Impact of a *Teacher Action Planner* that Captures Student Ideas on Teacher Customization Decisions

Libby Gerard, Korah Wiley, Allison Bradford, Jennifer King-Chen, Jonathan Lim-Breitbart and Marcia C. Linn
libbygerard@berkeley.edu, korah.wiley@berkeley.edu, allison_bradford@berkeley.edu,
jykchen@berkeley.edu, breity@berkeley.edu, mclinn@berkeley.edu
Graduate School of Education, University of California, Berkeley

Abstract: This design-based research takes advantage of advanced technologies to support teachers to rapidly respond to evidence about student ideas generated in their classrooms. Leveraging advances in natural language processing methods, the Teacher Action Planner (TAP) analyzes students' written explanations embedded in web-based inquiry projects to provide teachers with a report on student progress in developing the three-dimensional understanding called for by the Next Generation Science Standards. Based on the pattern in student scores, the TAP recommends research-based ways for teachers to customize instruction. This study examines how ten middle school teachers in 4 schools used the analysis of student ideas and suggestions for instructional customization presented in the TAP. This paper reports on how well their implemented customizations addressed student learning needs. It concludes with a discussion of the implications of the findings for redesign of the TAP.

Keywords: automated scoring, assessment, teaching, technology, science, knowledge integration

Major issue and potential significance

Teachers regularly customize curriculum to meet new requirements, incorporate new technologies, or serve new student populations. Teachers often need to make these customizations without evidence of student learning, time to interpret evidence, or pedagogical principles to inform their decisions (Brown, 2009; Kerr, Marsh et al., 2006; Remillard, 2005). *The Teacher Action Planner (TAP)* supports teachers to make timely, evidence-based curriculum customizations by analyzing automated scores of student's written explanations embedded in web-based inquiry units. The automated scores are generated using a natural language processing tool that assesses each written explanation for the multi-dimensional proficiency called for by the Next Generation Science Standards (NGSS). The TAP is intended to help teachers design instructional customizations that respond to their students' emerging ideas. To help teachers use the score analysis, the TAP *suggests* research-proven instructional strategies. In this design-based research, we report on how teachers, encountering the TAP for the first time, make use of both the analysis of student responses and the suggestions to customize their instruction. The goal is to collaborate with teachers to evaluate the impact of the TAP and identify ways to make it more effective.

Supporting teacher customization

Research demonstrates that when teachers adapt instruction, there can be powerful benefits for student learning (Remillard, 2005). However, teachers often make customization decisions without evidence of student thinking or a research-based pedagogical framework (Penuel & Gallagher, 2009; Kerr et al., 2006). Further, with the adoption of NGSS, many teachers report searching the internet for appropriate lessons or modifications to available materials without professional development. Thus, it is not surprising that customizations rarely improve outcomes and can even reduce opportunities for students to learn (Bismark, Arias, Davis & Palinscar, 2015).

Recent efforts to design educative curriculum materials to support customization have had limited success yet offer promising insights for the TAP design. Educative materials embed guidance designed to support teacher and student learning in the curriculum (Davis & Krajcik, 2005). A review found that teachers use educative materials in both effective and ineffective ways (Davis, Palinscar et al., 2017). Teachers sometimes maintained or augmented students' opportunities to learn (e.g., by giving students more time to work on scientific practices) and sometimes limited learning opportunities. For example, to support the design of an investigation, teachers limited learning by making many decisions for the students. In another example, teachers removed opportunities for students to make predictions due to time constraints. The TAP builds on these findings by supporting teachers to make effective use of embedded evidence of student learning and associating student progress indicators with tested customization suggestions.

This design-based research takes advantage of advanced technologies to support teachers to rapidly respond to their students' developing ideas. Using a technology-rich learning environment, the curricula in this study engages students in exploring scientific models, conducting virtual experiments, linking hands-on investigations to simulations, and explaining their thinking in essays. We leverage advances in natural language processing methods to analyze student written explanations and provide teachers with fine-grained information about strengths and weaknesses in student work. Based on the analysis of student responses, the TAP provides a visual representation of the student work and recommends research-based ways to customize instruction (Fig. 1). We examine how teachers respond to the analysis of student ideas, the suggestions for instructional customization, and how well the customizations teachers select address the learning needs of diverse students.

Theoretical framework and technologies

This study uses the *knowledge integration (KI) framework* to guide the design of the web-based inquiry units, assessments, scoring rubrics and supports for teacher customizations (Linn & Eylon, 2011). Knowledge integration (KI) involves a process of building on and strengthening science understanding by incorporating new ideas and sorting out alternative perspectives using evidence. The KI framework describes four processes central to developing robust understanding: supporting students to *articulate* and reflect on their initial ideas, *discover* relevant evidence that challenges or extends their view, *distinguish* among their initial ideas and the evidence they discover, and *revise* their perspective.

The automated scoring models that inform the TAP are based on human scoring of student responses using a KI rubric. The KI rubric rewards students for linking evidence to claims and for adding multiple evidence-claim links to their explanations. The recommended teacher customizations in the TAP, generated based on the distribution pattern of student scores, are designed to engage students in KI processes that will increase the coherence of their explanations. The suggested customizations support students to recognize the ideas they hold, discover the missing ideas identified by the automated scores, explore relevant evidence, distinguish which of the new ideas fill a gap in their initial view or clarify a connection, and revise their explanation to incorporate their new ideas.

Further, this research draws on KI to position teachers' ideas about curriculum and their students' ideas as a valuable starting point for planning customization. We frame teacher learning through customization as an iterative process of *eliciting their views* on the curriculum and student learning, *discovering new ideas* from evidence of students' progress on NGSS aligned embedded assessments and discussion with colleagues; *distinguishing among ideas* about curriculum design features and student ideas and, finally, *reflecting and integrating* ideas to plan customizations to instruction.

Teacher Action Report (TAP)

We developed the TAP report using a multi-step process. First, researchers designed embedded assessments for each unit that prompted students to write a coherent explanation by integrating their ideas related to multiple dimensions of the targeted NGSS performance expectation (PE). Units, embedded assessments, and NGSS scoring dimensions are shown in Table 1. Second, researchers hand-coded over 1,000 student responses to each assessment prompt, using rubrics that assess the degree to which each response integrates multidimensional knowledge into a coherent explanation (Riordan et al., 2020). These human-scored data were used to develop a natural language processing scoring model (using craterML). Third, once the model demonstrated sufficient agreement with human scorers, it was embedded into the inquiry unit to generate the analytics part of the TAP. Teachers used the inquiry unit as part of their regular curriculum. Fourth, the TAP report was automatically generated and made available in the teacher interface of the online learning platform as students submitted their explanations to the embedded assessment (typically on day 3 or 4 of each 10-day unit).

A sample TAP report from the Thermodynamics unit is shown in Fig 2 (King-Chen, 2020). As shown in Fig 2, the TAP provides teachers with graphs summarizing the automated scores in each NGSS dimension and recommends instructional customizations to help students improve their understanding.

Table. 1. Unit, NLP scored embedded assessment, and automated scoring dimensions

Unit *NGSS PE*	**Embedded Assessment Prompt**	**Automated Scoring Dimensions [DCI, CCC, or SEP]* + KI**
Thermo-dynamics Challenge *MS-PS3-3*	Explain WHY the experiments you [plan to test] are the most important ones for giving you evidence to write your report. Be sure to use your knowledge of insulators, conductors and heat	DCI:Thermal energy transfer SEP: Planning an investigation KI: Explanation accuracy and coherence

	energy transfer to discuss the tests you chose as well as the ones you didn't choose.	
Solar Ovens *MS-PS3-3*	Explain why David's claim is correct or incorrect using the evidence you collected from the model. Be sure to discuss how the movement of energy causes one solar oven to heat up faster than the other.	CCC: Energy and matter: Flows, cycles, and conservation SEP: Analyzing and interpreting data KI: Explanation accuracy and coherence
Photo & Cellular Respiration *MS-LS1-6*	Write an energy story below to explain your ideas about how animals get and use energy from the sun to survive. Be sure to explain how energy and matter move AND how energy and matter change.	DCI: details of photosynthesis CCC: flow of energy and cycling of matter KI: Explanation accuracy and coherence

*DCI: disciplinary core idea; CCC: crosscutting concept; SEP: science and engineering practice

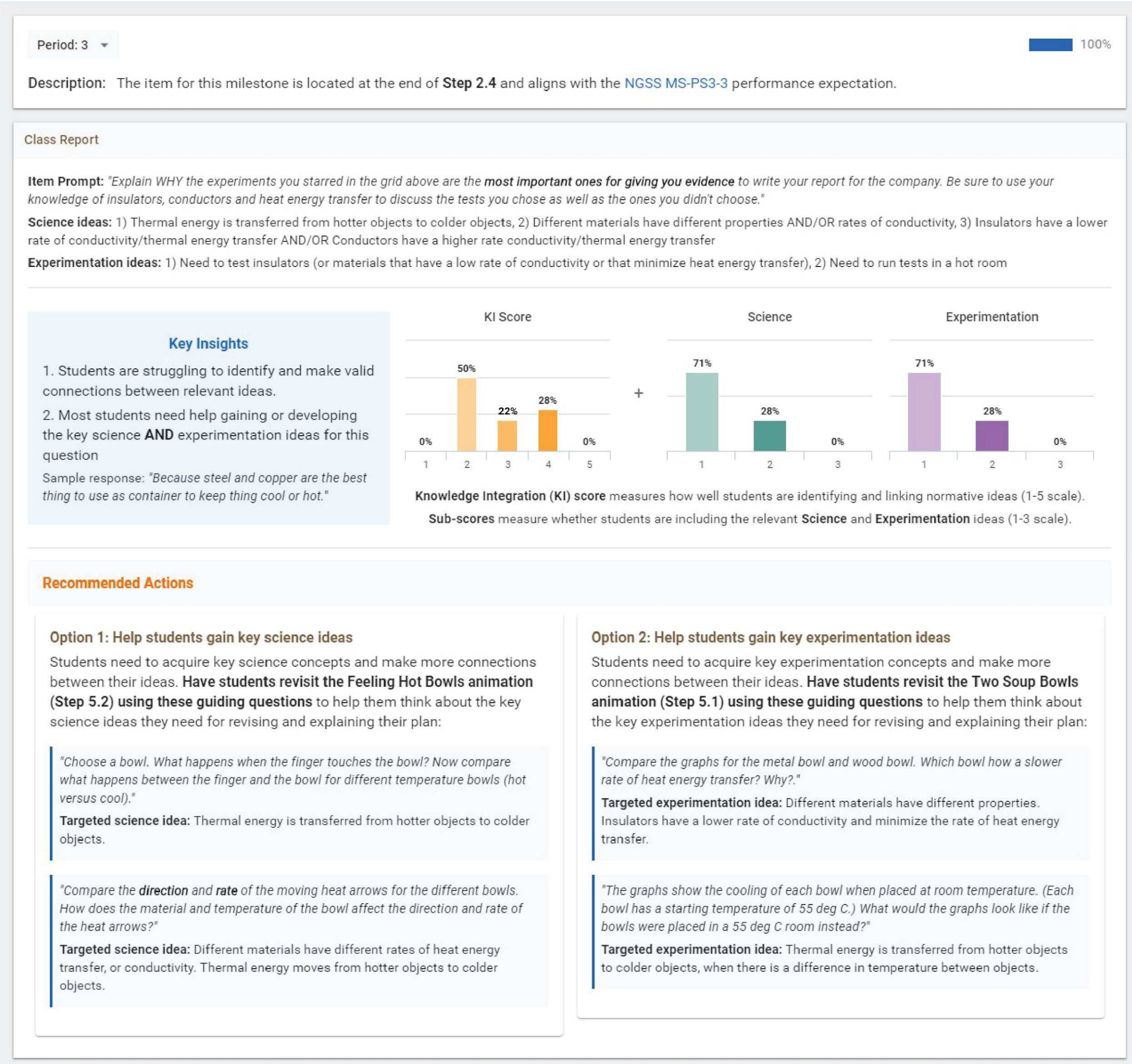

Figure 1. A sample TAP report generated for the Thermodynamics Challenge unit. The top section indicates the selected class period and specific NGSS PE targeted by the embedded assessment item. Next, the Class Report shows the assessment prompt and targeted DCI and SEP ideas. Key Insights (highlighted by the blue box) summarize for the teacher the main takeaways from the pattern of student scores for the three assessed

dimensions: KI, science DCI, and experimentation SEP. Lastly, Recommended Actions (which are adaptive, based on the specific pattern of student scores) provide the teacher with suggested instructional options for addressing their students' ideas. In this example, the recommended actions focus on helping students to discover new ideas by revisiting key models and considering guiding questions.

Methodology

Participants

Ten sixth and seventh grade teachers taught an online inquiry science unit that generated a TAP to help them guide customization of instruction during the 2018-2019 school year. Each unit contained one automatically scored explanation that provided the data for the TAP analytics. The teacher and the researcher met twice: before the unit to explore the unit and features of the TAP report; after the TAP was generated but prior to customization to review their class-specific TAP report.

In addition, most teachers met with researchers after customization to reflect on the experience. In the meeting after the TAP was generated, the researchers encouraged the teachers to respond to the TAP in any way they preferred, whether with inaction, by implementing or modifying a suggested customization, or by using a customization activity of their own.

Table 1. Participants

Unit	Teacher	Grade	Years experience teaching	Years teaching with On-line Inquiry	School Demographics
Thermo	A	6th	10	1	2% ELs, 2% F/R lunch
Thermo	B	6th	10	1	
Thermo	C	6th	8	5+	3% ELs, 12% F/R lunch
Photo	D	7th	20	18	
Photo	E	7th	< 1 (Pre-certification)		32% ELs, 72% F/R lunch
Photo	F	7th	25	0	
Solar Ovens		Newly Arrived to US Science			
Solar Ovens	G	6th	< 1 (Intern Teacher)	0	
Solar Ovens	H	6th	< 1 (Intern Teacher)	0	
Solar Ovens	I	STEAM Elective	10	0	8% ELs, 56% F/R lunch

Data sources and analysis

Data sources included recordings of two meetings with each teacher, classroom observations, and analysis of log data for unit activities. The meetings between the teachers and the researchers were semi-structured, using the same protocol across all 10 teachers, and ranged in duration from 30-60 minutes. In the first meeting, teachers commented as they reviewed the inquiry unit they chose to implement. In the second meeting, they reviewed the TAP generated based on their students' scores. Detailed classroom observations were collected by the lead designer and researcher associated with each unit. Observations were collected prior to use of the TAP (including the days students completed the embedded assessment), during the implementation of the customization, and after the implementation. The classroom observations, transcribed teacher interviews, and logs of interactions were used to capture how the teacher used information from the TAP to design their customization, how the customization supported students in KI processes, and what challenges arose. In terms of support for KI, we examined the degree to which the customization supported students to use the KI processes to: express the ideas they have, discover missing ideas by interacting with evidence, distinguish which of the new ideas they examined connect with their initial view, and refine their explanation to incorporate these new ideas. The lead researcher for each unit analyzed, categorized, and summarized their interview and observation data. Each presented their analysis to the research team, and categories were refined until reaching agreement. Each researcher then used these categories to analyze the data and generated reports on the specific unit (CITES; Author, 2020). This paper synthesizes these separate investigations.

Students' initial and revised explanations (before/after customization), and student responses to unit pretest/posttest items were logged and human-scored using 5-point KI rubrics. This paper focuses on the teacher-level data.

Major findings

Overall, most teachers used the TAP to design an instructional customization to respond to students' emerging ideas (Fig 2; 9/10 teachers). Most customizations focused on three KI processes: eliciting student's initial ideas, supporting students to add key ideas that were identified by the TAP as missing from their explanations (8/10 teachers); and prompting them to revise. The most commonly neglected KI process was distinguishing ideas. Adding ideas involved a general reteaching of the ideas identified as missing by the TAP. Some teachers presented these missing ideas to the whole-class, while others prompted their students to locate the missing ideas in the unit on their own and record them on a worksheet.

Four teachers aligned their customization with all four KI processes (Teachers A, C, F and I); two teachers aligned with all except distinguishing ideas; the remaining four omitted distinguishing ideas as well as other processes. The four teachers who designed customizations that aligned with *all* the KI processes, combined information from the TAP with their own activities. Teachers supported students to make connections between their initial set of ideas and the ideas in the unit that students might benefit from re-examining. For example, teachers guided students to articulate their ideas in worksheets or explanations; they encouraged students to analyze the dynamic model to discover key missing ideas; they supported students, either in pairs or a whole class discussion, to distinguish how these new ideas could be connected to explain the target topic; and they encouraged students to refine their initial explanation to incorporate the insights they gained from distinguishing among their ideas. These teachers framed students' revision of their explanation as a key part of doing science.

Teacher impressions. Most of the teachers reported in the interviews that they were surprised by their students' displayed level of understanding in the TAP. Their surprise varied by class period. Some were surprised that a class period they considered to be high performers displayed weaker understanding than they expected, while class periods they assumed were weaker demonstrated a more robust understanding than predicted. Based on the teachers' review of the TAP analysis, most expressed an interest in wanting to look more closely at each student's work, particularly to identify the students who were struggling the most.

Table 2. TAP-based Intervention Support for Knowledge Integration

Tchr/ Unit Years	Structure	Elicit	Discover	Distinguish	Connect
A/Thermo 10 years teaching	*Pair worksheets targeting TAP-identified NGSS dimensions + whole-class discussion*	Worksheet asking students to express their initial ideas	Partners revisit the model and extract information guided by worksheet; share with partner	In a student-led class discussion identify how evidence from the model explains key ideas	Prompted pairs to revise their explanations to incorporate ideas from discussion
C/Thermo 5+ years teaching	*Whole-class opener + pair revisit model*	Asking students to explain model elements	Drawing attention to model elements in whole-class discussion	Partners compare previous response to model insights, revisiting the model in the unit	Prompted students to revise, emphasizing refinement as key scientific practice
F/Photo[1] 25 years	*Storyboard targeting missing*	Connecting terms to students' ideas	Exploring possible ideas to include in	Determining which ideas to include in	Connect your ideas and storyboard

teaching	*student ideas identified by TAP*		their story	your revised story	ideas to make a coherent story
F/Ovens 25 years teaching	*Opener + energy diagram + pair revisit model*	Asking students to explain model elements	Drawing attention to model elements to include in energy diagram	Determining which ideas to include in energy diagram	Prompted students to revisit model and revise to incorporate ideas from diagram
I/Ovens 10 years teaching	*Opener + pair revisit model*	Asking students to explain model elements	Partners revisit the model to clarify key terms suggested by the teacher	Partners compare previous response to model insights	Prompted for multiple revisions based on automated guidance
D/Photo 20 years teaching	*Read TAP summary aloud and told to revise*	Sharing your response with peers	Listening to others' responses	--	Add score - improving ideas to explanation to improve response
B/Thermo 10 years teaching	*Whole-class opener*	Asking students to explain model elements	Drawing attention to model elements in whole class discussion	--	--
E/Photo <1 year teaching	*Worksheet emphasizing recall of key terms*	--	Looking for correct answer on specified unit step	--	Revise explanation to include ideas from the worksheet
H/Ovens <1 year teaching	*Circled the room with clipboard of ideas listed by TAP*	--	Asked questions about potentially missing ideas in their response	--	Prompted to revise
G/Ovens Yrs. experience =<1	*Decided not to intervene*	--	--	--	--
TOTAL per teacher		7	8	4	7

Teachers' decisions about how to take advantage of the information in the TAP reflected their pedagogical perspective. Teachers adapted the customization suggestions in the TAP to fit into their existing pedagogical structures. Decisions also reflect level of prior teaching experience to some extent. More experienced teachers were more likely to create customizations that engaged students in all KI processes, whereas the three new teachers tended to focus on reteaching and student recall of key terms.

For example, four teachers taught the Solar Ovens unit with the embedded TAP. Teachers' experience varied, from beginning and working towards receiving their credential, to many years of experience. Teacher G, who was a new teacher and taught a typical science class, reviewed the TAP, planned to do a customization, but ultimately did not. Teacher H, who also was a new teacher and taught a typical science class, reviewed the TAP and decided to let students continue to progress through the unit. However, as Teacher G circled the room to monitor student progress, Teacher G examined individual student explanations to ensure that they included the ideas highlighted in the TAP. This teacher also used the online teacher tools to give written feedback on other embedded assessments to several students.

In contrast to Teachers G and H, Teachers F and I planned customizations (drawing on the suggestions from the TAP) that supported students to connect their initial explanation to key ideas they were missing. Both Teachers F and I have 10+ years experience teaching and taught specialized science classes. They each prompted students to examine evidence presented in a dynamic model illustrating the key ideas. Teacher I's customization included a class opener consisting of a guided review of the model. Specifically, Teacher I asked students probing questions, similar to those listed in the TAP recommended actions, to help them recognize the key scientific mechanism in the model then directed students to explore the model with their partner. Teacher G instructed students to revise their explanation to incorporate the new ideas that they gathered as well as terms the teacher listed on the board.

Design implications

Analysis of the ways that 10 teachers used the TAP have implications for redesign to improve decision making.

Make visible the KI structure in each recommended action. A goal of the TAP is to encourage teachers to design customizations that respond to students' initial ideas (as analyzed in the TAP), and help students revise their explanations in light of new ideas identified as missing by the TAP. Effective customizations could engage students in each of the KI processes. This suggests that a redesigned TAP should make visible how each recommended action aligns with the KI framework. In other words, each suggested customization could be broken into four parts to illustrate how the customization engages students in each of the KI processes. In this way, teachers can modify the suggested customization to fit with their teaching approach while maintaining a focus on creating a robust pedagogical structure. Making the research-based pedagogical structure underlying curriculum visible has proven effective in prior work focused on helping teachers customize curriculum units for their specific student needs (Penuel & Gallagher, 2009).

Enable teachers to reflect on the impact of their customization. In this study the TAP updated in real-time. In some cases, a researcher supported the teacher to compare their students' initial scores to their revised scores, after they implemented the customization. These teachers were often surprised by their students' overall improvement, and they noticed opportunities for further refinement and exploration of ideas, wondering what they might do differently the next time. The redesigned TAP should enable teachers to more easily reflect on the impact of their customizations, so that they can generate their own ideas about customization design, try them in their classroom, and reflect on the results. As shown in Figure 2, a possible re-design would show students' initial explanation scores, beside the scores for their revised explanation. This quickly accessible evidence may provide a springboard for teachers' continued learning from the customization process. By designing customizations that instantiate teachers' conjectures about instruction, testing them in their classrooms, and reflecting on the results teachers can encourage the formation of new insights and begin to establish a process of continuing to improve instruction in their classroom.

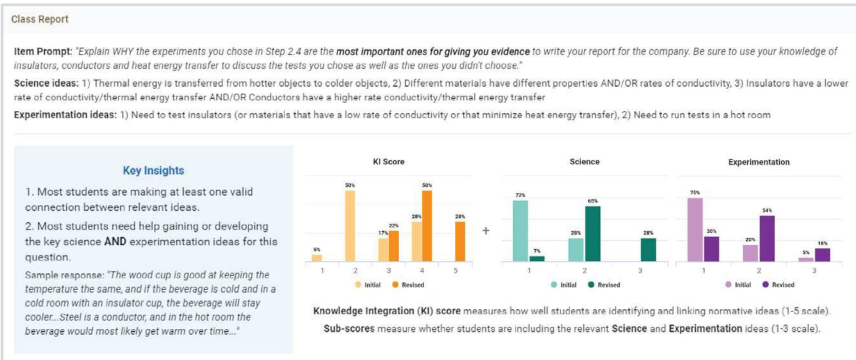

Figure 2. A prototype redesign of the TAP graphs, illustrating the change in student scores from their initial explanation to their revised explanation, after the teacher implemented a customization. Scores for students' initial explanation in the lighter shade, and the scores for their revised explanation are in the darker shade.

Support teachers to see individual student scores, so they can target assistance during class. When reviewing the TAP, many of the teachers expressed interest in knowing more about individual students' progress. This information can help the teacher target their support to those who need it most and enable the teacher to deliberately group students based on their scores to create collaborative learning opportunities. For the redesigned TAP, the team will explore ways to informatively display individual student scores, in addition to class-level data.

Conclusions

This research illustrates the wide range of customization decisions teachers make in response to evidence of their students' progress on the 3-dimensional integrated understanding called for by the NGSS. The findings indicate that nine of the ten teachers used the TAP to customize their instruction, demonstrating promise for promoting evidence-based instructional customizations. Refinements to the TAP are needed to help teachers design and refine instructional customizations that build on their students' ideas and guide them to progress in the KI process.

Further, this work points to promising research directions including: identifying a trajectory of teacher development in using evidence to customize instruction; distinguishing the value of using the TAP analytics relative to other forms of evidence (e.g. review of individual student work; checking-in with pairs during class) to support teachers in making curriculum customization within the constraints of classroom instruction; capturing teachers' effective customization decisions and generating more generalizable design principles.

References

Bismark, A. S., Arias, A. M., Davis, E. A., & Pallincsar, A. S. (2015). Examining student work for evidence of teacher uptake of educative curriculum materials. *Examining student work for evidence of teacher uptake of educative curriculum materials, 52*(6), 816-846. doi:10.1002/tea.21220

Brown, M. W. (2009). The teacher-tool relationship: Theorizing the design and use of curriculum materials. In Mathematics teachers at work: Connecting curriculum materials and classroom instruction. In J. T. Remillard, B. A. Herbel-Eisenmann, & G. M. Lloyd (Eds.), (pp. 17-36). New York: Routledge.

Davis, E. A., & Krajcik, J. (2005). Designing educative curriculum materials to promote teacher learning. *Educational Researcher, 34*(3), 3-14.

Davis, E., Palinscar, A. S., Smith, P. S., Arias, A. M., & Kademian, S. M. (2017). Educative curriculum materials: Uptake, impact, and implications for research. *Education Researcher, 46*(6), 293-304.

Kerr, K. A., Marsh, J. A., Ikemoto, G. S., Darilek, H., & Barney, H. (2006). Strategies to Promote Data Use for Instructional Improvement: Actions, Outcomes, and Lessons from Three Urban Districts. *American Journal of Education, 112*, 496-520. doi:10.1086/505057

King-Chen, J., (2020). Impact of autoscored student data reports on teacher instructional customizations for students' science learning. Paper to be presented at the National Association of Research in Science Teaching (NARST) Annual Meeting, Portland, OR.

Linn, M. C., & Eylon, B.-S. (2011). *Science Learning and Instruction: Taking Advantage of Technology to Promote Knowledge Integration*. New York: Routledge.

Next Generation Science Standards (NGSS, Accessed 2019). MS-ESS2-3 Earth's Systems. Retrieved from https://www.nextgenscience.org/pe/ms-ess2-2-earths-systems

Penuel, W. R., & Gallagher, L. P. (2009). Preparing teachers to design instruction for deep understanding in middle school earth science. *The Journal of the Learning Sciences*, *18*, 461-508

Remillard, J. T. (2005). Examining key concepts in research on teachers' use of mathematics curricula. *Review of Educational Research, 75*(2), 211-246.

Riordan, B. Wiley, K., King Chen, J., Bradford, A., Mulholland, M., Gerard, L., & Linn, M.C. (2020) Automated scoring of science explanations for multiple NGSS dimensions and knowledge integration. Paper accepted for *American Education Research Association Conference*. San Francisco, CA: AERA

Ruiz-Primo, M. A., & Furtak, E. M. (2007). Exploring teachers' informal formative assessment practices and students' understanding in the context of scientific inquiry. *Journal of Research in Science Teaching, 44*(1), 57-84.

Acknowledgements

Funding for this research was provided by the National Science Foundation awards STRIDES: Supporting Teachers in Responsive Instruction for Developing Expertise in Science (NSF project# 1813713).