How physics teachers model student thinking and plan instructional responses when using learning-progression-based assessment information

Alicia C. Alonzo¹ and Andrew Elby²

¹*Dept. of Teacher Ed., Michigan State University, 307 Erickson Hall, East Lansing, MI 48824*

²*Dept. of TLPL, University of Maryland, 2311 Benjamin Bldg., College Park, MD 20742*

One vision for the classroom use of learning progressions (LPs) involves using diagnostic assessments to determine a student's LP "level" in order to make instructional decisions. However, little is known about how experienced teachers reason about assessment information and thus how LP-based information might support their instructional decision-making. In this paper, we explore five experienced teachers' interactions with the same set of LP-based score reports to address the following questions: (1) What assumptions do teachers make about student thinking as they interact with LP-based assessment information? (2) What instructional reasoning is supported by these assumptions? We find that teachers conceptualize and use the LP levels differently from how the LP designers intended, but that the LP-based diagnostic information can be helpful to them in other, sometimes unanticipated, ways.

PACS: 01.40.-d

I. INTRODUCTION

Visions for the classroom application of learning progressions (LPs)—“descriptions of the successively more sophisticated ways of thinking about a topic that can follow one another as children learn” [1]—include using diagnostic assessments to determine a student's LP “level” in order to inform decisions about instructional interventions [2-4]. This idea has led to interest in commercially-developed LP-based instructional materials, such as assessments and curricula [5]. However, little is known about how experienced teachers reason about assessment information and therefore how LP-based information might support their instructional decision-making. In this paper, we use different cognitive models of student thinking as a theoretical framework through which to explore teachers' reasoning about LP-based assessment information. We argue that LP research, by treating student thinking as reasonably consistent with a particular LP level and by assuming that students can therefore be classified as being “at” a particular level, adopts a *theory-like* perspective [6]. In contrast, a *misconceptions* [7] or *knowledge-in-pieces (resources)* [8,9] perspective treats students' knowledge as more fragmented and, in the latter case, more context-dependent. These two perspectives are consistent with evidence suggesting that student thinking in the “messy middle” [10] of LPs may be particularly fragmented and context-dependent [11]. A pilot study of one physics teacher's interactions with LP-based score reports [12] revealed that he expressed all three views of student thinking, switching between reasoning about student ideas in terms of LP levels and finer-grained knowledge elements. His finer-grained interpretations led to more specific, actionable instructional ideas as compared to his LP-based analyses. Expanding on the pilot study, this paper explores an additional four teachers' interactions with

the same LP-based score reports. We address (1) What assumptions do teachers make about student thinking as they interact with LP-based assessment information? (2) What instructional reasoning is supported by these assumptions?

II. METHODS

A. Design & procedure

Our data come from a larger study [13] of seven experienced physics teachers' interactions with score reports derived from a force and motion (FM) LP [14] and set of 16 associated ordered multiple-choice (OMC) [15] items. The teachers were all recommended as employing high-quality formative assessment practices. The teachers all participated in three interviews. The first interview elicited teachers' descriptions of their own formative assessment practices. Before the second interview, teachers were sent a 3-page document describing the general LP construct and the FM LP.

In the second and third interviews, teachers were asked “think aloud” [16] about score reports. To ensure that the teachers interacted with the same information, each teacher worked with the same score report, which was constructed from actual (but pseudonymed) students' responses to the OMC items. Teachers were asked to treat the reports as if they represented data from their own students. The score reports provided 1) LP diagnoses for each student and 2) item text and information about students' responses to each OMC item. After the teachers had finished thinking aloud, the interviewer asked follow-up questions, both for clarification and to probe teachers' reactions to particular features of the score reports.

During the second interview, the teachers worked with a set of paper score reports; during the third interview, which occurred several months later, teachers worked with an

interactive computer-based reporting page, designed in light of the second interviews. The computer-based reporting page allowed teachers to access four types of score reports:

- 1) A class-level score report that provided a level diagnosis for each student;
- 2) A class-level report listing, for each item, the percentage of students choosing answers corresponding to each LP level;
- 3) Reports for each individual student including the student's answer and corresponding LP level for each individual OMC assessment item.
- 4) Item-level reports including the text of each item, with the LP level of each option and the percentage of students who chose each option.

Hyperlinks allowed teachers to navigate among the reports and between the reports and an abridged, expandable version of the FM LP description. Screen shots of the computer-based score report are available in [17].

Interviews were video-recorded, with screen capture of the computer interface added in interview 3. Interviews were transcribed prior to analysis.

B. Analysis

In the pilot study, a coding scheme based on the three models of student cognition described above (LP/theory-like, misconceptions, knowledge-in-pieces) worked well to characterize the teacher's interactions with the LP-based score reports. However, analyzing the transcripts for the other four teachers, we identified additional ways of carving up the conceptual terrain of student thinking about FM. Therefore, using a grounded theory approach [18], we refined our coding scheme to reflect the ways teachers reasoned about student thinking as they worked with the score reports. In addition, we coded each time the teacher proposed an instructional response triggered by information in the score report.

Once a stable coding scheme emerged, we re-coded earlier transcripts, including those from the original teacher. We applied codes to chunks of transcript that seemed to represent the same type of reasoning, ranging from single phrases to long conversational turns. In each round, both authors coded each transcript individually and then discussed the codes and our broader interpretations of the teacher's thinking. Because our primary result is these interpretations that emerged from our codes, rather than counts of specific coding instances, we consider this process of investigator triangulation [19]—in lieu of inter-rater agreement—as crucial to the validity of our results.

III. RESULTS

Based on the first set of interviews, we chose 5 of the 7 teachers in the larger study to include in this analysis. These five described engaging with student ideas in ways that

transcended a “gets it/ doesn't get it” perspective [20] and using their interpretations of student thinking to guide instruction. We judged these five teachers (including the pilot study participant [12]) to have sufficiently sophisticated existing formative assessment practices to make possible a study of how they reason about incorporating LP-based information into those practices. In this paper, we explore general features of these teachers' interactions with the score reports, rather than how variations in their existing formative assessment practices might influence variability within this group.

A. Reasoning about student thinking

Rather than adopting a single, consistent perspective about student thinking, each teacher expressed a variety of perspectives as he/she interacted with the score reports. Although all five teachers seemed to understand the intent of the LP and appropriated language from the LP-based materials to reason about student thinking, this seemed to stem largely from the interview context (i.e., trying to do what they perceived the interviewer wanted). All five teachers usually treated student thinking as less coherent than the LP perspective does. In interview 2, Henry expressed a common sentiment: “I guess I would have to go back and do a little bit of an item analysis and try to figure out why the people in the [level] twos are in the twos.” In other words, while the teachers were able to work with the LP-based information, they did not treat the levels as really telling them what students were thinking.

A full list of the codes we used to describe the assumptions that teachers appeared to be making about student thinking as they worked with the LP-based score reports, with example quotes, is available in [17]. In this subsection, we provide a brief overview of the variation in how teachers characterized the “conceptual terrain” covered by the FM LP. In the next subsection, we explore how these models of student thinking supported teachers' reasoning about instructional responses.

In addition to the three cognitive models described above, teachers made assumptions about student thinking that combined the LP perspective with finer-grained consideration of student thinking, representing a more fragmented view of student thinking than the LP/theory-like framework would suggest. While the FM LP assumes that students will have all of the characteristics at a given level, and that these are likely to manifest in specific common errors, teachers often assumed that students would have some (but not all) of the characteristics and/or some of the common errors associated with a given level. For example, Level 2 of the FM LP includes the ideas that “motion implies a force in the direction of motion,” “non-motion implies no motion,” and “force implies motion in the direction of force,” as well as the well-documented impetus idea, that force is “carried with a moving object” [14]. In interview 2, Tim provided the following diagnosis:

So that's what I'm seeing is that [impetus] seems to be the largest level 2 problem. Whereas the other level 2 problems, which are if there's no motion there are no forces acting on it, it seems like that's not as big of a deal.

We used the code *LP levels divisible* to capture the idea that each LP level consists of separable conceptions rather than a unified conceptual whole. Other codes were used to describe alternative ways of describing student thinking, all characterizing student ideas at a grain size smaller than that of the LP levels.

B. Reasoning about Instructional Responses

When using *LP* assumptions about student thinking, the teachers' reasoning was quite general (e.g., "He must be struggling with some of those [ideas]"), and their instructional responses were limited to topics (e.g., "I would focus on this idea that force causes motion as opposed to force changing motion [the description of level 3]") or general strategies (e.g., "I would focus on getting them out of the level 3 area."). As indicated by the last example, teachers sometimes interpreted the LP as intended: a "map" to guide teachers' work with students. In interview 3, Julia explained: "So, I mean, I guess I look at the purpose of this data is to do interventions to move students along the learning progression and further their learning." She suggested using the LP level diagnoses to "pair my level 4 kids with my level 2 kids to remediate and help them at least to a level 3." As indicated in Julia's statement (and endorsed by LP researchers), teachers often suggested trying to move students from their current level to the one just above. For example, in interview 2, Aaron concluded, "So... a little over a third of the class... needs to be brought up from a level 2 to a level 3. And then the bulk of the class... needs support from that level 3 to level 4." However, Aaron and the other teachers did not take the next step assumed by researchers: comparing the levels to determine what support students might need to make those transitions.

Instructional reasoning based on finer-grained analyses were sometimes vague, too, focusing on ideas to address rather than specific strategies for addressing them. However, by identifying specific ideas to target, even these somewhat vague conclusions provided more actionable information than simply moving from one level to the next without a sense for the differences between levels. For example, in contrast to the strategy of pairing students on the basis of their LP level, Julia (also in interview 3) identified a specific idea to "fix" after conducting a finer-grained (*misconceptions*) item analysis: "Okay, so this idea of force dissipating seems to be tripping up Molly. Okay, so that's a matter of fixing the misconception." Similarly, in interview 2, Aaron took a *misconceptions* perspective to

identify the specific conception (impetus) that students were struggling with:

So right now, looking at this, I have an overwhelming idea in my head that I need to somehow, get the instruction across that the force is independent after the object leaves the hand. After the impulse is done, that force is no longer applied. That's the biggest conception at this point, looking at those.

In both cases, although still not proposing specific instructional responses, the teachers had more to act on than simply "move students to the next level."

Furthermore, although most instructional responses were vague, the rare instances of specific instructional ideas appeared *only* when teachers were attributing finer-grained knowledge (e.g., *misconceptions* or *LP levels divisible*) rather than LP levels to students. Thus, when dividing the conceptual terrain of the LP into smaller units, the teachers' interpretations of student thinking *and* proposed implications for instruction were more specific and actionable. In these finer-grained analyses, teachers tended to look at student responses to specific assessment items, hypothesizing about factors that produced particular patterns of responses and how these factors could be addressed instructionally. For example, using a *LP levels divisible* perspective, Henry identified a specific bullet point of the FM LP level 3 description as the focus of his instruction: "However, student doesn't believe that objects can continue moving with a constant speed without applied force." In response, he thought about what he might do in class, as well as what might underlie the students' thinking:

So I think if we do something, maybe with an airtrack or an air hockey table or something like that and we talk about, "Okay, what's pushing?" Perhaps the students are still not really understanding what a force is. We have to try to work that in there somehow.

Tim, based on responses to multiple items, identified a common difficulty not explicitly addressed in the learning progression. We classified his reasoning with a general *smaller than LP* code but saw hints of a *knowledge-in-pieces* perspective as he thought about why students' reasoning about gravity differed across contexts. He proposed an instructional response drawing on his interpretation of students' difficulty with gravity and his knowledge of their other ideas:

This idea that gravity is this holding force that doesn't let something move... keeps popping up. So we would definitely have to talk about gravity... They have a good intuitive sense of heavier things are affected more by friction. So I would try and base off of that idea... They're telling me they think that gravity,

heavier things affect... friction, which is a good place to start.

In both cases, the teachers used LP-based assessment data *but not the LP levels* to propose specific instructional moves. Thus, while teachers at times displayed their ability to reason with LP-based information as researchers envision, their analyses based on different assumptions about student cognition were more supportive of instructional decision-making.

IV. DISCUSSION & CONCLUSION

In this paper, we question the common assumption that information about students' LP levels will provide useful information for teachers' instructional decision-making. Evidence from teachers with high-quality formative assessment practices indicates that they use different sorts of reasoning about their students' ideas. One might argue that, with further professional development, these teachers could come to understand and use the LP levels more consistently. But this is not necessarily a desirable goal. The LP levels (and associated reasoning about student thinking) seemed to be less useful than a more detailed analysis of students' item responses. Of course, if time for detailed item analyses isn't available, LP diagnoses might be better than nothing. However, if a student is diagnosed at level 3 because he/she understood some areas of FM at a

level 2 and some areas of FM at a level 4, this could result in inappropriate instruction.

It is also possible that LPs may be more useful for teachers less sophisticated with formative assessment. Moving from a "gets it/doesn't get it" perspective to a more nuanced view of student thinking is crucial, and LPs may be well-suited for that. However, we must also pay attention to the effects of research-based tools on teachers with more expertise.

There are, of course, limitations to this study. Teachers worked with a standardized score report, rather than with results for their students. They might reason differently, and perhaps more specifically, with additional knowledge about the students and their learning needs and resources. In addition, the FM LP, associated assessment items, and psychometric model used to create the score reports are by no means perfect. Different results may be obtained with another LP. However, we are concerned that all LP research tends to gloss over the context-dependence of students' thinking and that it is finer-grained information (such as how student ideas vary with context) that teachers may need to truly understand their students' thinking and respond productively.

ACKNOWLEDGMENTS

Funding was provided by NCS Pearson, Inc. Elizabeth Xeng de los Santos conducted the interviews.

-
- [1] NRC, *Taking science to school : Learning and teaching science in grades K-8*. (NAP, Washington, D.C., 2007).
 - [2] A. C. Alonzo, Measurement: Interdisc. Res. & Perspective **9** (2-3), 124-129 (2011).
 - [3] P. Black, M. Wilson and S.-Y. Yao, Measurement: Interdisc. Res. & Perspective **9** (2-3), 71-123 (2011).
 - [4] E. M. Furtak, J. Res. Sci. Teach. **49** (9), 1181-1210 (2012).
 - [5] A. Cromwell and K. McClarty, *Learning progressions for enhanced personalization* (Pearson, 2013).
 - [6] C. Ioannides and S. Vosniadou, Cognitive Science Quarterly **2** (5-61) (2002).
 - [7] J. Clement, Am. J. Phys. **50**, 66 (1982).
 - [8] A. A. diSessa, Cogn. Instr. **10** (2-3), 105-225 (1993).
 - [9] D. Hammer, Am. J. Phys. **68** (S1), S52-59 (2000).
 - [10] A. W. Gotwals and N. B. Songer, Sci. Ed. **94** (2), 259-281 (2010).
 - [11] J. T. Steedle and R. J. Shavelson, J. Res. Sci. Teach. **46**, 699-715 (2009).
 - [12] A. Alonzo and A. Elby, in *Learning and becoming in practice: The International Conference of the Learning Sciences (ICLS)*, edited by J. L. Polman, E. A. Kyza, et al., (ISLS, Boulder, 2014), Vol. 2, pp. 1037-1041.
 - [13] A. C. Alonzo, E. X. de los Santos and J. L. Kobrin, presented at the AERA, Philadelphia, 2014.
 - [14] A. C. Alonzo and J. T. Steedle, Sci. Ed. **93** (3), 389-421 (2009).
 - [15] D. C. Briggs, A. C. Alonzo, C. Schwab and M. Wilson, Educational Assessment **11** (1), 33-63 (2006).
 - [16] K. A. Ericsson and H. A. Simon, *Protocol analysis: Verbal reports as data*. (MIT Press, Cambridge, 1993).
 - [17] A. C. Alonzo and A. Elby, presented at NARST, Chicago, IL, 2015. DOI: 10.13140/RG.2.1.3334.1922
 - [18] K. Charmaz, in *Handbook of qualitative research (2nd ed.)*, edited by N. K. Denzin and Y. S. Lincoln (Sage, Thousand Oaks, CA, 2000), pp. 509-535.
 - [19] N. Denzin, *The research act: A theoretical introduction to sociological methods*. (Prentice Hall, Englewoods Cliffs, 1989).
 - [20] V. K. Otero, J. Teacher Ed. **57** (3), 247-255 (2006).