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
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
Context: US & many others face a STEM crisis

Int'l rank on science (PISA), 2018

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

U.S.A.
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)
Our Goal

• We wanted to rigorously assess science practices, at scale

• We wanted to better support teachers:
  – In their instruction of science practices
  – In their assessment of science practices

• We wanted to better support students:
  – In their learning of science practices
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)
  - 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

- **Data-mined and knowledge-engineered rules**
  - to assess & scaffold students’ practices, and alert teachers on who needs help & how
Inq-ITS overview & components

FOR TEACHERS: ACTIONABLE REPORTS & IMMEDIATE ALERTS

REPORTS ALERTS & TIPS

FOR STUDENTS: REAL TIME FEEDBACK JUST WHEN IT IS NEEDED

VIRTUAL TUTOR

Gobert, Sao Pedro, Baker, & Betts, US Patent nos. 9373082, 9564057, 10186168
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*
  – *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).
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, taken together, these two frameworks provide 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
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’ processes (Aulls, 2002)
  - Real time assessment and monitoring can help this, and improve students’ learning
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 (HPL)**
  - 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 use
  - 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

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

It looks like you did great at designing a controlled experiment, but let me remind you to collect data to help you test your hypothesis.

- How do I do that?
- Show me my hypothesis.
- OK.
Components for Teachers

Real-Time Reports for Assessment

Real-Time Alerts for Instruction
Alerts are generated in real-time, as your students work in Inq-ITS. If an alert is resolved it will disappear from this list after 5 minutes.
Components for Teachers: TIPS (Teacher Inquiry Practice Supports)

Remember, you should be running controlled trials. Controlled trials is when you only change one variable at a time.
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

Study 1: Identifying the messy middle
“Doing” Science vs Explaining Science

Doing

• Generating a research question
• Forming a hypothesis
• Collecting, analyzing, & interpreting data
• Selecting data to warrant a claim
• Etc.

NGSS, 2013

Writing

• Claim
• Evidence
• Reasoning
• etc.

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

212.5 g

212.5 g

212.5 g
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)
Analyses

- K-means clustering was used to group students
- 144 of 293 students (49%) were in the messy middle:
  - Low Doing – High Writing (N = 131)
  - High Doing – Low Writing (N = 13)

False Positives ("Johnnys")

False Negatives ("Billys")

Billys (False Negatives) cont.

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

<table>
<thead>
<tr>
<th>CER</th>
<th>Writing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Claim</td>
<td><em>I already did this part</em></td>
</tr>
<tr>
<td>Evidence</td>
<td><em>the answer to this is in my claim. I think</em></td>
</tr>
<tr>
<td>Reasoning</td>
<td><em>I don't know</em></td>
</tr>
</tbody>
</table>
### Johnnys (False Positives)

- Wrote about an experiment that he did not conduct:

<table>
<thead>
<tr>
<th>Select</th>
<th>Trial #</th>
<th>Container Shape</th>
<th>Liquid Type</th>
<th>Container Filled To</th>
<th>Liquid Mass (g)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>narrow</td>
<td>oil</td>
<td>quarter</td>
<td>212</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>narrow</td>
<td>water</td>
<td>quarter</td>
<td>212</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>narrow</td>
<td>alcohol</td>
<td>quarter</td>
<td>12</td>
<td></td>
</tr>
</tbody>
</table>

**Claim:**

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

**Evidence:**

If you change the shape of the container to wide to square the
Johnnys (False Positives)

- Miscalculated data, but reported correct data:

<table>
<thead>
<tr>
<th>Trial #</th>
<th>Container Shape</th>
<th>Liquid Type</th>
<th>Container Filled To</th>
<th>Liquid Mass (g)</th>
<th>Liquid Volume (ml)</th>
<th>Liquid Density (g/ml)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>narrow</td>
<td>oil</td>
<td>quarter</td>
<td>212.5</td>
<td>250</td>
<td>62500</td>
</tr>
<tr>
<td>2</td>
<td>narrow</td>
<td>oil</td>
<td>quarter</td>
<td>212.5</td>
<td>250</td>
<td>53125</td>
</tr>
</tbody>
</table>

Written Claim

When the container was both narrow and wide the density was 0.85.
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)
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)
  • 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:

Animal Cell + Scaffolding → Plant Cell → Genetics → Natural Selection

40 days 40 days 90 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 Animal Cell Virtual Activity
Ask Questions/Hypothesize

**Goal**
The golgi body is not receiving enough protein. Investigate how you can fix this problem.

**What I Will Change**
- endoplasmic reticulum
- produces proteins
- transports materials outside cell
- golgi body
- ribosome
- transports protein to golgi bodies

**What Will Happen**
- transports protein to golgi bodies
- ribosome
- golgi body
- transports materials outside cell
- produces proteins
- endoplasmic reticulum
Collect Data

**Goal**
The golgi body is not receiving enough protein. Investigate how you can fix this problem.

**My Hypothesis**
If I change the endoplasmic reticula so they increase, then the transportation of protein to golgi bodies will increase.

---

**Locked In**
nucleus: 1  lysosome: 2
mitochondrion: 4  nucleolus: 1
vacuole: 1

endoplasmic reticulum: 8

golgi body: 9
ribosome: 9

Run Trial
Analyze and Interpret Data

**WHAT I CHANGED**
- endoplasmic reticulum
- produces proteins
- transports materials outside cell
- golgi body
- ribosome
- transports protein to golgi bodies

**WHAT HAPPENED**
- transports protein to golgi bodies
- ribosome
- golgi body

**MY ANALYSIS**
- What I observed:
  - Supports my hypothesis
  - Does not relate to my hypothesis
  - Refutes my hypothesis
## Evidence

These trials are evidence of my claim: 4, 5, 6,

<table>
<thead>
<tr>
<th>Select</th>
<th>Trial #</th>
<th>Endoplasmic Reticulum</th>
<th>Golgi Bodies</th>
<th>Ribosomes</th>
<th>Production of Proteins</th>
<th>Transportation of Protein to Golgi Bodies</th>
<th>Transportation of Nutrients Outside Cell</th>
<th>Cell Health</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>8</td>
<td>100%</td>
<td>133%</td>
<td>400%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>5</td>
<td>63%</td>
<td>133%</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>113%</td>
<td>267%</td>
<td>900%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>113%</td>
<td>267%</td>
<td>900%</td>
<td></td>
</tr>
<tr>
<td>✓</td>
<td>5</td>
<td>5</td>
<td>9</td>
<td>9</td>
<td>113%</td>
<td>167%</td>
<td>900%</td>
<td></td>
</tr>
<tr>
<td>✓</td>
<td>6</td>
<td>2</td>
<td>9</td>
<td>9</td>
<td>113%</td>
<td>67%</td>
<td>900%</td>
<td></td>
</tr>
</tbody>
</table>
Explain Findings (not discussed here)

**CLAIM**

Write a sentence that states what you found out about the scientific question you just investigated. Provide enough detail so that a friend who did not do the experiment could learn from your description.

**EVIDENCE**

Provide and describe scientific evidence from your data table that supports (or refutes) your claim. Remember to provide enough detail so that a friend who did not do the experiment could learn from your description.

**REASONING**

Explain why your evidence (what you wrote in Box 2) supports your claim (what you wrote in Box 1). Remember to provide enough detail so that a friend who did not do the experiment could learn from your description.
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.
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 based on AI-based assessments, important for real time instruction of NGSS practices.
Components for Teachers: Alerts

- **Recent Alerts**
  - **John Marcone**
    - Hypothesizing
    - now
  - **Victoria Fowler**
    - Hypothesizing
    - 1 min
  - **Daniel Waters**
    - Collecting Data
    - 3 min
  - **Sebastian Bloom**
    - Collecting Data
    - 5 min
  - **Miles Dearborn**
    - Collecting Data
    - 5 min
  - **Homer Wells**
    - Collecting Data, Hypothesizing
    - 5 min
  - **Sarah Bree**
    - Collecting Data
    - 6 min
  - **Felicia Page**
    - Hypothesizing
    - 8 min
  - **Erasmus Trey**
    - Hypothesizing
    - 10 min
  - **Simon Cole**
    - Hypothesizing
    - 12 min

Alerts are generated in real-time, as your students work in Inq-ITS. If an alert is resolved it will disappear from this list after 5 minutes.

**Alert - Hypothesizing | now**

John Marcone is struggling to understand what an independent variable is.

- **Mark as Resolved**

**About this lab**

<table>
<thead>
<tr>
<th>Phase Change</th>
<th>amount of ice vs melting point</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent Variables</td>
<td>Dependent Variables</td>
</tr>
<tr>
<td>Container Size</td>
<td>Melting Time</td>
</tr>
<tr>
<td>Amount of Ice</td>
<td>Melting Point</td>
</tr>
<tr>
<td>Amount of Heat</td>
<td>Boiling Time</td>
</tr>
<tr>
<td>Boiling Point</td>
<td></td>
</tr>
</tbody>
</table>

**Today's Performance**

<table>
<thead>
<tr>
<th>Hypothesizing</th>
<th>Collecting Data</th>
<th>Analyzing</th>
</tr>
</thead>
<tbody>
<tr>
<td>12%</td>
<td>86%</td>
<td>56%</td>
</tr>
</tbody>
</table>

**Today's Alerts**

- **Phase Change | amount of ice vs melting point**
  - 13:24 **Hypothesizing: Identifying IV**
    - Struggling to understand what an independent variable is
  - 13:12 **Hypothesizing: Identifying DV**
    - Struggling to understand what a dependent variable is

- **Phase Change | container size vs boiling time**
  - 13:24 **Hypothesizing: Identifying IV**
    - Struggling to understand what an independent variable is
  - 13:12 **Hypothesizing: Identifying DV**
    - Struggling to understand what a dependent variable is
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 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.

Slow Progress Alerts
If a student spend too much time on one phase, the teacher is alerted. This limit is set in settings.
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.
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.
- These were mined from our previous study.
TIPS Development

• **Teacher Inquiry Practice Supports** – 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
Press for TIPS (Teacher Inquiry Practice Supports)

John Marcone is struggling to understand what an independent variable is.

TIPS

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%
Here are some suggestions for guiding questions/comments you can use:

1) What is your independent variable?
2) Remember, an independent variable is the thing that you want to change.
3) Look at the goal and think about which variable you are going to change in your investigation. That is your independent variable.
4) For this investigation, your independent variable should be the amount of ice because this is what you are going to change.

“What’s your independent variable?”
Here are some suggestions for guiding questions/comments you can use:

1) What is your independent variable?
2) Remember, an independent variable is the thing that you want to change.
3) Look at the goal and think about which variable you are going to change in your investigation. That is your independent variable.
4) For this investigation, your independent variable should be the amount of ice because this is what you are going to change.

“Remember, an independent variable is the thing that you want to change.”
Here are some suggestions for guiding questions/comments you can use:

1. What is your independent variable?
2. Remember, an independent variable is the thing that you want to change.
3. Look at the goal and think about which variable you are going to change in your investigation. That is your independent variable.
4. For this investigation, your independent variable should be the amount of ice because this is what you are going to change.

“For the independent variable, select which variable you are going to change in your investigation.”
Here are some suggestions for guiding questions/comments you can use:

1) What is your independent variable?
2) Remember, an independent variable is the thing that you want to change.
3) Look at the goal and think about which variable you are going to change in your investigation. That is your independent variable.
4) For this investigation, your independent variable should be the amount of ice because this is what you are going to change.
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 to explain to me why this permits them to know how the change in the independent variable lead to changes in the dependent variable.”
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)

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

• On-going analyses will look at log data to determine how students' performances changed as a result of teachers using TIPS in their instruction (Adair et al., ICLS, 2022)
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.
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
• Thank you!
• Janice.gobert@gse.rutgers.edu
• Janice@apprendis.com
• Inqits.com