Collaborative Research: Supporting Instructional Decision Making: The Potential of Automatically Scored Three-dimensional Assessment

PASTA — The Potential of An Automatically Scored Three-dimensional Assessment System

Xiaoming Zhai[1], Joe Krajak[2], Yue Yin[3], Christopher Harris[4], Xinyu He[1], Ehsan Latif[1], Peng He[2], Namsoo Shin[2], David McKinney[4]
[1]University of Georgia; [2]Michigan State University; [3]University of Illinois Chicago; [4]WestEd

Project Goals

This four-year Level II Teaching Strand project, targeting early-stage Design and Development, has three goals:

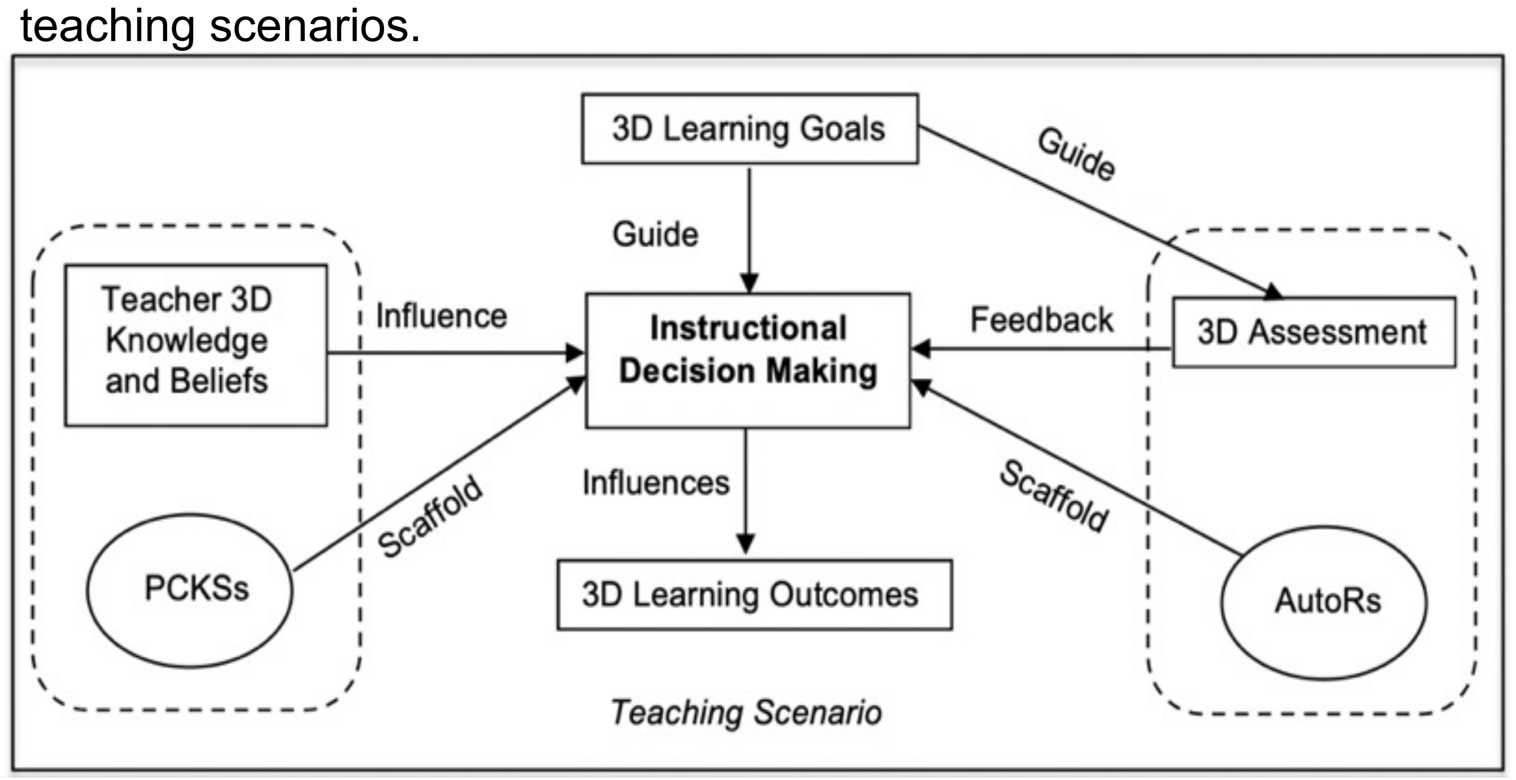
(a) develop automatically generated student reports (AutoRs) for three-dimensional (3D) science assessments to assist middle-school teachers to notice, attend to, and interpret information in ongoing classroom teaching;

(b) develop effective pedagogical content knowledge supports (PCKSs) to improve teachers' use of AutoRs to make effective decisions for instructional moves;

(c) examine the effectiveness of AutoRs and PCKSs to support teachers' decision making and student 3D learning.

Theoretical Framework

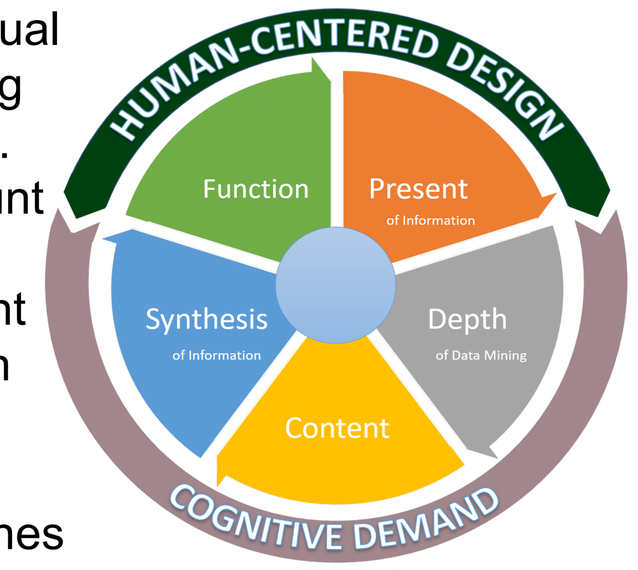
We develop a theoretical model accounting for scaffolding teachers' instructional decision-making with AutoRs and PCKSs. In this model, AutoRs and PCKSs scaffolds teachers to make instructional decisions that influence students' 3D learning outcomes. AutoRs, as part of the 3D assessment, helps teachers interpret students' responses to 3D assessments. PCKSs are expected to improve teachers' transferable capacity because they can eventually improve teachers' 3D knowledge and beliefs so that teachers can apply PCK and relevant skills to new teaching scenarios.



AutoR Design

Inspired by cognitive load theory, we propose a conceptual framework for teacher-centered AutoRs design, including cognitive demands and human-centered design support.

Cognitive demands. This dimension examines the amount and level of integration of the information in the report. Three characteristics are recognized to count the amount and the level of the information: the content presented in the report, the synthesis level of presented information, and the depth of data mining.

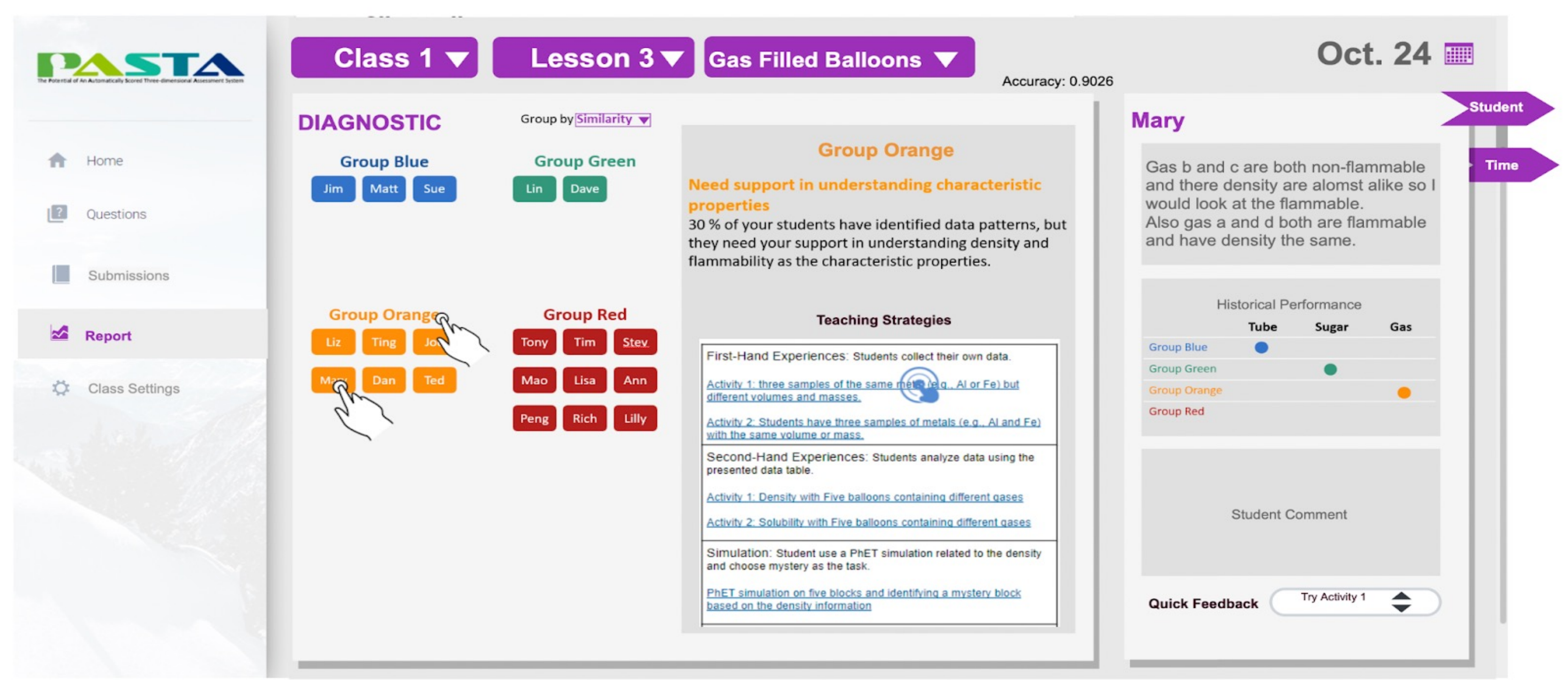


Human-centered design support. This dimension examines the design features that can assist teachers in interpreting and using the information. We identify two sub-dimensions: user functionality and information presentation.

To provide grouping information, we develop a grouping rubric based on students responses achievement on 3D.

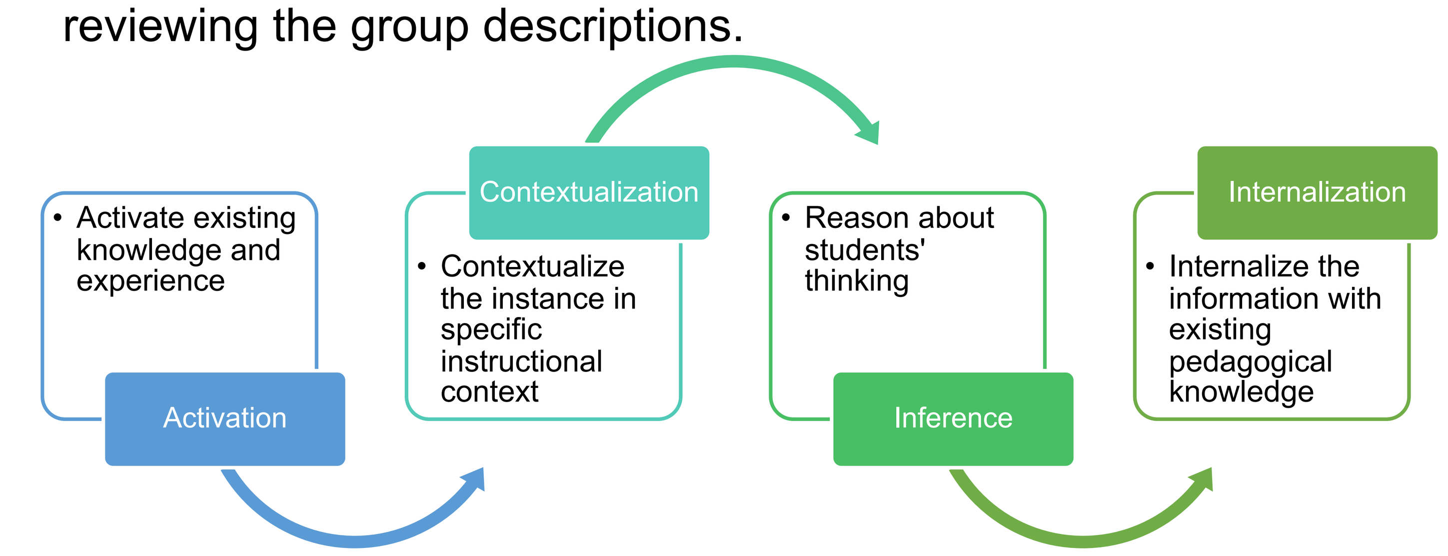
Group	Achieved	Not Yet Achieved
A	Students **have achieved in 3D**. • DCI: understanding the characteristic properties vs. non-characteristic properties of substances • SEP: analyzing and interpreting data • CCC: identifying patterns in data	Students have achieved the task and **need support to move forward**.
B	Students have **achieved in DCI**. • DCI: understanding the characteristic properties vs. non-characteristic properties of substances	Students **need support in SEP and CCC**. • SEP: analyzing and interpreting data; and • CCC: identifying patterns in data.
C	Students have **achieved in SEP and CCC**. • SEP: analyzing and interpreting data; and • CCC: identifying patterns in data	Students need **support in DCI**. • DCI: understanding the characteristic properties vs. non-characteristic properties of substances.
D	Students have **achieved in N/A**	Students need **support in 3D**. • DCI: understanding the characteristic properties vs. non-characteristic properties of substances; • SEP: analyzing and interpreting data; and • CCC: identifying patterns in data.

We plan to present three types of information on screen: a) grouping information, b) instructional strategies suggestions; and c) individual task performance. Then, we have teachers involved to provide feedback of the AutoR and revise the AutoR design in an iterative loop.



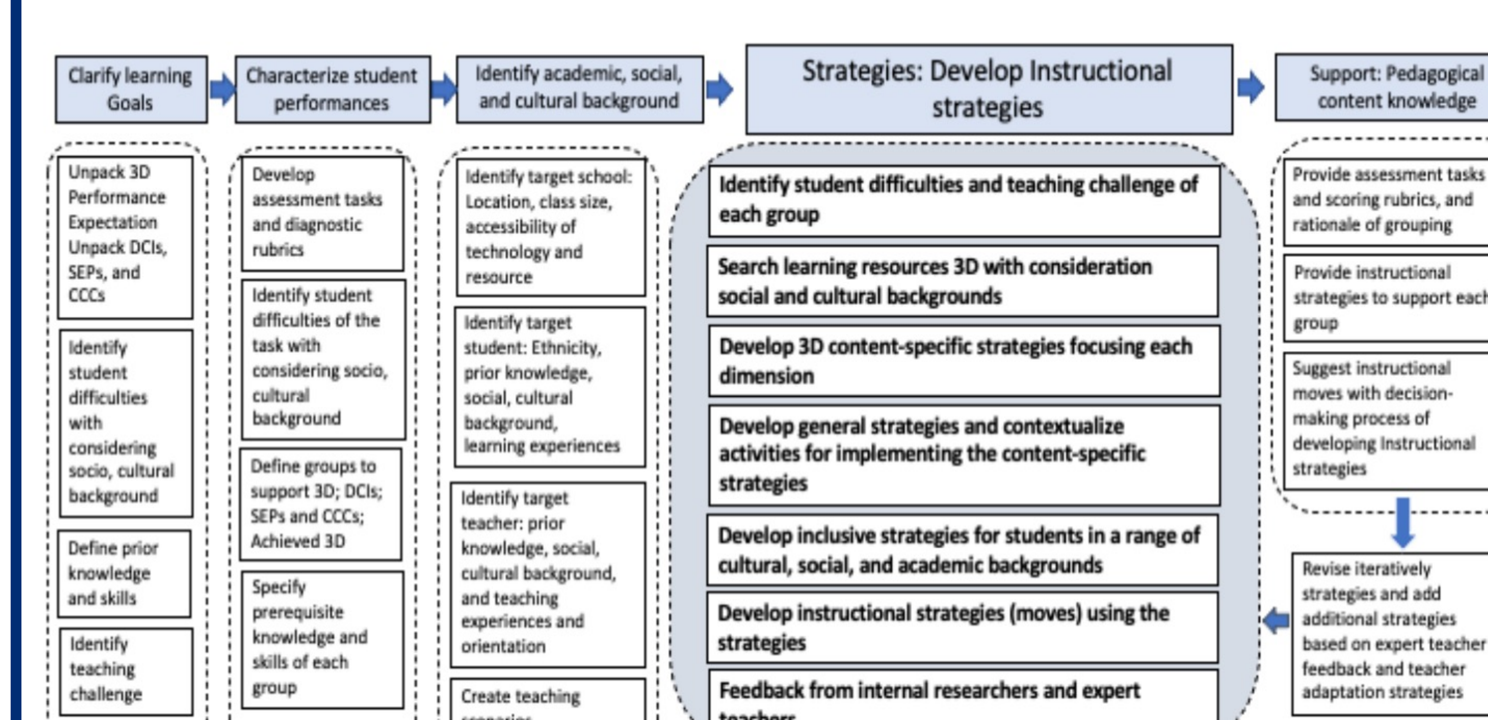
Teacher Interpretation of AutoR

- *Knowledge activation.* The original AutoRs activate teachers' curiosity about students' learning. Teachers connected the information received with their experience in class (e.g., a specific student).
- *Contextualization.* Teacher abstract the contextual information to understand the situation and process. For example, Teachers A and B paid attention to the scores of some students who were highlighted in the diagram of responding time and showed their concerns in figuring out the learning condition of the students.
- *Cross-validating information for reasoning.* In reading the individual scores table, teachers inferred the highlighted students' thinking who were identified before.
- *Prioritizing information for reasoning.* Teacher C paid attention mainly to the individual scores table and made decisions directly based on the scores. However, teacher A spent most of the time reviewing the group descriptions.



Instructional Strategies for Using AutoRs

A second challenge for teachers is how to effectively transform student performance on 3D assessments into effective instruction. Effective instructional moves require not only the ability to interpret assessment results but also actionable knowledge (Bennett, 2018). If teachers have limited PCK to transform their interpretation of student performance into meaningful 3D instructional activities, 3D assessments might still end up with a limited impact. Therefore, essential PCKSs such as pedagogical scaffolds might help teachers effectively use information from 3D assessments (Bybee, 2014). This project will develop PCKSs to help teachers make instructional decisions to effectively promote 3D learning.



Automatic scoring development

We select nine items about chemical reaction from NGSA and develop 3D-based analytic rubrics for each item. Each item has 5-10 dimensions. We use a pre-trained natural language processing model – BERT to do a binary text-classification task for automatic scoring. According to the analytic scores, we further divide student responses into four groups, which show the responses achievement on 3D.

Gas filled balloons

Alice did an experiment that caused four balloons to fill with gas, as shown in the figure to the right. Alice tested the flammability of each gas. She also measured the volume and mass of each gas to calculate the density. The tests and measures all occurred under the same conditions. The data is in the Table 1 below.

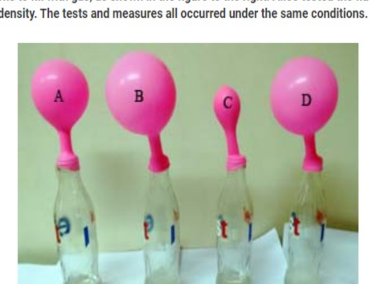


Table 1. Data of four gases in the balloons.

Sample	Flammability	Density	Volume
Gas A	Yes	0.089 g/L	180 cm^3
Gas B	No	1.422 g/L	270 cm^3
Gas C	No	1.981 g/L	35 cm^3
Gas D	Yes	0.089 g/L	269 cm^3

Which, if any, of the gases listed in the data table could be the same? Using information from the table, explain your answer.

Student response:
Gases A and D could be the same gas. Because they have the same density even though they don't have the same volume.

Dimensions	Elements	Score
DCI	Student clearly states that Gases A and D could be the same substance.	1
SEP+CCC	Student supports a claim by referring to a data pattern that the flammability of Gas A and Gase D is the same in the table. (comparisons of data in different columns).	0
SEP+CCC	Student supports a claim by referring to a data pattern that the density of Gas A and Gas D is the same in the table. (comparisons of data in different columns).	1
DCI	Student indicates flammability is one of the characteristic properties to identify substances.	0
DCI	Student indicate density is one of the characteristic properties to identify substances.	0

Question	% of students correct-human scoring	Testing Algorithm-Tensor Flow (Transformer- Bert Based)		
		Aspect	Accuracy	Kappa
c01Gas filled balloons (034.02-c01)	74.30%	1	95.98	0.885
	56.90%	2	96.77	0.934
	65.90%	3	95.54	0.902
	10.10%	4	95.54	0.783
	11.30%	5	96.76	0.837

This study was made possible by National Science Foundation grants Award ID: 2101104 (PI Zhai), 2100964 (PI Krajcik), 2101166 (PI Harris), 2101112 (PI Yin). The findings, conclusions, or opinions herein represent the views of the authors and do not necessarily represent the view of personnel affiliated with the National Science Foundation.