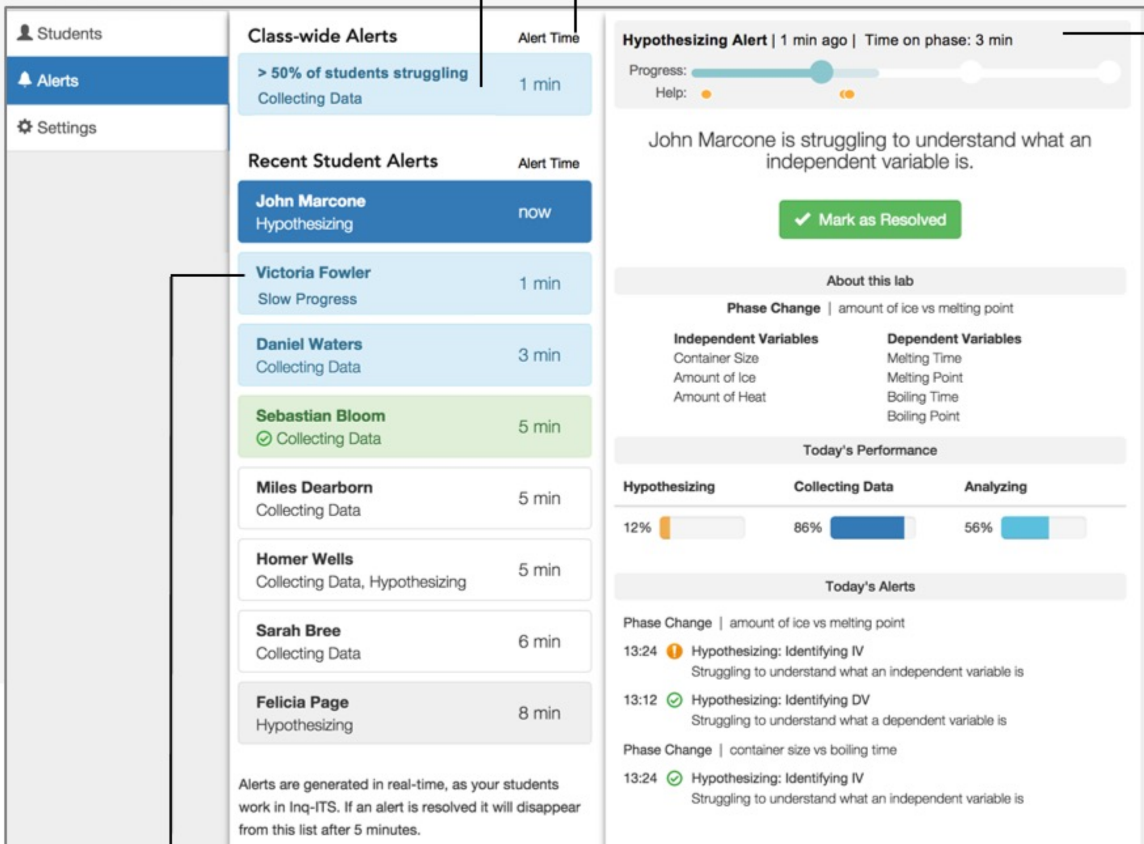
Class-wide Alerts

Triggers when any subskill falls below the limit for enough students. The limit and percentage of students required to trigger an alert is set in the settings, and defaulted to "50% of students falling below 60% performance"

If there are no class-wide alerts, this section should not be visible (no header, etc)

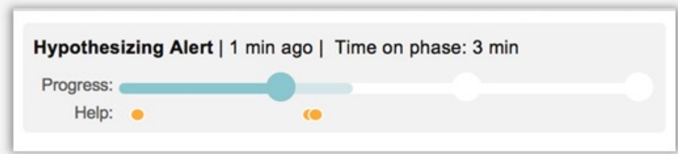
New "Alert Time" label

Added to help clarify what the time displayed is. This should be tested for usability, and may need iteration



Updated Header

- Added "time on phase" which should reflect the amount of time the student has been working on the phase in question
- Added Progress Bar, showing where a student is in the activity set
- Added help indicator dots, show how much help the students received and when.



Shows completed phases within an activity

Indicates a completed activity

Indicates the current phase

Student received help from rex (dots spaced by phase)

Student received help from rex twice that phase

Student received help from rex 3 times that phase (etc…)

Slow Progress Alerts

If a student spend too much time on one phase, the teacher is alerted. This limit is set in settings.

RUTGERS

Study 3: Testing Inq-Blotter

RQ1) Are real-time alerts for inquiry practices associated with student improvement?

RQ2) Does the pattern of teacher support provided to students differ in relation to performance on practices?

(Dickler, Gobert, & Sao Pedro, JLA,2021)

Methods

- Participants:
  - 2 middle school teachers
  - 211 middle school students

- Procedure:
  - Students completed three Inq-ITS lab activities
  - Teachers used Inq-Blotter as students completed Inq-ITS labs
    - *Audio data of interactions were recorded*
    - *Inq-ITS triangulates all teacher blotter data with all students' Inq-ITS data + voice recording (for research)*

Measures: Inq-ITS and Inq-Blotter Log Data

- Students' competencies were scored using our assessment algorithms for each practice *
- Log data from Inq-Blotter was examined in terms of:
  - *alerts that appeared for the teacher*
  - *the student alerts accessed by the teacher*
  - *the content of alerts*
  - *timestamps*
- *N* = 35 recordings were captured and transcribe
- Teacher turns were coded by two raters for types of supports provided
  - *i.e., science practices v. content, evaluative, etc.*

* Gobert et al., 2013, 2018; Dickler et al., JLA 2020

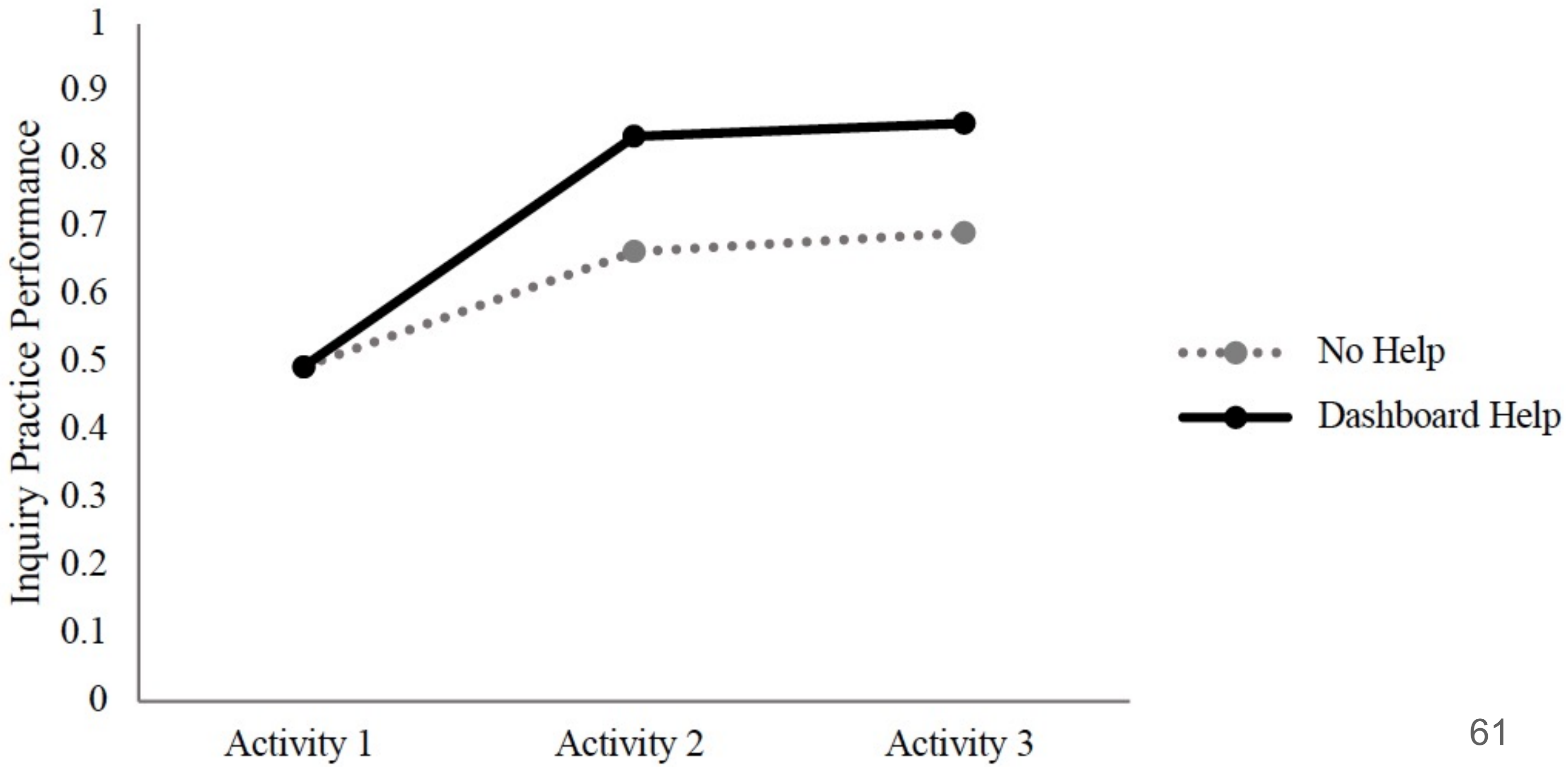
Analyses: RQ1

RQ1) Are real-time alerts for inquiry practices associated with student improvement*?

- Triangulated log data from Inq-ITS and Inq-Blotter
  - *Identified students who were helped (n = 35 students) and matched students who were **not** helped (n = 35 students)*
- A Mixed Model Analysis of Variance (MM ANOVA) was used to compare student performance across activities between conditions
  - *i.e., help versus no help*

Results: RQ1

- The MM ANOVA revealed that students helped based on an alert had marginally significantly greater improvement across activities on the practice on which they were helped
  - i.e., interaction effect, $F(2, 136) = 2.60$, $p = 0.078$



Analyses: RQ2

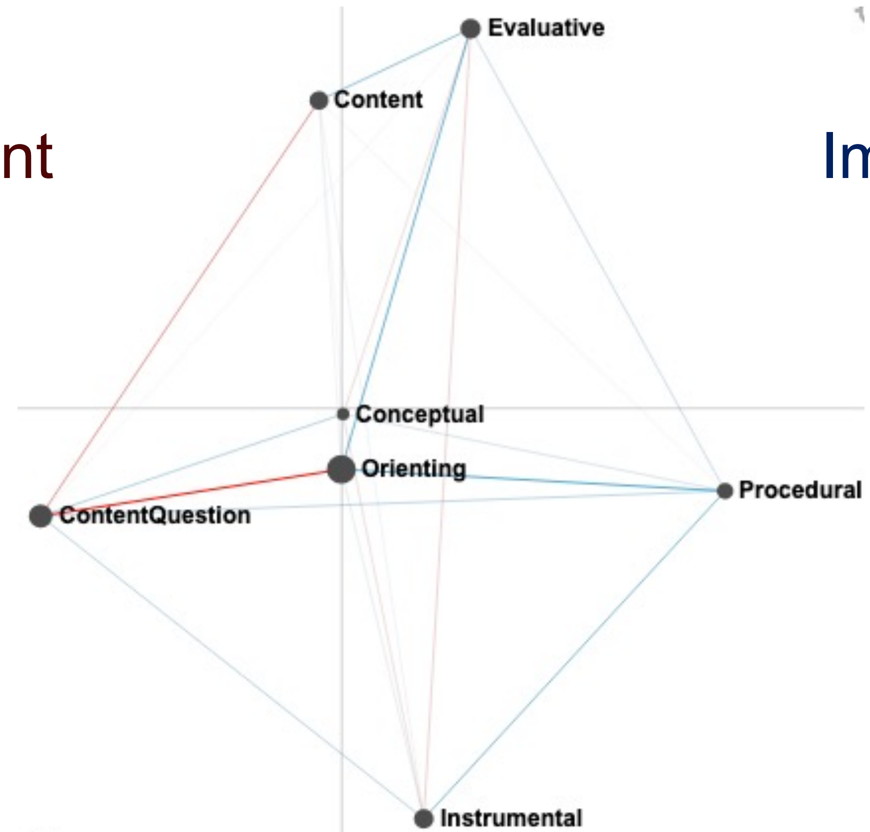
RQ2) Does the pattern of teacher support provided to students differ in relation to performance on practices?

- Triangulated log data with coded audio transcripts
- Compared patterns in support when helped students improved or did not improve on their next activity using ENA

Results: RQ2 (continued)

- Students who did *not* improve received combinations of lower-level/content supports more frequently

No Improvement



Improvement

Dickler, Gobert, & Sao Pedro, JLA, 2020

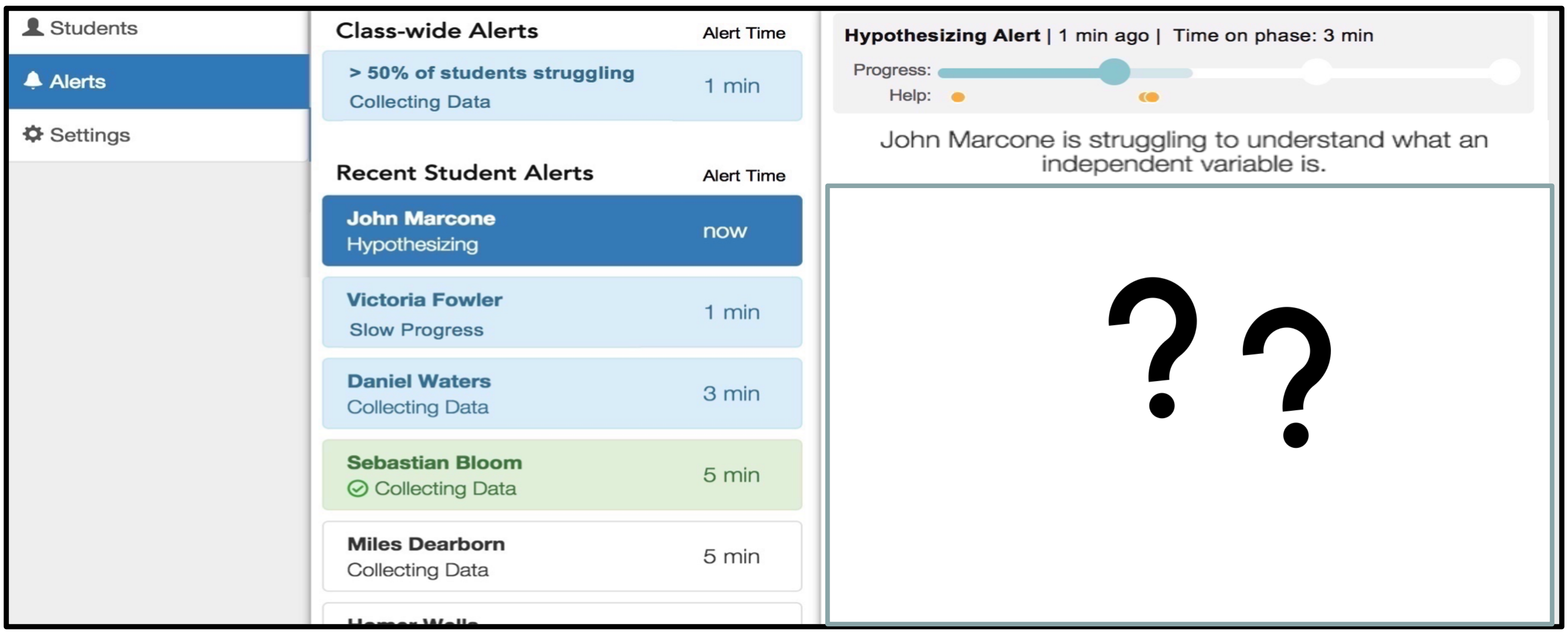
Results: RQ2

- ENA revealed the pattern of support associated with improvement was significantly different, $t(34) = 2.45$, $p = .04$
- *Those who showed improvement received more high level support, i.e. procedural & conceptual support.*
- *Those who did not improve received* combinations of lower-level & *content* support.
- These findings have important implications for designing *alerts to promote explicit practice support*
  - *Prior studies indicate potential of providing teachers with example prompts to guide interactions (e.g., Morris & Chi, 2010)*

Dickler, Gobert, & Sao Pedro, JLA, 2020

Supporting Teachers with TIPS

- Since Study 3 showed that discourse support of specific types can lead to student improvement, we added TIPS (Teacher Inquiry Practice Supports) to our alerting dashboard
- Productive discourse was mined from our previous study to develop TIPS.



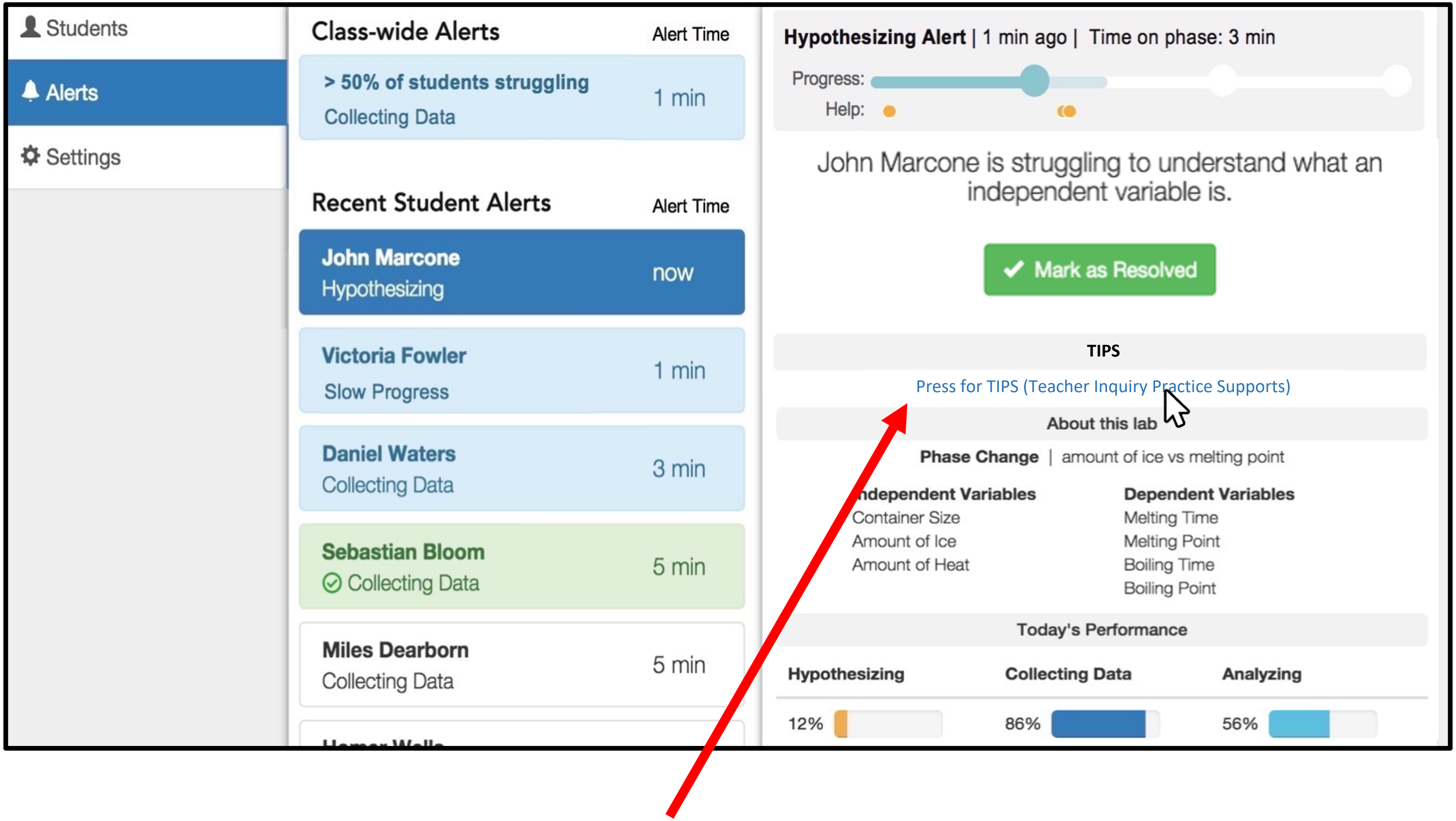
TIPS Development

- **T**eacher **I**nquiry **P**ractice **S**upports – prompts for teachers to support the student's inquiry practices
  - *TIPS are sent directly to the teacher within alerts in Inq-Blotter*

(Adair et al., 2020; Gobert et al., 2018)

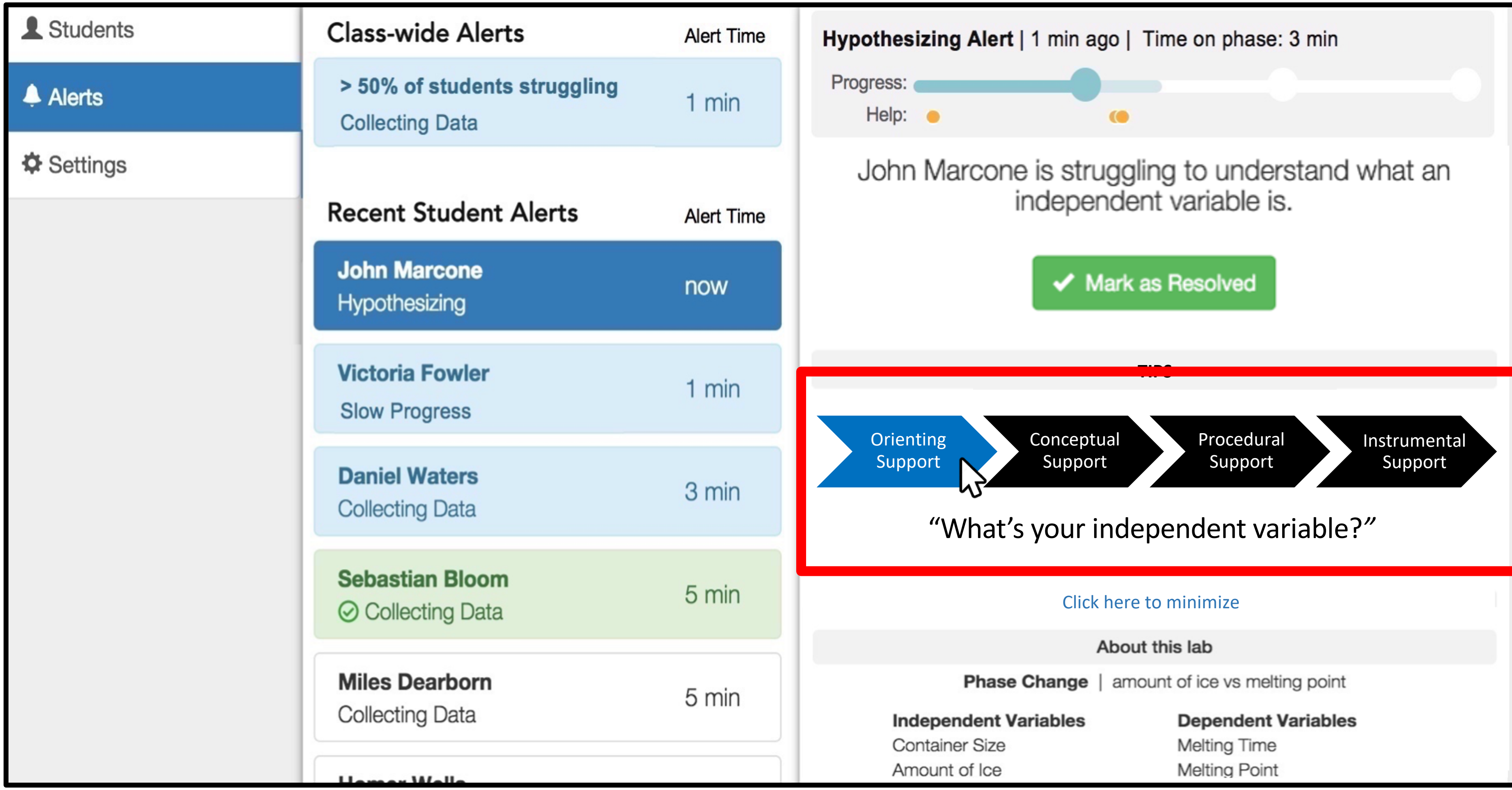
TIPS were added to Inq-Blotter; Development included

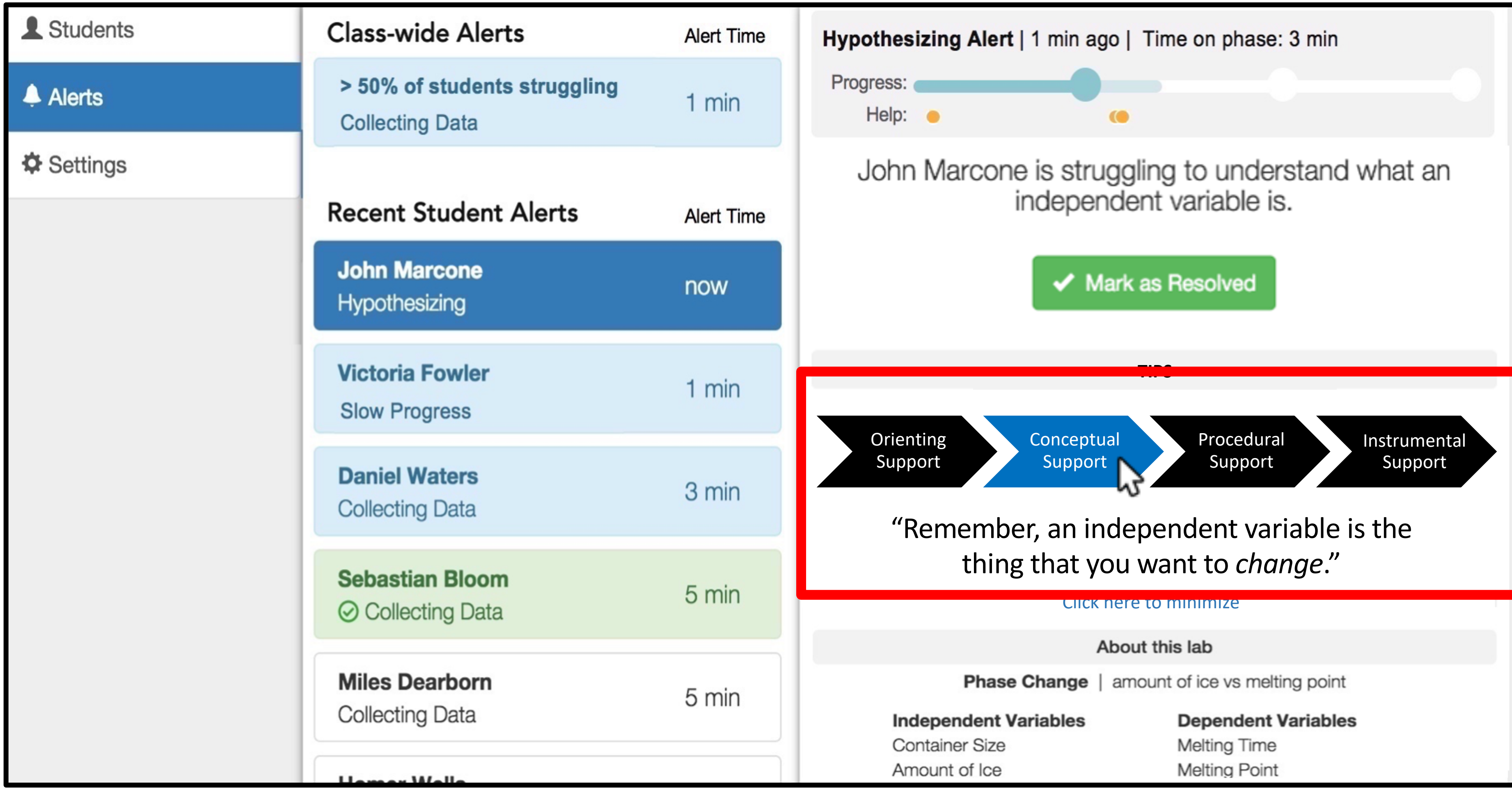
- Obtained 219 teacher-spoken segments from recorded conversations with the 2 middle school teachers from Study 2
- Used segments that had previously been coded for four categories of support (i.e., orienting, conceptual, instrumental, procedural)
- Filtered segments for which students improved on the practice after receiving support from the teacher
- Constructed TIPS for each category of support based on filtered teacher segments
- Embedded TIPS into the Inq-Blotter system

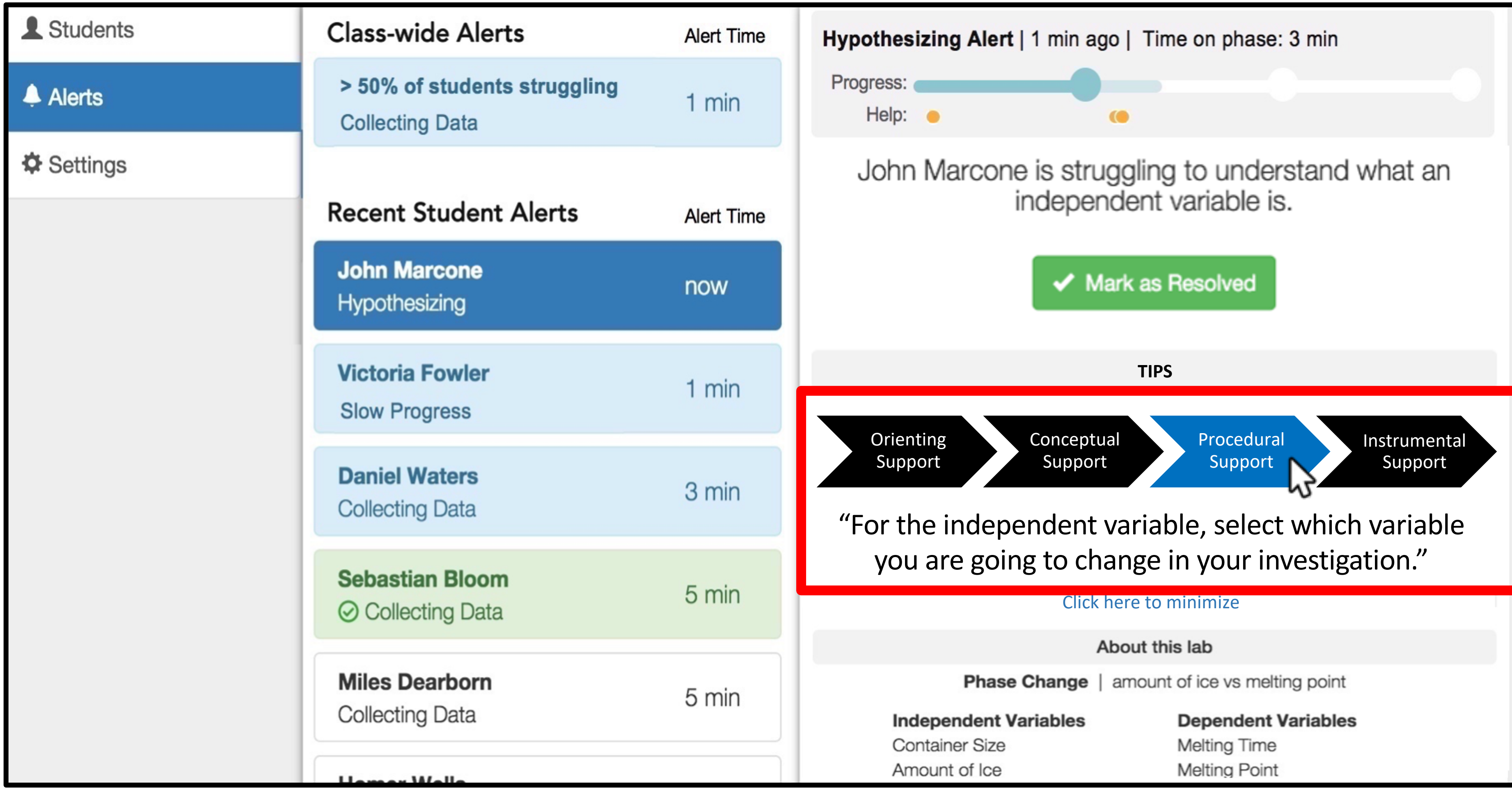


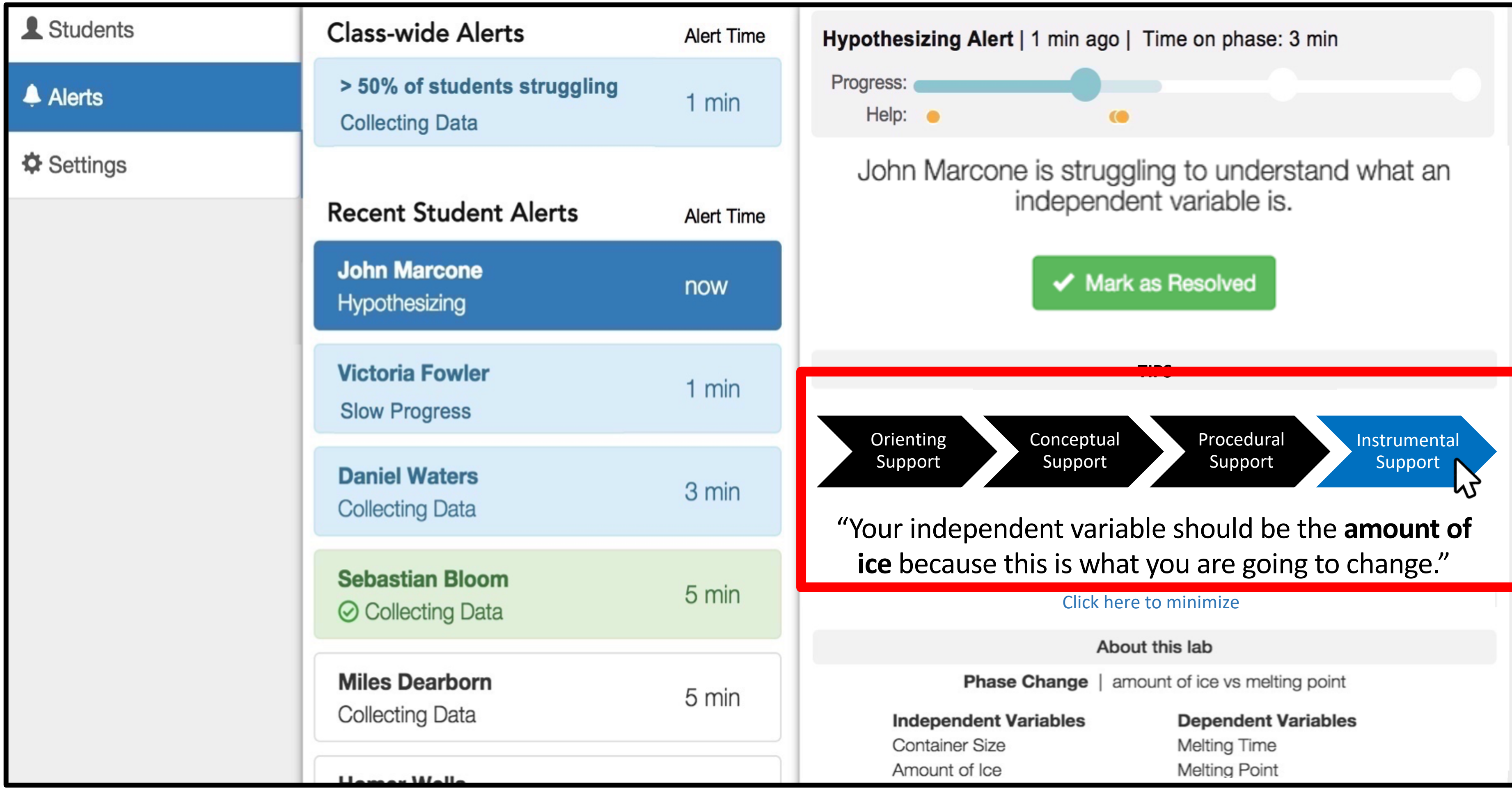
Students
Alerts
Settings
Class-wide Alerts
Alert Time
> 50% of students struggling
Collecting Data
1 min
Recent Student Alerts
Alert Time
John Marcone
Hypothesizing
now
Victoria Fowler
Slow Progress
1 min
Daniel Waters
Collecting Data
3 min
Sebastian Bloom
Collecting Data
5 min
Miles Dearborn
Collecting Data
5 min
Hypothesizing Alert | 1 min ago | Time on phase: 3 min
Progress:
Help:
John Marcone is struggling to understand what an independent variable is.
Mark as Resolved
TIPS
Press for TIPS (Teacher Inquiry Practice Supports)
About this lab
Phase Change | amount of ice vs melting point
Independent Variables
Container Size
Amount of Ice
Amount of Heat
Dependent Variables
Melting Time
Melting Point
Boiling Time
Boiling Point
Today's Performance
Hypothesizing
12%
Collecting Data
86%
Analyzing
56%

RUTGERS


Students
Alerts
Settings
Class-wide Alerts
Alert Time
> 50% of students struggling
Collecting Data
1 min
Recent Student Alerts
Alert Time
John Marcone
Hypothesizing
now
Victoria Fowler
Slow Progress
1 min
Daniel Waters
Collecting Data
3 min
Sebastian Bloom
Collecting Data
5 min
Miles Dearborn
Collecting Data
5 min
Hypothesizing Alert | 1 min ago | Time on phase: 3 min
Progress:
Help:
John Marcone is struggling to understand what an independent variable is.
Mark as Resolved
Orienting Support
Conceptual Support
Procedural Support
Instrumental Support
"What's your independent variable?"
Click here to minimize
About this lab
Phase Change | amount of ice vs melting point
Independent Variables
Container Size
Amount of Ice
Dependent Variables
Melting Time
Melting Point


Students
Alerts
Settings
Class-wide Alerts
Alert Time
> 50% of students struggling
Collecting Data
1 min
Recent Student Alerts
Alert Time
John Marcone
Hypothesizing
now
Victoria Fowler
Slow Progress
1 min
Daniel Waters
Collecting Data
3 min
Sebastian Bloom
Collecting Data
5 min
Miles Dearborn
Collecting Data
5 min
Hypothesizing Alert | 1 min ago | Time on phase: 3 min
Progress:
Help:
John Marcone is struggling to understand what an independent variable is.
Mark as Resolved
Orienting Support
Conceptual Support
Procedural Support
Instrumental Support
“Remember, an independent variable is the thing that you want to change.”
About this lab
Phase Change | amount of ice vs melting point
Independent Variables
Container Size
Amount of Ice
Dependent Variables
Melting Time
Melting Point


Students
Alerts
Settings
Class-wide Alerts
Alert Time
> 50% of students struggling
Collecting Data
1 min
Recent Student Alerts
Alert Time
John Marcone
Hypothesizing
now
Victoria Fowler
Slow Progress
1 min
Daniel Waters
Collecting Data
3 min
Sebastian Bloom
Collecting Data
5 min
Miles Dearborn
Collecting Data
5 min
Hypothesizing Alert | 1 min ago | Time on phase: 3 min
Progress:
Help:
John Marcone is struggling to understand what an independent variable is.
Mark as Resolved
TIPS
Orienting Support
Conceptual Support
Procedural Support
Instrumental Support
"For the independent variable, select which variable you are going to change in your investigation."
Click here to minimize
About this lab
Phase Change | amount of ice vs melting point
Independent Variables
Container Size
Amount of Ice
Dependent Variables
Melting Time
Melting Point

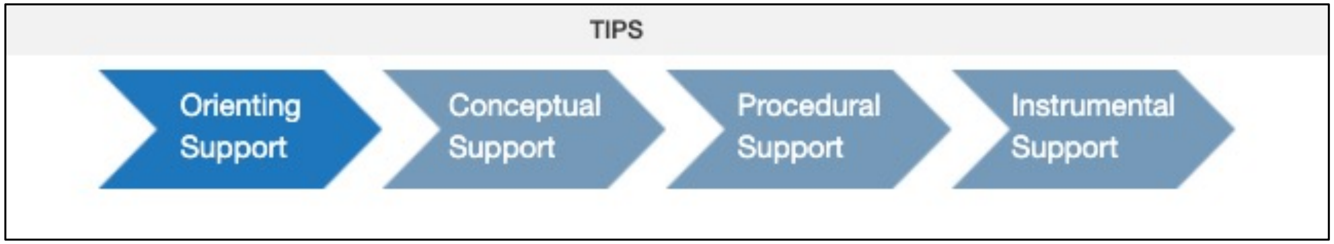

Students
Alerts
Settings
Class-wide Alerts
Alert Time
> 50% of students struggling
Collecting Data
1 min
Recent Student Alerts
Alert Time
John Marcone
Hypothesizing
now
Victoria Fowler
Slow Progress
1 min
Daniel Waters
Collecting Data
3 min
Sebastian Bloom
Collecting Data
5 min
Miles Dearborn
Collecting Data
5 min
Hypothesizing Alert | 1 min ago | Time on phase: 3 min
Progress:
Help:
John Marcone is struggling to understand what an independent variable is.
Mark as Resolved
Orienting Support
Conceptual Support
Procedural Support
Instrumental Support
"Your independent variable should be the amount of ice because this is what you are going to change."
Click here to minimize
About this lab
Phase Change | amount of ice vs melting point
Independent Variables
Container Size
Amount of Ice
Dependent Variables
Melting Time
Melting Point

Methods

- Participants:
  - 4 teachers from different schools
    - 2 Remote (Fully Online, Synchronous)
    - 1 In-Person/Traditional
    - 1 Hybrid
- Procedure:
  - Teachers used **Inq-Blotter with TIPS** as students completed Inq-ITS labs; Clickstream data of the types of alerts and supports that teacher selected and timestamps; Audio recordings
  - Teachers were interviewed about their experiences

Teacher Interviews – Theme 1

- TIPS helped teachers differentiate levels of support
  - *“In general, it was **helpful to remind me to not jump straight to giving kids the answer**. I had a few kids surprise me. They figured things out on their own using the TIPS more often than I thought they would.”*
  - *"I talk to my kids all the time, but it made it easier to identify like a **laser what I needed to talk to them about**."*



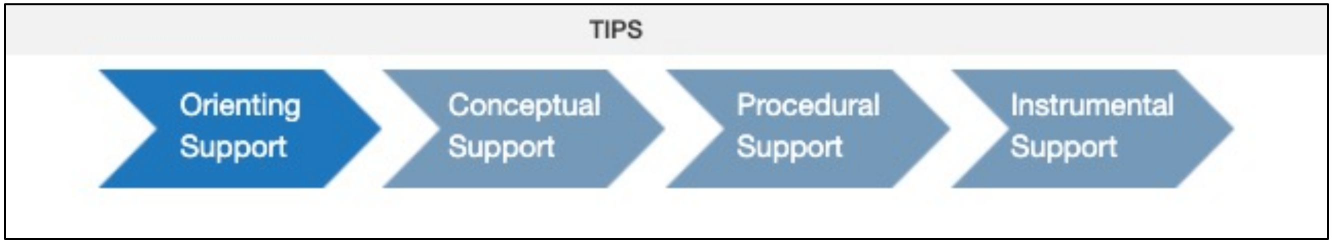
Teacher Interviews– Theme 1, cont'd

- TIPS helped teachers differentiate levels of support
- Teacher quotes:
  - *"I use conceptual TIPS to* ***students who are more skilled*** *at science to prompt them to think at the next level. For example, if they have changed too many variables at once, I say "How will you know what caused the change in your dependent variable?"*
  - *"I typically use instrumental-type hints for students who are in Special Ed or on IEPs. For example, I might say, change only one variable at a time, then I ask them t explain to me why this permits them to know how the change in the independent variable lead to changes in the dependent variable."*

RUTGERS

Teacher Interviews– Theme 2

- TIPS helped teachers with timeliness
  - "The **TIPS saved me time** to clarify what is going on...I was able to make my way around the room to more students. When you add [that] up...it really saves me time."
  - "[TIPS] helped me with **starting that communication with the students**. How much did that decrease the amount of time? Probably 1-2 minutes. I get those TIPS, and that's what I would send the kids online."



Discussion

- These preliminary results suggest that adding TIPS to the alerting dashboard helps inform teachers' instruction of inquiry practices, as expected by NGSS.

- Our early findings also suggest the TIPS allows teachers to:
  - *help more students,*
  - *help them more efficiently,*
  - *give them precisely the help they need (Sao Pedro et al., 2019)*
  - *Teachers' help improves students' transfer on the practice on which they were helped (Dickler et al., JLA)*

- Important to NGSS instruction, TIPS gives teachers powerful data and actionable TIPS to differentiate help to students who have different needs with science inquiry.

Overall Implications for Assessment & Instruction

- This approach, based on the assessment triangle and Evidence Centered Design & data mining, can inform the design of future *scalable* assessments for science inquiry practices (Mislevy et al, 2012; Mislevy, Yan, Gobert & Sao Pedro, 2020), important to rigorous performance assessment & instruction.

- Is also very relevant & useful to teachers' formative assessment and instruction for NGSS, and in turn, students's learning of the practices.

RUTGERS

Inq-ITS Teams

@ RU
Mike Brigham, Software Engineer
Amy Adair, Doctoral Student
Mariel O'Brien, Doctoral Student
Jeremy Lee, Doctoral Student
Jessica Owens, Doctoral Student
Ellie Segan, Doctoral Student

Haiying Li, Former Post Doc
Victor Brusca, Former Developer
Rachel Dickler, Former Ph.D. Student

@ Apprendis
Michael Sao Pedro, CTO
Cam Betts, Managing Director
Jason Kleban, Software Engineer

Brett Craig & Kaitlyn Cuff, & Parker Shannon, Partnerships
Denise Sao Pedro, Project Success

RUTGERS

- Thank you!
- Janice.gobert@gse.rutgers.edu
- Janice@apprendis.com

- For research papers, see Inqits.com