Uncovering the unknown unknowns of Peer Instruction questions

Rica Sirbaugh French[1,2] and Edward Prather[2,3]

[1]*Department of Physical Sciences, MiraCosta College, One Barnard Drive, Oceanside, CA, 92056*
[2]*Center for Astronomy Education, University of Arizona, 933 N. Cherry Avenue, Tucson, AZ, 85721*
[3]*Department of Astronomy, University of Arizona, 933 N. Cherry Avenue, Tucson, AZ, 85721*

"Wow! I wish I'd written that question..." Many instructors struggle to write Peer Instruction questions that can drive the intellectually engaging discussions necessary to fully develop learners' discipline fluency. Does your question bank contain the questions you need? How would you go about evaluating this and even if you did, how would you know what you're missing? We present a framework for uncovering the variety in the discipline representations, intellectual tasks, and difficulty levels employed in hundreds of multiple-choice questions produced by faculty in our workshops over the years. We then exploit this framework to generate new questions using underutilized representations and tasks. Through this work, we illustrate a process for creating *fluency-inspiring questions*. Learning environments that make use of fluency-inspiring questions afford learners more robust opportunities to unpack complex concepts, practice critical discernment, and develop discipline fluency.

I. INTRODUCTION

For 15 years members of the Center for Astronomy Education (CAE) have worked with thousands of college faculty and instructors during professional development workshops that focus on the use of Peer Instruction (PI, also called Think-Pair-Share, TPS, in the college-level introductory astronomy community) as one of many active-learning strategies [1, 2]. In these workshops faculty collaborate in small groups to craft their own multiple-choice questions for use with PI/TPS. We observe, to first order, reasonable diversity in these questions' levels of complexity, the ways the questions convey information, and the kinds of cognitive tasks a learner is required to do to answer any given question. A more rigorous examination, however, reveals the questions may lack the variety needed to help learners develop robust understandings of the key ideas of the discipline. Documenting how to characterize the existing variety and demonstrating how this information can be used to generate more effective questions should be of great use to instructors.

Here we describe efforts to unpack the variables used to characterize the PI/TPS questions and illustrate how this work can shed light on what is missing. Specifically, we describe a framework for (1) characterizing and cataloging the questions in terms of the representations used, intellectual tasks required, and complexity levels reached; (2) systematically identifying the variables missing from the questions on a particular topic; (3) generating new questions to fill in the gaps; (4) generating what we have coined *fluency-inspiring questions*: those making use of multiple representations and intellectual tasks that require the learner to develop one or more complex lines of reasoning and, if necessary, integrate them together to answer the question; and (5) generating new ways of thinking about how to move learners toward fluency by creating richer opportunities for the learners to practice critical discernment, unpacking their reasoning, and being more reflective.

II. BACKGROUND AND THEORY

A. A Database of Proven Peer Instruction Questions

CAE members and collaborators have developed hundreds of multiple-choice questions for use in various research, assessment, and instructional purposes (PI, in particular). As the first step in creating a database of classroom-tested PI/TPS questions freely available to instructors via the CAE website, we sorted approximately 500 of these questions based on primary astronomical topic. Initially, we intended to also communicate the specific conceptual and reasoning difficulties each question was designed to address. It was not clear, however, that instructors choosing to implement PI/TPS in their teaching would perceive this as the most helpful information when searching for good questions. Recognizing that the questions contain a wealth of information far beyond simply the topic addressed helped us see that not only was a more descriptive and informative characterization necessary, but also *understanding how to characterize this information* was equally important. We believe that instructors using PI/TPS questions in their classrooms and curriculum development as well as AER/PER researchers alike will find these results of great value.

B. Theoretical Foundations

This work is, in large part, inspired by the work of Cedric Linder and collaborators on *representations* (ways of conveying information), *affordances* (pieces of disciplinary information that a representation allows access to), *unpacking* (disassembling a package of information and making the various pieces and connections explicit), and *discernment* (coming to recognize and understand what to focus on and interpreting it or making meaning using the appropriate context). Indeed, Linder [3] and Airey and Linder [4] contain the basis for our initial rubric on representations and as our analyses continue, we are beginning to see evidence to support some of their theoretical postulates. One such example is the notion that to achieve discipline fluency, it is necessary (though not sufficient) for learners to become fluent with a collection of representations [3, 4]. Fredlund et al. even note that since different representations allow access to different types of information and often emphasize different aspects of a concept, it is pedagogically more powerful to use multiple representations than to rely on a single one to do the work of many [5].

This is also borne out in our work as we attempt to create more complex questions that make use of multiple representations and require learners to engage in multiple intellectual tasks, e.g. fluency-inspiring questions. As Linder points out, instructors who employ only a limited set of representations cannot move their students to discipline fluency because learners' abilities to unpack their reasoning and practice discernment are hampered by the lack of variety in the representations [3]. Students will naturally default to the representations most often used rather than those most suitable for the task(s) at hand [6]. Thus, it is up to the instructors to inject more variety and combine multiple representations with pedagogically appropriate intellectual tasks in ways that facilitate developing learners' discipline fluency. Fredlund et al. also suggest that the power of research-validated active-learning methods such as PI/TPS may lie within a theoretical framing in which the method itself, by its very nature, naturally facilitates the unpacking and disambiguation of the affordances of a set of representations [5].

III. DATA AND RESEARCH PROGRESSION

A. The Data

We began with 353 multiple-choice questions created by faculty during 45 multi-day professional development workshops. The majority (293) are from 41 CAE Teaching Excellence Workshops [1, 2] held from 2005-2015. These workshops target instructors of general education introductory Earth, astronomy, and space science courses with question topics spanning the Earth-Sun-Moon system, Renaissance astronomy, solar system, light and atoms, stars, exoplanets and life in the universe, and galaxies and cosmology. The remaining 60 questions are from four of the AAPT/APS/AAS Workshops for New Faculty in Physics and Astronomy [7] held from 2015-2017. These professional development experiences support physics and astronomy instructors who are in the first few years of their initial teaching appointments. Questions from these workshops cover the topics of work and kinetic energy, inelastic collisions, rotational motion, heat and temperature, Gauss's law for electric fields, Faraday's and Lenz's laws, simple harmonic motion, and the Bohr model of the atom.

B. Research Progression

As previously described, our early efforts focused on "simply" cataloging the existing database questions in terms of the primary topics addressed. With the goal of expanding this database, we began to examine the questions generated by faculty in the professional development workshops previously described. This confirmed our earlier insights: we needed a more insightful and systematic way to characterize the abundance of information contained within these questions. Specifically, we looked to develop a scheme that could highlight the affordances of the representations and tasks used as well as their potential for promoting critical discernment and helping learners develop their discipline fluency. We ultimately settled on a three-pronged approach grounded in the answers to the following.

- What kinds and how many different ways of conveying information are used?
- What kinds and how many different cognitive exercises must the learner engage in?
- How complex will the discourse be among learners attempting to explain the reasoning behind their answers?

Preliminary work on a topically diverse subset of the data illustrated a marked lack of diversity and complexity in the questions. It became apparent, however, that while certain topics are notably limited in the representations and tasks used, this is not the case across all topics. Additionally (though not unsurprisingly), there are significantly fewer high-complexity questions across all topics.

Eventually, we came to recognize these deficits as opportunities to (1) systematically identify gaps, (2) generate questions to fill those gaps, (3) generate more complex questions that combine multiple representations and tasks in ways not seen in the data, and (4) explore the potential for more pedagogically interesting questions, ones that do some of the unpacking for the learner rather than leave him/her to start from a "blank slate."

IV. THE FRAMEWORK

We created rubrics for classifying how information is represented and the kinds of intellectual tasks one must engage in to answer a question. We refined the Question Complexity Rubric (QCR) of Cormier, Prather, and Brissenden [8] and use it to rate the complexity involved in unpacking and explaining one's reasoning. Applying the initial set of rubrics to a subset of questions led to multiple iterations until we achieved consistency in applying them to additional questions. These rubrics comprise a framework that is easily generalized to other disciplines, types of curricular materials, and active-learning practices.

Modes of Representation	Intellectual Tasks
1. words	1. visualize
a. written	2. draw/sketch
b. spoken	3. model
2. pictures & diagrams	4. compare
a. photographs	5. identify
b. static images	6. predict
c. figures	7. extrapolate
d. sketches	8. count
3. graphs & charts	9. rank
4. tables	10. sort
5. mathematical formalism	11. match
6. numbers	12. quantitative reasoning
7. animations & simulations	13. calculate
a. moving pictures and/or diagrams; with no user interaction	14. apply/analyze
b. simulators with user interaction mechanisms	15. write
8. recordings of reality	
a. video	
b. audio	
9. gestures	
a. facial expressions	
b. body movements	

FIG. 1. Full list of representations and tasks used in the rubrics.

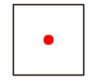
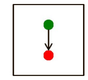
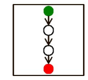
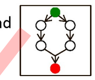
Code	Degree of Cognitive Complexity	To convince someone else of the correct answer, unpacking and explaining requires…	Type of Reasoning	Schematic Visualization
1	trivial	...stating only a single fact or element of declarative knowledge.	recall	
2	low	...only one step of reasoning.	simple	
3	medium	...multiple sequential steps of reasoning.	chain	
4	high	...multiple pathways of sequential reasoning steps involving two or more concepts or topics.	compound	

FIG. 2. The Question Complexity Rubric (QCR).

A. Modes of Representation

A "mode of representation" is a way of conveying information. The numbered modes in Fig. 1 are general types of information delivery widely used in teaching college level science. Some modes have lettered subtypes that serve as common examples and are included to assist in identifying the general mode of representation. PI/TPS questions typically do not make use of numbers 7, 8, or 9 but they are included in the rubric since the framework can easily be extended to other types of instructional materials and methods.

B. Intellectual Tasks

An "intellectual task" is a specific cognitive action that one may perform to arrive at the correct answer to a question. Tasks that are inherent to virtually all questions, such as "recall" or "interpret," are not uniquely helpful when distinguishing differences among questions and therefore not included in this work. While PI/TPS questions do not make use of task number 15 in Fig. 1, it is included in our rubric because the framework is easily generalized to other types of instructional materials and methods.

Most questions naturally require multiple intellectual tasks so it is necessary to differentiate between the chief or *intended* task and any *supporting* tasks. The wording of some questions automatically reveals the intended task. For example "How many of the following..." implies "count," while a question that asks for objects to be arranged in a particular order signals that "rank" is the intended task. Supporting tasks are those that, while not dominant, may still be essential when reasoning about the question. Tasks such as "quantitative reasoning," "compare," and/or "visualize," for instance, might necessarily precede the intended task "rank" and therefore must be included as supporting tasks.

The list of supporting tasks for any given question should include all tasks that are *reasonably likely among a population of learners in the discipline*. Note, however, that a question that includes the supporting tasks "visualize," "model," and "sketch" does not mean that all three tasks are required to answer the question. Rather, one person might find it sufficient to "visualize" while another may need to "model" and/or "sketch" instead of, or in addition to, "visualize." Still, all three tasks must be included when characterizing that particular question.

C. The Question Complexity Rubric (QCR)

Rather than rate the difficulties of the questions themselves, we wanted to understand and describe the questions in terms of their conceptual complexities. A question's QCR code (Fig. 2) ranks both its degree of conceptual and cognitive complexity and the level of complexity involved in unpacking and explaining the reasoning behind the correct answer. In this way, the QCR code represents the level of intellectual engagement required to convince someone else of the correct answer.

V. DISCUSSION AND CONCLUSIONS

We have found the framework described here to be truly *generative* in that it facilitates creating (1) new questions that make use of representations and tasks that are underutilized in the faculty-generated questions, (2) questions that combine multiple representations and intellectual tasks in ways not seen in the data, and (3) more meaningful opportunities for learners to unpack complex concepts, practice critical discernment, and develop discipline fluency.

We created the identifier *fluency-inspiring questions* for those questions that address points (2) and (3) above. One example of a fluency-inspiring question is a unique form of fill-in-the-blank that forces the learner to extract information from one representation and map it onto one or more others while engaging in multiple cognitive tasks. Another ex-

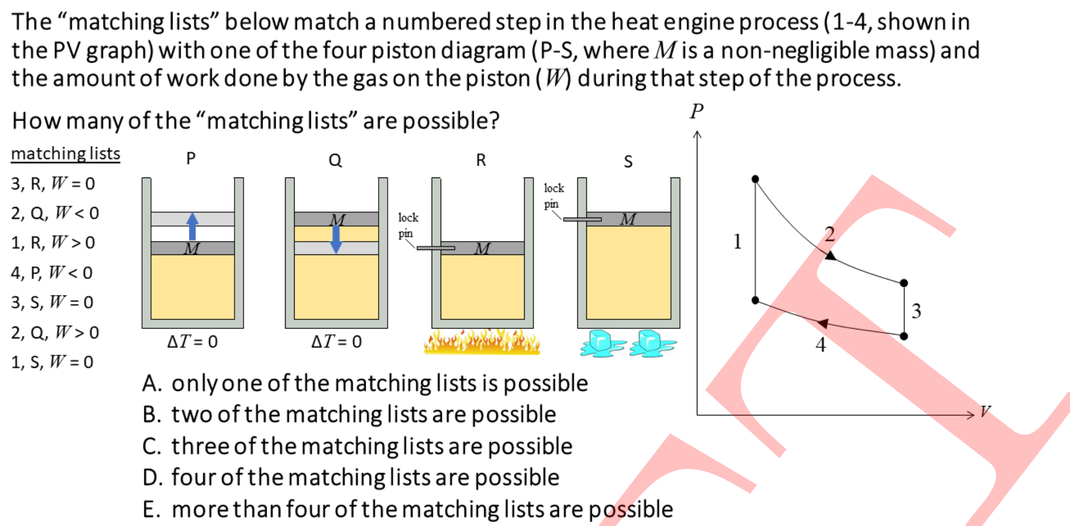


FIG. 3. A fluency-inspiring question created by applying this framework.

ample involves "matching lists" where multiple intellectual tasks are required to match different aspects of multiple representations together. These interesting question constructs automatically force the unpacking of information by leading the learner from one representation to another while simultaneously activating one intellectual exercise after another. Oftentimes, this also means we end up with a question that has the highest possible QCR rank of 4. A specific example of

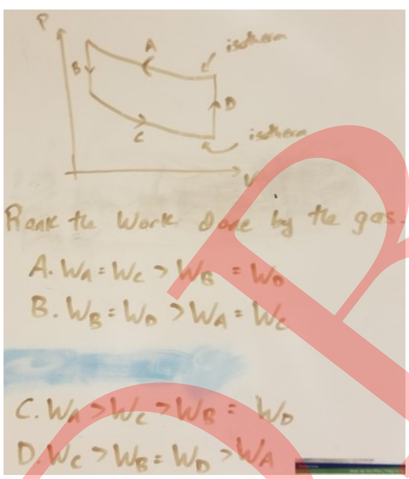


FIG. 4. A typical question created by faculty.

this is seen in the fluency-inspiring question shown in Fig. 3. To create this question, we analyzed the faculty-generated question shown in Fig. 4, identified corresponding tasks and representations, and then applied this framework. The application of this framework to conduct a complete analysis of all 353 multiple-choice questions from § III A is ongoing and includes the development of additional fluency-inspiring questions.

By investigating the variety in faculty-generated PI/TPS questions we have come to realize that our framework is not simply a tool for making new questions: it informs how we think about our disciplines and provides a pathway for exploring more pedagogically interesting and powerful opportunities to help learners develop fluency in the key ideas of the disciplines. We are moving from a place of "we didn't know what we didn't know" to "now we know what we didn't know" and "we know what we need to do about what we didn't know."

ACKNOWLEDGMENTS

We thank Nate Goss for his meaningful input into many sessions iterating on and refining the rubrics and Gina Brissenden for helping track down all of the questions.

[1] E.E. Prather and G. Brissenden, *Astr. Ed. Rev.*, **7**(2), 1 (2008).
[2] https://astronomy101.jpl.nasa.gov/workshops. Retrieved 7/4/2018.
[3] C. Linder, *Eur. J. Sci. and Math. Ed.*, **1**(2), 43 (2013).
[4] J. Airey and C. Linder, *J. Res. in Sci. Teach.*, **46**(1), 27 (2009).
[5] T. Fredlund, C. Linder, J. Airey, and A. Linder, *Phys. Rev. ST - Phys. Ed. Res.*, **10**, 020129 (2014).
[6] T. Fredlund, J. Airey, and C. Linder, *Eur. J. Phys.*, **33**, 657 (2012).
[7] http://www.aapt.org/Conferences/newfaculty/nfw.cfm. Retrieved 7/4/2018.
[8] S. Cormier, E. Prather, and G. Brissenden, in *Earth and Space Science: Making Connections in Education and Public Outreach, ASP Conf. Ser. Vol. 443*, edited by J.B. Jensen, J.G. Manning, and M.G. Gibbs (Astronomical Society of the Pacific, San Francisco, 2011), p. 439.