Error Analysis as a Validation of Learning Progressions

Brent Morgan
University of Memphis
202 Psychology Bldg
Memphis, TN 38152
+1 (901) 338-5041
brent.morgan@memphis.edu

William Baggett
University of Memphis
Department of Computer Science
Memphis, TN 38152
+1 (901) 678-5465
wbaggett@memphis.edu

Vasile Rus
University of Memphis
375 Dunn Hall
Memphis, TN 38152
+1 (901) 678-5259
vrus@memphis.edu

ABSTRACT

Learning progressions (LPs) are a recent educational theory pertaining to student modeling. LPs argue that students with equal test scores may nonetheless have different conceptualizations of the material, with varying degrees of maturity. However, there is little empirical validation for LPs. To this end, we mapped two physics LPs (one predefined, one described in the paper) onto the answer choices of a popular conceptual physics test (the Force Concept Inventory; FCI). We then assessed 444 high school physics students using a pretest-posttest design. Students with more mature incorrect answers on the pretest performed better on the posttest than their less mature counterparts. We discuss implications for theorists and practitioners in learner modeling.

Keywords

Learning Progressions, Learner Models, FCI, Intelligent Tutoring Systems

1. INTRODUCTION

Students arrive at a learning session with a wide variety of backgrounds, skills, knowledge, and abilities. These individual differences imply that each person cannot be expected to learn the same way and at the same speed as another [3]. A set of learner characteristics which impact learning is referred to as a learner model. A learner model allows instruction to be tailored to each individual student with the goal of maximizing learning for that student. As a student progresses through a learning session, her learner model is updated based on the quality of her contributions.

One novel approach to learner modeling relies on a research framework called Learning Progressions (LPs), developed recently by the science education research community as a way to increase adaptivity in traditional instruction. LPs have been defined as "descriptions of the successively more sophisticated ways of thinking about an idea that follow one another as students learn" [12, 13]. LPs provide a promising means to organize and align content, instruction, and assessment strategies to give students the opportunity to develop deep and integrated understanding of concepts. LPs can be viewed as incrementally more sophisticated ways to think about an idea that emerge naturally while students move toward expert-level understanding of the idea [4]. LPs define qualitatively different levels of understanding of big ideas. The levels can be sequentially related or the relation could be more complex. For instance, topic A may develop the ideas from a less sophisticated topic B but also connect to other topics. LPs are organized in levels of understandings which reflect major milestones in learners' journey towards mastery. The lower level, called the Lower Anchor, represents naïve thinking typically associated with novices. The top level, called the Upper Anchor, represents the mastery/expert level of understanding.

Each student's LP level can presumably be determined from the quality of her contributions, just as with a generic learner model. Traditional assessment typically treats all wrong answers as equally wrong. An LP framework, in contrast, would argue that not all wrong answers are equivalent, and that two answers, while both incorrect, may reflect vastly different milestones on the path to mastery. Thus, students with similar assessment scores may still have vastly different understandings of the topic. In this case, the LP theory would expect a student at a higher LP level would reach mastery faster than the student at a lower LP level. Despite an increase in the popularity of learning progressions, there is surprisingly scant empirical evidence to support these claims [16].

In this paper, we describe a new LP (i.e., relative LP levels of various responses) for two physics topics (Force & Motion, Newton's 3rd Law) on a popular physics assessment tool, the Force Concept Inventory (FCI). Each incorrect answer choice was classified as a higher LP level answer or lower LP level answer. Using pretest and posttest FCI scores collected from 444 high school physics students, we also report a preliminary investigation into the efficacy of these FCI LP maps. A learning progression framework would predict that students with generally 'better' incorrect answers on a given pretest topic would be closer to mastery than those with 'worse' incorrect answers, and this would be reflected by higher posttest scores for those closer to mastery. If this is the case, it would provide data-driven (or bottom-up) support of our conceptually-driven (top-down) LP map.

2. MATERIALS

2.1 Force Concept Inventory

The Force Concept Inventory (FCI) is a 30-item multiple-choice "test" designed to assess students' understanding of the *most basic* concepts in Newtonian mechanics [7]. The FCI presents students with various situations and asks them to choose between Newtonian explanations for the phenomena, versus common-sense alternatives [8]. The FCI has been widely used to measure learning in introductory physics courses. For example, Hake [6] in combination with Coletta and Phillips [2] used the FCI to measure learning in 73 university and college introductory physics classes. Its popularity among researchers was a major motivation for developing an LP map for the FCI as part of our DeepTutor project whose aim is to develop the first intelligent tutoring system based on learning progressions [15]. Another attraction was that the FCI was designed to identify known misconceptions that students often possess [8]. The vast majority of these "lures" can easily be classified and incorporated into an LP framework.

2.2 Developing LPs

The FCI covers multiple concepts in Newtonian mechanics, including Free Fall Near Earth, Circular Motion, and Newton's 3rd Law. The predominant concept, however, is Force & Motion. The Force & Motion LP used for this paper was identical to the one developed by Alonzo and Steedle [1]. The higher levels were defined so as to meet national standards for 8th grade students. The 8th grade Force & Motion standards are applicable to the high school standards and match the top level of understanding expressed in the FCI. The lower levels were then populated by compiling student's ideas about Force & Motion reported in the literature. These ideas were then ordered by relative difficulty. The LP was then iteratively revised based on data collected from physics novices.

Although the FCI is predominantly focused on Force & Motion, it does address other Newtonian concepts, such as Newton's 3rd Law, which are not mapped by Alonzo and Steedle [1]. Hence, a Newton's 3rd Law Learning Progression was developed with the direction of two physics professors. The method used to develop this LP was based on Alonzo and Steedle's [1] process described above.

First, we defined the top level of the hypothetical Newton's Third Law LP as the knowledge needed to articulate and apply Newton's Third Law. The knowledge of Newton's Third Law specified by the top level of our LP matches that specified for grade levels 9-12 in the National Science Education Standards [11]. The lower levels were student's ideas about Newton's Third Law reported in the literature [e.g., 9, 10, 17] and ordering these ideas based on suggestions from the literature and/or the intuitions of two physics professors regarding the relative difficulty.

Next, we collected responses from 30 paid workers from Amazon Mechanical Turk, each of whom answered 22 open-ended Newton's 3rd Law questions. The responses were coded according to the LP, and refinements to the LP were made as necessary to accommodate student's responses. Table 1 presents the revised LP for Newton's Third Law.

A team of two physics professors and two authors of this paper collaborated to map each of five answer choices from each of nine FCI questions according to the LPs. The LP map is shown in Table 2. Five of the questions corresponded to the Force & Motion topic, and four addressed Newton's 3rd Law. These questions were specifically selected for this analysis for two reasons. First, each question and answer choice was related only to one specific topic. Second, the incorrect answer choices exhibited a distinct and unambiguous separation in quality of comprehension according to the physics professors.

Table 1. Newton's 3rd Law LP

5	The student understands that all forces arise out of an interaction between two objects and that these forces are equal in magnitude and opposite in direction.
4	The student identifies equal force pairs, but indicates that both forces act on the same object. (For the example of a book at rest on a table, the downward gravitational force exerted by the earth on the book and the upward normal force exerted by the table on the book are identified as an action-reaction pair.)
3	The student uses the effects of a force as an indication of the relative magnitudes of the forces in an interaction.
2.5	The student indicates that the forces are equal because of the properties of the objects involved.
2	The student indicates that the forces in a force pair do not have equal magnitude because the objects are dissimilar in some property (e.g., bigger, stronger, faster).
1	The student believes that inanimate/passive objects cannot exert a force.
0	No statement about relevant interaction forces or Newton's 3rd Law

Table 2. LP map for upper and lower levels

Newton's Third Law Question and Answer maps

	N3L1	N3L2	N3L3	N3L4
Upper	a, d	b, c	b, c	d
Lower	b, c	d, e	d, e	a, b, c

Force & Motion Question and Answer maps

	FM1	FM2	FM3	FM4	FM5
Upper	a, d	c, e	d	a, b	a, b
Lower	c, e	b, d	a, b, e	c, d	d, e

note: We have removed the actual question numbers to avoid making the FCI answer key public knowledge. Please contact one of the authors for the actual FCI question numbers.

2.3 Data Collection

We administered the FCI to 444 students at three public and two private high schools in the mid-south region, across six teachers and 26 classrooms. The students completed the FCI twice, once at the beginning of the semester (pretest) and once at the end (posttest). Between those two time periods, each classroom covered topics relevant to the FCI, though individual course content varied. Students completed the FCI via provided scantron sheets, which were then collated and processed. The results of the scantron sheets were then compared to direct markings on the actual FCI test in the case of blank or unidentifiable scantron responses. There were five students with perfect scores on the five Force & Motion questions and 19 students with perfect scores on the four Newton's Third Law questions. These students were not included in the respective analyses below.

3. RESULTS

Prior student knowledge was assessed using the initial administration of the FCI (pretest). The mean proportion score of the FCI pretest was 0.26, with a standard deviation of 0.15 (see Table 3). Six students recorded the minimum observed score (1/30), whereas one student attained a perfect score (30/30). A one-sample t-test indicated the mean pretest score was higher than chance (0.2), $t(443) = 7.57$, $p < .001$, though only slightly. Low prior knowledge was ideal for our purposes, of course, providing a large sample of incorrect answers.

Table 3. FCI Pre-Post descriptives by school

Course	N	Pre M (SD)	Post M (SD)	*d*
Public 1 (Pu1)	116	**0.20** (0.10)	**0.22** (0.11)	0.13
Private 1 (Pr1)	25	**0.19** (0.09)	**0.29** (0.10)	1.03
Public 2 (Pu2)	94	**0.27** (0.14)	**0.42** (0.16)	0.94
Private 2 (Pr2)	128	**0.33** (0.20)	**0.47** (0.22)	0.68
Public 3 (Pu3)	81	**0.21** (0.10)	**0.29** (0.13)	0.74
Total	444	**0.26** (0.15)	**0.35** (0.19)	0.55

After assigning student answer choices to the corresponding LP levels described above, we compared the posttest performance of students with more incorrect answers corresponding to the upper level of the LP map on the pretest with students who had more incorrect answers corresponding to the lower level. For example, a student with one upper level answer and three lower level answers would be assigned to the lower level group (irrespective of number of answers correct). To be conservative, students with an equal number of upper and lower level answers were assigned to the lower level.

The comparison of posttest scores, including descriptive statistics, independent samples *t*-tests, and Cohen's *d*, is displayed in Table 4. Across both topics, students in the upper level had higher posttest scores than students in the lower level. Additionally, the findings were associated with small to medium effect sizes. Although this provides initial support for the LP hypothesis, it is possible that these differences in posttest scores are actually being driven by prior knowledge (i.e., pretest scores). Accordingly, an analysis of covariance (ANCOVA) was used, with pretest scores as a covariate to control for prior knowledge. Even when taking pretest scores into account, there were still differences in overall posttest scores between the upper and lower levels for the Force & Motion LP, $F(1, 418) = 12.78$, $p < .001$, $\eta_p^2 = .03$. Differences on overall posttest scores between the upper and lower levels for the Newton's 3rd Law LP were marginally significant, $F(1, 405) = 2.92$, $p = .088$, $\eta_p^2 = .01$. The marginal significance is likely due to the fact that the five Force & Motion questions are more relevant to the rest of the FCI than the four Newton's 3rd Law questions.

Table 4. Posttest descriptives for lower vs. upper level

	Lower Level		Upper Level			
Topic	N	M (SD)	N	M (SD)	*t*	*d*
N3L	115	0.30 (0.17)	310	0.35 (0.18)	**2.50**	**0.28***
FM	131	0.30 (0.15)	308	0.37 (0.19)	**3.70**	**0.40***

* *p* =< .01; N3L = Newton's 3rd Law; FM = Force & Motion

4. DISCUSSION

In this paper, we presented a method used to develop a novel Newton's 3rd Law Learning Progression. This LP as well as a previously developed Force & Motion LP were then mapped onto a popular physics assessment tool, the Force Concept Inventory. FCI pretest and posttest scores were collected from over 400 high school physics students. Unlike traditional assessment, an LP-based student model assumes that not all wrong answers are equally wrong. Hence, we predicted that students whose incorrect pretest answer choices corresponded to a relatively lower LP level would be further away from mastery, and this would then be reflected in students' posttest scores. The results provided support for this claim, and are also among the earliest evidence-based support for learning progressions in general.

It should also be noted that the FCI LP map was able to predict overall (30-item) posttest scores based only on *incorrect* pretest answers of, at most, five questions. The results were also agnostic to instruction type and quality. Relatively stronger effects were observed with the Force & Motion topic than with Newton's 3rd Law. This is likely due to the fact that many other FCI questions draw on Force & Motion comprehension, whereas none of the other twenty-six FCI questions apply to Newton's 3rd Law outside of the four discussed in this paper.

As this is a preliminary report, there are of course many limitations. To begin, the FCI is perhaps not an ideal tool to investigate learning progressions. It was not specifically developed with learning progressions in mind, and the answer choices for most questions do not represent all possible LP levels. For example, all 20 answer choices for the Newton's 3rd Law topic only represented three out of the possible six LP levels. Also, many FCI questions contain answer choices which apply to topics not directly related to the topic addressed by the question (though again, none of these questions were included in this paper). Despite these flaws, however, the FCI is popular, and many researchers may be able to apply a learning progressions framework to future or even past FCI datasets [14].

Also, the analyses included in this brief report are relatively simple and not comprehensive. As such, although the findings provide support for the learning progressions hypothesis, much more evidence is needed to fully validate our Force & Motion and Newton's 3rd Law LPs.

Given the limitations of the FCI mentioned previously, there is also a market for a comprehensive physics assessment which features answers at a variety of LP levels for each question. For example, a question with the correct answer at LP level 4 and all four incorrect answers at LP level 1 offers nothing to the student model for students with incorrect answers.

Another next step is to incorporate learning progressions into the learner models of Intelligent Tutoring Systems (ITS; [15]). One of the multiple advantages of ITSs over traditional classroom instruction is the capacity to adaptively tailor instruction to meet the needs of each and every learner [18, 5]. These systems can then be scaled up to teach many users at once. This would allow for a more thorough investigation into learning progressions, including experimental evidence as to whether adaptive instruction is beneficial for students at different LP levels. Furthermore, advances in natural language processing may allow us to detect LP differences in the text of student contributions. This could perhaps eliminate the need for a pretest for ITSs.

Finally, we encourage physics education researchers to consider incorporating learning progressions into their learner models. To that end, we are currently preparing a manuscript which will report the full LP map for all 30 FCI questions.

5. ACKNOWLEDGMENTS

This research was supported in part by Institute for Education Sciences under award R305A100875. Any opinions, findings, and conclusions or recommendations expressed in this material are solely the authors' and do not necessarily reflect the views of the sponsoring agencies.

6. REFERENCES

[1] Alonzo, A. C., and Steedle, J. T. Developing and assessing a force and motion learning progression. *Science Education*, 2009, 93, 389-421.

[2] Coletta, V. P., and Phillips, J. A. Interpreting FCI scores: Normalized gain, preinstruction scores, and scientific reasoning ability. *American Journal of Physics*, 2005, 73, 1172-1182.

[3] Corcoran, T., Mosher, F. A., and Rogat, A. 2009. Learning progressions in science: An evidence-based approach to reform. *Consortium for Policy Research in Education Report #RR-63*. Philadelphia, PA: Consortium for Policy Research in Education.

[4] Duschl, R., Maeng, S. and Sezen, A. Learning progressions and teaching sequences: a review and analysis, *Studies in Science Education*, 2011, 47, 123-182.

[5] Graesser, A. C., D'Mello, S. K., Hu, X., Cai, Z., Olney, A., and Morgan, B. AutoTutor. In P. M. McCarthy, & C. Boonthum (Eds.), Applied natural language processing and content analysis: Identification, investigation and resolution (pp. 169-187). Hershey, PA: IGI Global, 2012.

[6] Hake, R. R. Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American journal of Physics*, 1998, 66, 64-74.

[7] Halloun, I., Hake, R. R., Mosca, E. P., and Hestenes, D. 1995. Force Concept Inventory (Revised, 1995); online (password protected) at http://modeling.asu.edu/R&E.Research.html

[8] Hestenes, D., Wells, M., and Swackhamer, G. Force concept inventory. *The physics teacher*, 1992, 30, 141-158.

[9] Kolokotronis, D., and Solomonidou, C. A Step-by-Step Design and Development of an Integrated Educational Software to Deal with Students' Empirical Ideas about Mechanical Interaction. *Education and Information Technologies*, 2003, 8, 229-244.

[10] Minstrell, J. (n. d.). Facets of students' thinking. Retrieved May 12, 2014, from http://depts.washington.edu/huntlab/diagnoser/facetcode.html

[11] National Research Council. National Science Education Standards. (NSES) Washington, DC: National Academy Press, 1996.

[12] National Research Council. Systems for state science assessment. (M. R. Wilson & M. W. Bertenthal, Eds.). Washington, DC: National Academy Press, 2005.

[13] National Research Council. Taking science to school: Learning and teaching science in grades K–8. (R. A. Duschl, H. A. Schweingruber, and A. W. Shouse, Eds.). Washington: The National Academies Press, 2007.

[14] Neumann, I., Fulmer, G. W., Liang, L. L., and Neumann, K. Analyzing the FCI based on a force and motion learning progression. *Science Education Review Letters*, 2013, 8-14.

[15] Rus, V., D'Mello, S. K., Hu, X., and Graesser, A. C. Recent Advances in Conversational Intelligent Tutoring Systems, AI Magazine, 2013.

[16] Sikorski, T. R., and Hammer, D. 2010. A critique of how learning progressions research conceptualizes sophistication and progress. In K. Gomez, L. Lyons, & J. Radinsky (Eds.), *Learning in the Disciplines: Proceedings of the 9th International Conference of the Learning Sciences* (ICLS 2010) - Volume 1, Full Papers. Chicago, IL: International Society of the Learning Sciences, Chicago, IL, 2010.

[17] Smith, T. I., and Wittmann, M. C. Comparing three methods for teaching Newton's third law. *Physical Review Special Topics-Physics Education Research*, 2007, 3(2), 020105.

[18] VanLehn, K. The behavior of tutoring systems. *International Journal of Artificial Intelligence in Education*, 2006, 16, 227-265.