RUTGERS
UNIVERSITY | NEW BRUNSWICK

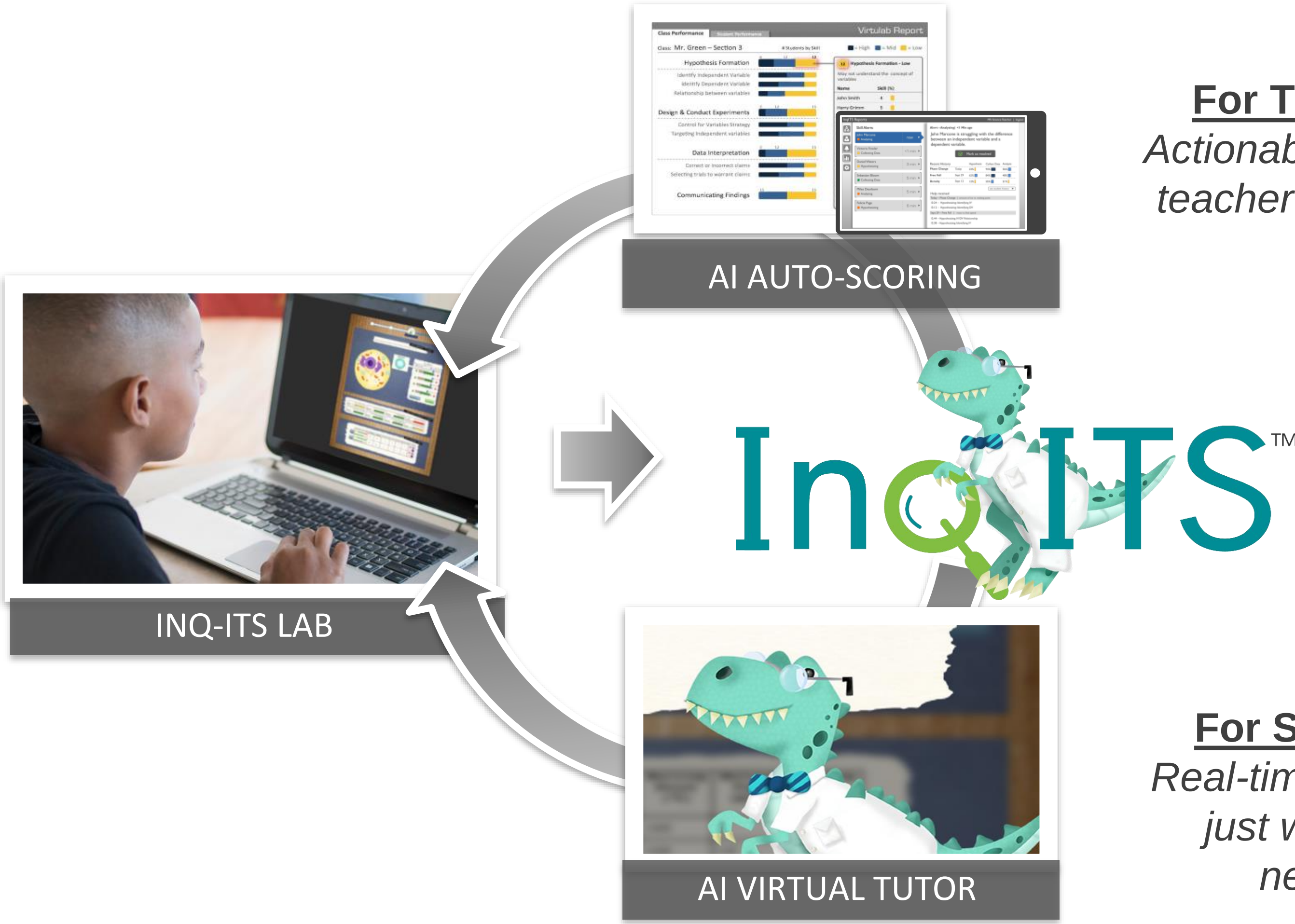
InQITS
Real Time Science Assessment & Tutoring

Automated Analyses of Students' Difficulties with Explanations in Science Inquiry

Jessica Owens[1] Amy Adair[1] Ellie Segan[1] Janice Gobert[1,2]

[1]Rutgers University
[2]Apprendis

Inquiry Intelligent Tutoring System (Inq-ITS)



For Teachers:
Actionable alerts via teacher dashboard

For Students:
Real-time feedback just when it is needed

Gobert, Sao Pedro, Baker, & Betts,
US Patents 9373082, 9564057, 10186168;
See inqits.com/research for papers

Next Generation Science Standards (NGSS)

Asking Questions	Developing & Using Models	Planning & Carrying Out Investigations	Analyzing & Interpreting Data
Using Math & Computational Thinking	Constructing Explanations	Engaging in Argument from Evidence	Communicating Findings

(NGSS, 2013; NRC, 2012)

Next Generation Science Standards (NGSS)

- Asking Questions
- Developing & Using Models
- Planning & Carrying Out Investigations
- Analyzing & Interpreting Data
- Using Math & Computational Thinking
- Constructing Explanations
- Engaging in Argument from Evidence
- Communicating Findings

(NGSS, 2013; NRC, 2012)

Claim - Evidence - Reasoning Framework
(McNeill & Krajcik, 2011)

Claim, Evidence, and Reasoning Rubric

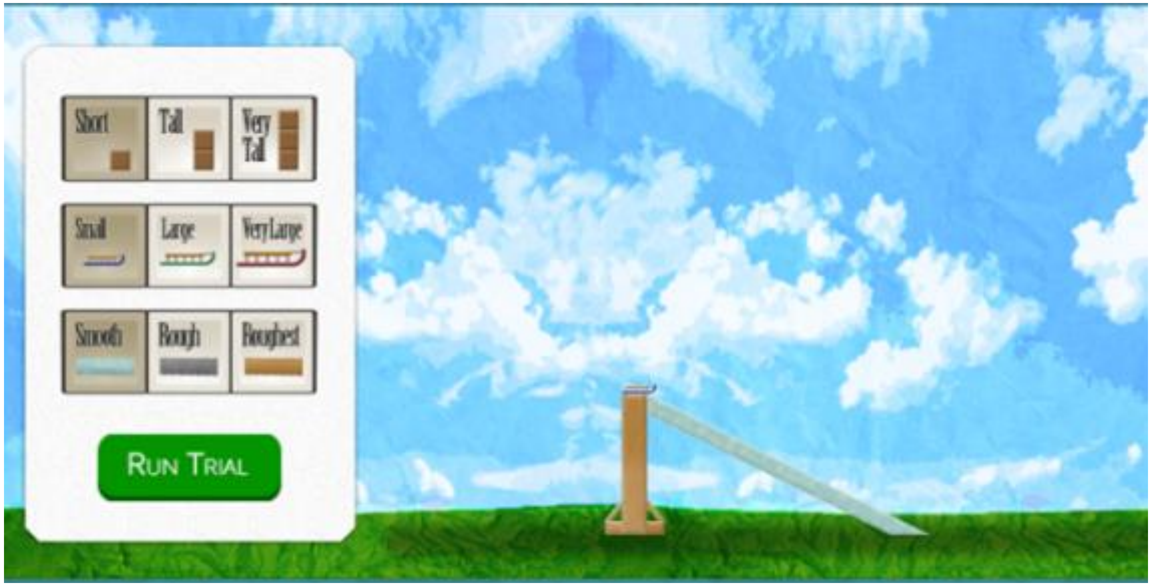
Claim	Evidence	Reasoning
A statement or conclusion that answers the original question (hypothesis).	*Appropriate and sufficient scientific data to support the claim.*	*A scientific principle that connects the evidence to the claim. It shows why the data counts as evidence by using appropriate and sufficient scientific principles.*
Example: *In my hypothesis I stated that if I changed the amount of liquid so that it increases, the density of the liquid would stay the same. My data supported my hypothesis.* *IV: The amount of liquid* *DV: The density of the liquid* *IV(R): so that it increases* *DV(R): would stay the same*	***Example:*** *My data clearly shows that as the liquid increased from 250 ml to 500 ml then to 1000 ml, the liquid's density stayed the same at 1.*	***Example:*** *The reason that my evidence supports my claim is because the density of a liquid does not change just by adding more liquid. Water always has a density of 1 whether it is in a little cup or in a giant lake. You can't change its density unless you change the liquid itself, like when fresh water becomes salt water in estuaries.*

(inqits.com)

Research Questions

1. Which sub-components of the Claim-Evidence-Reasoning process did students struggle with the most?

2. Were difficulties consistent across the two driving questions?



(inqits.com)

Study Overview

Participants

- 76 eighth-grade students from four science classes taught by one teacher in the Northeastern United States

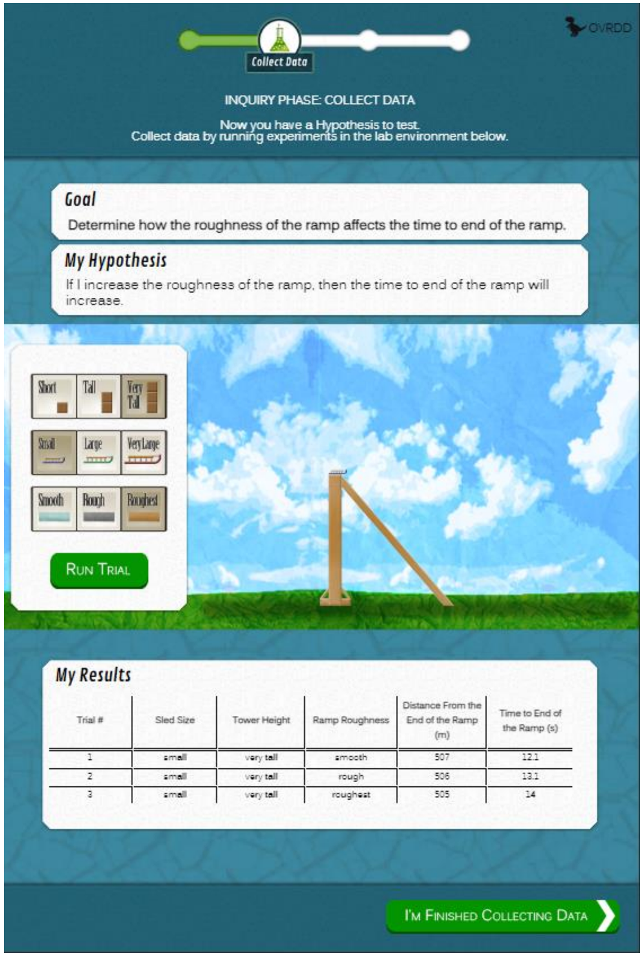
Materials

- Inq-ITS online intelligent tutoring system (Gobert et al., 2023)

- One Forces & Motion Virtual Lab was used as a formative assessment in which students answered two different driving questions:

  1. How does **ramp roughness** affect **time** to the end of the ramp?

  2. How does **sled size** affect **distance** traveled from the end of the ramp?



(inqits.com)

Scored with Automated NLP Algorithms

- Scores for each C-E-R component are automatically calculated as a sum of their sub-components

C-E-R	Sub-Component	Description	Possible Point Values
Claim	Claim IV	Did the student state the target independent variable (IV)	No Credit: 0, Max Credit: 1
	Claim IVR	Did the student say how they changed the independent variable (i.e., the independent variable relationship; IVR)?	No Credit: 0; Partial Credit: 0.5, 0.8; Max Credit: 1
	Claim DV	Did the student state the target dependent variable (DV)?	No Credit: 0, Max Credit: 1
	Claim DVR	Did the student say how the dependent variable changed in the experiment (i.e., the dependent variable relationship; DVR)?	No Credit: 0; Partial Credit: 0.5, 0.8; Max Credit: 2
Evidence	Sufficient	Did the student state data for at least two trials (i.e., a sufficient amount of data)?	No Credit: 0; Partial Credit: 0.5, 1; Max Credit: 2
	Appropriate IVR	Did the student state the appropriate data for the independent variable?	No Credit: 0; Partial Credit: 0.5, 0.8; Max Credit: 1
	Appropriate DVR	Did the student state the appropriate data for the dependent variable?	No Credit: 0; Partial Credit: 0.5, 0.8; Max Credit: 1
Reasoning	Connection	Did the student state how the claim relates to the evidence?	No Credit: 0; Partial Credit: 0.5, 0.8; Max Credit: 1
	DV/DVR	Did the student state the dependent variable and/or say how the dependent variable changed?	No Credit: 0, Partial Credit: 0.5. Max Credit: 1
	IV/IVR	Did the student state the independent variable and/or say how they changed the independent variable?	No Credit: 0, Partial Credit: 0.5, Max Credit: 1
	Theory	Did the student explain the scientific principle behind the phenomena?	No Credit: 0, Max Credit: 1

(Li et al., 2017)

RUTGERS Inq-ITS

Focus on Reasoning

- Using the C-E-R sums from the automated scoring, we analyzed students' performance across the two driving questions to see which sub-components students struggled with most frequently
- Three paired samples t-tests showed:
  - Claim and Evidence did *not* have a statistically significant change between the two trials
  - Reasoning *did* have a significant change between the two trials
    - The students' scores **decreased significantly** from Driving Question 1 to Driving Question 2
    - DQ1 $M = 3.49$, $SD = 1.37$ to DQ2 $M = 2.96$, $SD = 1.75$; $t(75) = 2.68$, $p = .009$
- Previous research shows that students often demonstrate difficulties with incorporating scientific theories and principles into their reasoning across science domains both in Inq-ITS (Adair et al., 2023) and elsewhere (McNeill et al., 2006)

C-E-R	Sub-Component	Description
Reasoning	Connection	State how the claim relates to the evidence
	DV/DVR	State the dependent variable and/or say how the dependent variable changed
	IV/IVR	State the independent variable and/or say how they changed the independent variable
	Theory	Explain the scientific principle behind the phenomena

Fine-Grained Hand Scoring

1. Students were grouped based on the overall **change in their score** for **Reasoning**
2. Their scores were assessed for **correctness** and **frequency** to determine their outcome

Reasoning Sub-Component	Overall Change in Reasoning	Correct Both Times	Partially Correct Both Times	Incorrect Both Times
Connection	Increased	9	2	1
	No Change	15	0	2
	Decreased	6	1	6
DV/DVR	Increased	8	0	2
	No Change	17	0	3
	Decreased	12	2	4
IV/IVR	Increased	7	3	1
	No Change	17	0	3
	Decreased	8	0	6
Theory	Increased	1	0	9
	No Change	0	0	20
	Decreased	2	0	30

Fine-Grained Hand Scoring

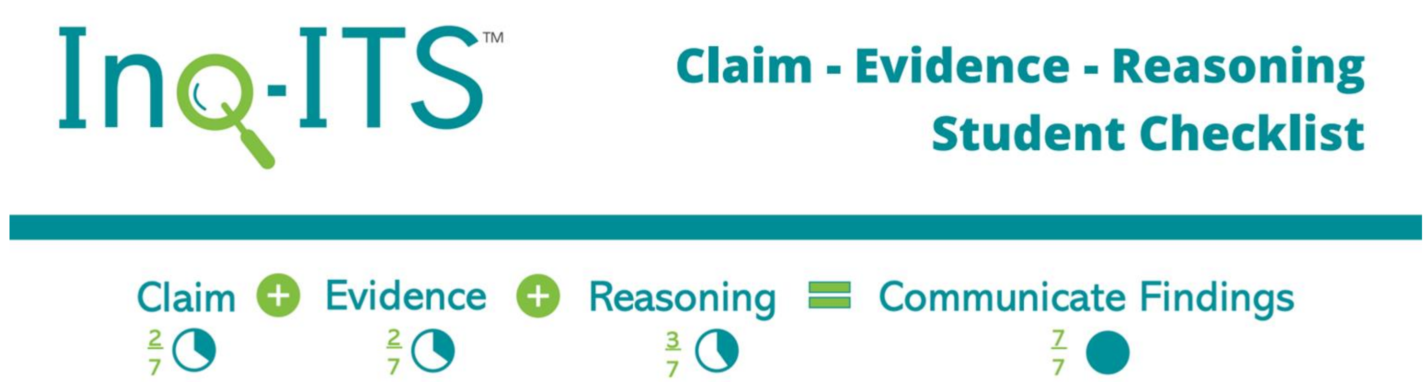
1. Students were grouped based on the overall **change in their score** for **Reasoning**
2. Their scores were assessed for **correctness** and **frequency** to determine their outcome
3. Their scores were assessed based on whether they **increased** or **decreased**

Reasoning Sub-Component	Overall Change in Reasoning	Correct Both Times	Partially Correct Both Times	**Increased from DQ1 to DQ2**	**Decreased from DQ1 to DQ2**	Incorrect Both Times
Connection	Increased	9	2	**4**	**2**	1
	No Change	15	0	**2**	**2**	2
	Decreased	6	1	**7**	**17**	6
DV/DVR	Increased	8	0	**7**	**1**	2
	No Change	17	[illegible]	[illegible]	[illegible]	3
	Decreased	12	[illegible]	[illegible]	[illegible]	4
IV/IVR	Increased	7	[illegible]	[illegible]	[illegible]	1
	No Change	17	[illegible]	[illegible]	[illegible]	3
	Decreased	8	[illegible]	[illegible]	[illegible]	6
Theory	Increased	1	[illegible]	[illegible]	[illegible]	9
	No Change	0	[illegible]	[illegible]	[illegible]	20
	Decreased	2	[illegible]	[illegible]	[illegible]	30

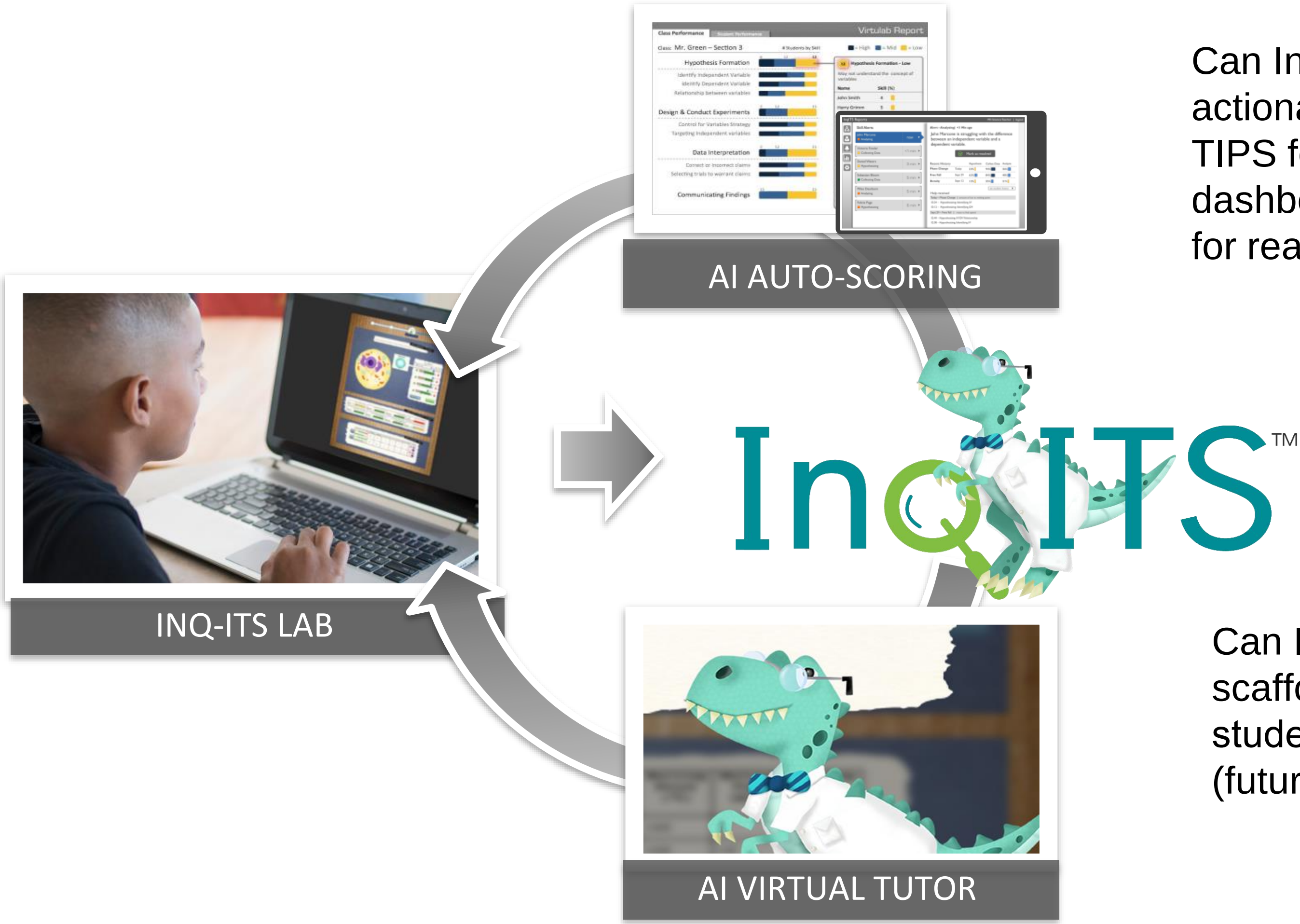
77.6% incorrect both times

Discussion

- C-E-R and their respective sub-components are difficult for students, as evidenced by our fine-grained NLP scoring
- The *Theory* subcomponent is particularly difficult because students must explain the scientific principle behind the phenomena
- Looking at the two driving questions:
  1. How does **ramp roughness** affect **time** to the end of the ramp?
     - Students likely have prior knowledge of *roughness* making the task a bit easier
  2. How does **sled size** affect **distance** traveled from the end of the ramp?
     - Requires more content knowledge, which may have been why the scores decreased



Implications for Future Work



Can Inform detailed, actionable alerts and TIPS for teachers on our dashboard, Inq-Blotter, for real time instruction

Can Inform our fine-grained scaffolds so Rex can target students' specific difficulties (future work)

Thank you!

Jessica A. Owens Amy Adair Ellie Segan
Dr. Janice Gobert

jo503@rutgers.edu amy.adair@gse.rutgers.edu ers173@scarletmail.rutgers.edu
janice.gobert@gse.rutgers.edu

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A210432 to Rutgers University. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

References

Adair, A., Dickler, R., Gobert, J., Sao Pedro, M., Olsen, J., Owens, J.A., & Lott, C. (2023). Examining students' mathematical evidence in CER explanations during science inquiry contexts. [Manuscript in preparation].

Chinn, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. Review of Educational Research, 63(1), 1-49.

Comparison of automated scoring methods for scientific explanations. Presented at American Educational Research Association (AERA): Sig-Cognition and Assessment, Toronto, Canada. doi: 10.13140/RG.2.2.36534.96325

Dickler, R., Li, H., & Gobert, J. (2019). Examining the generalizability of an automated scoring method and identifying student difficulties with scientific explanations. Presented at American Educational Research Association (AERA): Learning and Instruction, Toronto, Canada. doi: 10.13140/RG.2.2.33179.52006

Dickler, R., Gobert, J., & Sao Pedro, M. (2021). Using innovative methods to explore the potential of an alerting dashboard for science inquiry. *Journal of Learning Analytics, 8*(2), 1-18.

Gobert, J. D., Sao Pedro, M.A., Betts, C.G. (2023). An AI-Based Teacher Dashboard to Support Students' Inquiry: Design Principles, Features, and Technological Specifications. In N. Lederman, D. Zeidler, & J. Lederman (Eds.), Handbook of Research on Science Education, (Vol. 3, pp. 1011-1044). Routledge.

Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. Cognitive science, 12(1), 1-48.

Kubsch, M., Czinczel, B., Lossjew, J., Wyrwich, T., Bednorz, D., Bernholt, S., Fiedler, D., Strauß, S., Cress, U., Drachsler, H., Newmann, K., & Rummel, N. (2022). Toward learning progression analytics—Developing learning environments for the automated analysis of learning using evidence centered design. In, Frontiers in Education (p. 605). Frontiers.

Li, H., Gobert, J., & Dickler, R. (2017). Dusting off the messy middle: Assessing students' inquiry skills through doing and writing. In E. André, R. Baker, X. Hu, M. Rodrigo, & B. du Boulay (Eds.), Artificial Intelligence in Education. AIED 2017. Lecture Notes in Computer Science (pp. 175-187). Cham: Springer.

Liu, O. L., Rios, J. A., Heilman, M., Gerard, L., & Linn, M. C. (2016). Validation of automated scoring of science assessments. Journal of Research in Science Teaching, 53(2), 215-233.

McNeill, K. L., Lizotte, D. J., Krajcik, J., & Marx, R. W. (2006). Supporting students' construction of scientific explanations by fading scaffolds in instructional materials. The Journal of the Learning Sciences, 15(2), 153-191.

Myford, C. M., & Wolfe, E. W. (2009). Monitoring rater performance over time: A framework for detecting differential accuracy and differential scale category use. Journal of Educational Measurement, 46(4), 371-389.

Next Generation Science Standards Lead States. (2013). Next Generation Science Standards: For states, by states. The National Academies Press.

Osborne, J. F., Henderson, J. B., MacPherson, A., Szu, E., Wild, A., & Yao, S. Y. (2016). The development and validation of a learning progression for argumentation in science. Journal of Research in Science Teaching, 53(6), 821-846.

Ruiz-Primo, M. A., & Furtak, E. M. (2006). Informal formative assessment and scientific inquiry: Exploring teachers' practices and student learning. Educational Assessment, 11(3-4), 237-263.

Sao Pedro, M., Gobert, J., & Dickler, R. (2019). Can an Alerting Teacher Dashboard Improve How Teachers Help Their Students Learn Science Inquiry Practices?. *Grantee Submission*.