Physics Students' Implicit Connections Between Mathematical Ideas

Trevor I. Smith
Rowan University

Suzanne White Brahmia
University of Washington

Alexis Olsho
University of Washington

Andrew Boudreaux
Western Washington University

The Physics Inventory of Quantitative Literacy (PIQL) aims to assess students' physics quantitative literacy at the introductory level. PIQL's design presents the challenge of isolating types of mathematical reasoning that are independent of each other in physics questions. In its current form, PIQL spans three principle reasoning subdomains previously identified in the research literature: ratios and proportions, covariation, and signed (negative) quantities. An important psychometric objective is to test the orthogonality of these three reasoning subdomains. We present results that suggest that students' responses to PIQL questions do not fit this structure. Groupings of correct responses identified in the data provide insight into the ways in which students' knowledge may be structured. Moreover, questions with multiple correct responses may have different responses in different data-driven groups, suggesting that the both the answer choice and the context of the question may impact how students (implicitly) relate various ideas.

Keywords: Physics, Quantitative Literacy, Psychometric Analyses

One major goal of university-level physics courses is the development of mathematical reasoning skills, but despite decades of research on the complex interplay between physics conceptual understanding and mathematical reasoning (Boudreaux, Kanim, & Brahmia, 2015; Brahmia, Boudreaux, & Kanim, 2016a, 2016b; Rebello, Cui, Bennett, Zollman, & Ozimek, 2007; Sherin, 2001; Thompson, 2010), measuring these skills has not gained as much popularity as strictly conceptual assessments such as the Force Concept Inventory or Force and Motion Conceptual Evaluation (Madsen, McKagan, & Sayre, 2017; Madsen, McKagan, Sayre, & Paul, 2019).

Assessing conceptual understanding is inherently easier than assessing generalized mathematical reasoning. The former is tied to specific physics contexts taught over a finite period, while the latter is ubiquitous across contexts and time. Physics education researchers have conducted qualitative case studies to probe students' mathematical reasoning and their transitions to expert-like reasoning (c.f., Hayes & Wittmann, 2010; Hu & Rebello, 2013; Smith, Thompson, & Mountcastle, 2013). While this method of research provides a rich view into how specific students reason in a particular context, little has been published that characterizes the process of emerging expert-like mathematical reasoning across multiple topics and for a large group of students.

We have developed the Physics Inventory of Quantitative Literacy (PIQL) to meet the need fora robust and easily administered multiple-choice assessment to measure students' mathematical reasoning in the context of physics (a.k.a., physics quantitative literacy, PQL) (Olsho, White Brahmia, Boudreaux, & Smith, 2019).The PIQL is intended to test three key components of PQL: ratio and proportion (Cohen & Kanim, 2005), covariation (Carlson, Oehrtman, & Engelke, 2010; Hobson & Moore, 2017; Moore, Paoletti, & Musgrave, 2013;

Paoletti & Moore, 2017), and negativity (Brahmia & Boudreaux, 2017; Vlassis, 2004; White Brahmia, Olsho, Smith, & Boudreaux, 2018, 2019, in press).

We consider PQL to be a conceptual blend between physics concepts and mathematical reasoning (Fauconnier & Turner, 2002). In order to measure the complexity of ideas that students bring from both of these input spaces, we have chosen to include some multiple-choice multiple-response (MCMR) questions in which students are instructed to "select all statements that **must be true**" from a given list, and to "***choose all that apply***" (emphasis in the original text). The MCMR question format has the potential to reveal more information about students' thinking than standard single-response (SR) questions, but it also poses problems with data analysis, as typical analyses of multiple-choice tests assume SR questions. We have previously compared two different methods that could be used to identify groups of questions evident in students' responses to PIQL questions and compare them to the groups defined by our three PQL constructs (Smith et al., in press). In this paper we briefly describe these methods and present our preliminary results from using module analysis for multiple-choice responses, which allows each correct response to MCMR questions to be included separately (Brewe, Bruun, & Bearden, 2016).

Data for this study were collected in two different terms at a comprehensive public university in the Northwestern United States. The PIQL was given as a pretest during the first week of the term in three different calculus-based introductory physics classes: Mechanics ($N = 821$), Electricity and Magnetism ($N = 701$), and Thermodynamics and Waves ($N = 585$). These data do not form a matched set, but we take them as three snapshots in time, which may be representative of a progression through the introductory course sequence: before mechanics (PreMech), after mechanics and before electricity & magnetism (PostMech), and after electricity& magnetism (PostEM). Previous analyses of these data have shown that students' overall scores on the PIQL increase over time, so we can consider students progressing toward expertise through our data (Smith et al., 2018, in press).

We have previously reported preliminary results from applying module analysis to identify coherent groups of PIQL questions and responses (Smith et al., in press). In the current study we interpret these results in new ways that focus specifically on the questions related to covariation. Our work is guided by the research question: In what ways do patterns of students' responses to PIQL questions reveal insights into the ways in which they think about and use covariation in physics contexts?

Previous Results: Coherent Modules Identified from Students' Responses

We have previously used both confirmatory and exploratory factor analysis to identify groups of questions evident in students' response patterns to PIQL questions and compare those groups with our pre-defined PQL constructs (Smith et al., in press). These results showed a poor alignment between response patterns and our three primary constructs of PQL: ratio and proportion (questions 1–6), covariation (questions 7–14), and negativity (questions 15–20); however, a major limitation of factor analysis is that questions must be scored dichotomously as either completely correct or incorrect. This is problematic for MCMR questions because a student who chose one correct response to a question with two correct responses would be coded the same way as a student who chose multiple incorrect responses; therefore, we choose not to emphasize these results here.

To preserve the nuance and complexity of students' response patterns within (and between) questions we used module analysis for multiple-choice responses to examine the network of student responses to PIQL questions (Brewe et al., 2016). Module analysis uses community detection algorithms to identify modules (a.k.a. communities, clusters, etc.) within networks of responses to multiple-choice questions. We have chosen to analyze a network of only correct responses to PIQL questions. The benefit of this method is that we can examine the patterns that arise from students' selections of each individual correct response, which preserves some of the complexity of module analysis questions by recognizing any time a student chooses a correct response. A limitation of the way that we have used module analysis is that we are ignoring whether or not a student chooses incorrect responses in addition to correct responses. Expanding the network to include correct and incorrect responses could address this limitation, but is beyond the scope of the current study.

In network analysis studies, the choice of community detection algorithm seems to depend a lot on personal preference of the researchers. Unfortunately, the InfoMap algorithm used by Brewe et al. (2016) did not yield useful results.[1] In the absence of clear guidelines regarding which community detection algorithm would be most relevant, we chose to compare the modules identified by six different algorithms. We feel confident that modules that are identified by multiple community detection algorithms are representative of the data.

A major result from our analyses was that the modules were not consistent in our three time-dependent data sets. The results went from four modules in PreMech (two with only two responses each) to six modules in PostEM (most with only 2–3 responses). Contrast this to what might be expected for a hypothetical group of experts: true experts would answer all questions correctly, resulting in strong links between all correct responses, and all responses being in one coherent module. Our data show that as students progress toward expertise during the introductory sequence, modules become less coherent, not more. Additional data from upper-division students are needed to examine the continuation of this progression.

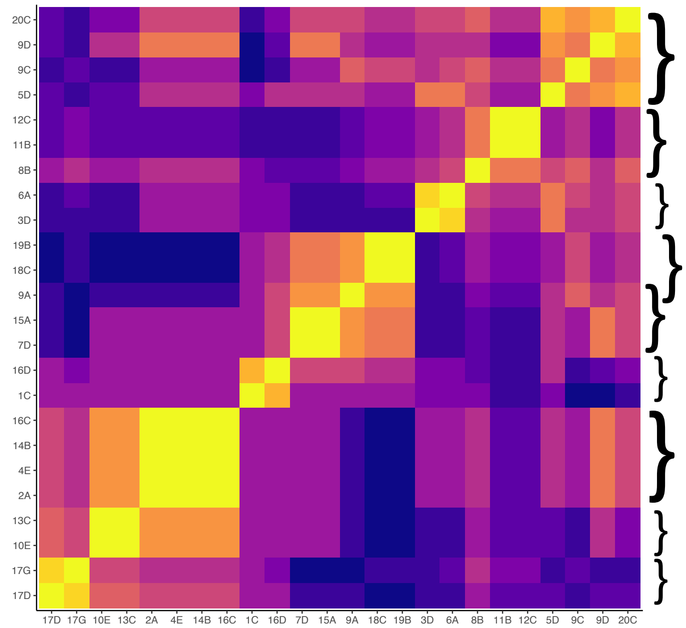
The changes in module definitions over time led us to look for consistent patterns across the results, which may represent stable elements of student reasoning. Figure 1 shows the average likelihood that each question pair occurs in the same module as well as the "submodules" that we have identified as being consistent across our analyses. Each of these submodules may be seen as a bright yellow/orange square along the diagonal in Figure 1, with submodule i (in the upper right corner) being the least cohesive (least bright). Some submodules are subsets of our PQL constructs: ratio and proportion (iii), covariation (ii and viii), and negativity (ix). Others include questions from two or three of these constructs (i, iv, v, vi, and vii), emphasizing the connections between these constructs.

The MCMR questions with more than one correct response show some particularly interesting trends. Question 17 has two correct answers (17D, 17G) that group very strongly together. Question 16 also has two correct responses (16C, 16D), but they do not group into the same submodule. Question 9 has three correct responses: 9C and 9D are in submodule i (which is the least coherent submodule mentioned above), and 9A groups equally well with two different submodules, neither of which connects with 9C or 9D.

New Insights: Covariation Questions Group Together and Split Apart

To answer our research question we consider the questions intended to assess students' PQL regarding covariation (7–14). Some of these items were taken directly (with permission) from the Precalculus Concept Assessment (PCA) (Carlson et al., 2010): question 8 asks students to interpret the slope of a graph of a function (modified from the PCA by graphing speed as a function of time), questions 11 and 12 are bottle questions for which students need to either

[1] Wells et al. (2019) report similar difficulty using the InfoMap algorithm.



Submodule Definitions

i) **5**, *9C*, *9D*, 20

ii) *8*, *11*, *12*

iii) **3**, **6**

iv) 18, 19, (*9A*)

v) *7*, 15, (*9A*)

vi) **1**, 16D

vii) **2**, **4**, *14*, 16C

viii) *10*, *13*

ix) 17D, 17G

Figure 1. Previous results from applying module analysis to PIQL data (Smith et al., in press). The heat map shows the average cooccurrence matrix from the three data sets (yellow indicating responses that are always in the same module, dark blue indicating responses that are never in the same module). The bright squares along the diagonal are used to identify the consistent "submodules" across data sets. Questions intended to probe ratio and proportion are shown in bold in the definition list, questions for covariation are in italics, and questions for negativity are underlined. Response 9A fits equally with two submodules.

select a bottle to match a given graph (11) or select a graph to match a given bottle (12), and question 14 asks students to interpret a ratio expression with a variable in the numerator and denominator to make claims about both the rate of change of the ratio and its limiting value (in the context of fish in a lake). Question 7 is a variation on the students-professors question in which students must select an equation to represent the statement "There are three times as many quarks as nucleons." Question 9 tells students that the length and width of a flag both increase by a factor and asks which, if any, of the following quantities also increase by the same factor: perimeter, area, length of diagonal, and/or length of curve superimposed on the diagonal. For question 10 students must compare the distances traveled by two joggers who run at different speeds for different amounts of time, and for question 13 students are given the equation $m = k\frac{p}{3n^2}$ and asked what happens to m if both n and p double (with k held fixed).

As seen in Figure 1, these questions show up in five out of the nine submodules, but only two submodules include only covariation questions (ii and viii). We identify commonalities between the questions that appear in covariation-only submodules. Questions 8, 11, and 12 (submodule ii) all require students to interpret graphs, and these are the only such questions on the PIQL. Questions 10 and 13 (submodule viii) both require students to determine the output of a known function with two input variables. The questions in each of these submodules test students' abilities to use a unique and somewhat sophisticated type of reasoning.

It is more difficult to identify why other questions/responses group together (or don't). It makes sense that students would choose 9C (the curve along the diagonal of the flag) at similar rates as they choose 9D (the diagonal itself). But why is 9A (the perimeter of the flag) more strongly connected to questions 7 and 15 (both of which require students to select an equation based on a description) or to questions 18 and 19 (which involve interpreting negative vector quantities)? Questions 2, 4, and 14 all involve ratios, but 16C requires students to correctly compare a positively charged sphere and a negatively charged sphere, which does not (on the surface) seem to be related. More work is needed to reveal why students' responses group in these particular ways.

Summary and Future Directions

As mentioned above, the submodules identified by module analysis do not correspond with the groups defined by our PQL constructs. This suggests that either a) students' PQL cannot be separated into skills regarding ratio and proportion, covariation, and negativity, or b) their skills in these areas have developed similarly such that they are functionally equivalent. Regardless of the interpretation, module analysis reveals complexity and structure that changes over time as students progress through the introductory physics course sequence.

Several questions still remain:

- How do the modules identified by students' responses to PIQL questions change over time throughout the undergraduate curriculum?
- How sensitive are these modules to different forms of instruction? Are they the same at different institutions or in different courses?
- How do students' choices of incorrect responses relate to their choices of correct responses (especially for MCMR questions)? How often do students contradict themselves?
- Are there underlying commonalities that we can identify in each module? How do these relate to previous literature on quantification and quantitative reasoning or students' conceptual understanding of physics?

Module analysis opens the possibilities for future work that goes beyond analysis of only correct responses by identifying modules of incorrect responses as well (Brewe et al., 2016). Including incorrect responses in the module analysis could provide evidence to explain why some questions and responses group together in unexpected ways. This is particularly important for interpreting responses to MCMR questions, as different (in)correct responses can reveal insights into different aspects of a student's understanding regarding both physical concepts and mathematical reasoning. This interplay is essential to measuring PQL.

We plan to look more closely at the dynamics of the defined modules over time by using matched sets of responses collected from the same students at different times, and by expanding data collection beyond the introductory sequence. These longitudinal data will allow us greater confidence in claims regarding how students' response patterns change over time. Future work will also include expanding data collection beyond a single university. The coupling of PIQL MCMR questions with module analysis shows promise for finding patterns of emergent expertise in mathematical reasoning in introductory physics, and beyond, on a scale that cannot be achieved using qualitative research methods.

Acknowledgments

This work is supported by the National Science Foundation under grants DUE-1832836, DUE-1832880, and DUE-1833050.

References

Boudreaux, A., Kanim, S., & Brahmia, S. (2015). Student facility with ratio and proportion: Mapping the reasoning space in introductory physics. Retrieved from https://arxiv.org/abs/:1511.08960

Brahmia, S., & Boudreaux, A. (2017). Signed Quantities: Mathematics Based Majors Struggle to Make Meaning. In A. Weinberg, C. Rasmussen, J. Rabin, M. Wawro, & S. Brown (Eds.), *Proceedings of the 20th Annual Conference on Research in Undergraduate Mathematics Education.* San Diego, CA.

Brahmia, S., Boudreaux, A., & Kanim, S. E. (2016a). Developing mathematical creativity with physics invention tasks. Retrieved from https://arxiv.org/abs/1602.02033

Brahmia, S., Boudreaux, A., & Kanim, S. E. (2016b). Obstacles to Mathematization in Introductory Physics. Retrieved from https://arxiv.org/abs/1601.01235

Brewe, E., Bruun, J., & Bearden, I. G. (2016). Using module analysis for multiple choice responses: A new method applied to Force Concept Inventory data. *Physical Review Physics Education Research*, *12*(2), 20131. doi: 10.1103/PhysRevPhysEducRes.12.020131

Carlson, M., Oehrtman, M., & Engelke, N. (2010). The precalculus concept assessment: A tool for assessing students' reasoning abilities and understandings. *Cognition and Instruction*, *28*(2), 113–145. doi: 10.1080/07370001003676587

Cohen, E., & Kanim, S. E. (2005). Factors influencing the algebra "reversal error". *American Journal of Physics*, *73*(11), 1072–1078. doi: 10.1119/1.2063048

Fauconnier, G., & Turner, M. (2002). *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*. New York: Basic Books.

Hayes, K., & Wittmann, M. C. (2010). The role of sign in students' modeling of scalar equations. *The Physics Teacher*, *48*(4), 246–249. doi: 10.1119/1.3361994

Hobson, N. L. F., & Moore, K. C. (2017). Exploring Experts' Covariational Reasoning. In A. Weinberg, C. Rasmussen, J. Rabin, M. Wawro, & S. Brown (Eds.) *Proceedings of the 20th Annual Conference on Research in Undergraduate Mathematics Education.* (pp. 664–672). San Diego, CA.

Hu, D., & Rebello, N. S. (2013). Understanding student use of differentials in physics integration problems. *Physical Review Special Topics – Physics Education Research, 9*(2), 20108. doi: 10.1103/PhysRevSTPER.9.020108

Madsen, A., McKagan, S. B., & Sayre, E. C. (2017). Resource Letter RBAI-1: Research-Based Assessment Instruments in Physics and Astronomy. *American Journal of Physics*, *85*(4), 245–264. doi: 10.1119/1.4977416

Madsen, A., McKagan, S. B., Sayre, E. C., & Paul, C. A. (2019). Resource Letter RBAI-2: Research-based assessment instruments: Beyond physics topics. *American Journal of Physics*, *87*(5), 350–369. doi: 10.1119/1.5094139

Moore, K. C., Paoletti, T., & Musgrave, S. (2013). Covariational reasoning and invariance among coordinate systems. *Journal of Mathematical Behavior*, *32*(3), 461–473. doi: 10.1016/j.jmathb.2013.05.002

Olsho, A., White Brahmia, S., Boudreaux, A., & Smith, T. I. (2019). The physics inventory of quantitative reasoning: Assessing student reasoning about sign. In A. Weinberg, D. Moore-Russo, H. Soto, & M. Wawro (Eds.), *Proceedings of the 22nd Annual Conference on Research in Undergraduate Mathematics Education.* (pp. 992–997). Oklahoma City, OK.

Paoletti, T., & Moore, K. C. (2017). The parametric nature of two students' covariational reasoning. *Journal of Mathematical Behavior*, *48*, 137–151. doi: 10.1016/j.jmathb.2017.08.003

Rebello, N. S., Cui, L., Bennett, A. G., Zollman, D. A., & Ozimek, D. J. (2007). Transfer of learning in problem solving in the context of mathematics and physics. In D. H. Jonassen (Ed.), *Learning to Solve Complex Scientific Problems* (pp. 223–246). Mahwah, NJ: Lawrence Erlbaum Associates.

Sherin, B. L. (2001). How Students Understand Physics Equations. *Cognition and Instruction*, *19*(4), 479–541. doi: 0.1207/S1532690XCI1904_3

Smith, T. I., Thompson, J., & Mountcastle, D. (2013). Student understanding of Taylor series expansions in statistical mechanics. *Physical Review Special Topics – Physics Education Research, 9*(2), 020110. doi: 10.1103/PhysRevSTPER.9.020110

Smith, T. I., Brahmia, S. W., Olsho, A., Boudreaux, A., Eaton, P., Kelly, P. J., Louis, K. J., Nussenbaum, M. A., & Remy, L. J. (2018). Developing a reasoning inventory for measuring physics quantitative literacy. Retrieved from https://arxiv.org/abs/1901.03351

Smith, T. I., Eaton, P., Brahmia, S. W., Olsho, A., Boudreaux, A., DePalma, C., LaSasso, V., Straguzzi, S., & Whitener, C. (in press). Using psychometric tools as a window into students' quantitative reasoning in introductory physics. In Y. Cao, S. Wolf, & M. Bennett (Eds.), *Physics Education Research Conference 2019*. Provo, UT. Retrieved from http://arxiv.org/abs/1907.05491

Smith, T. I., White Brahmia, S., Olsho, A., & Boudreaux, A. (2019). Developing a reasoning inventory for measuring physics quantitative literacy. In A. Weinberg, D. Moore-Russo, H. Soto, & M. Wawro (Eds.), *Proceedings of the 22nd annual conference on research in undergraduate mathematics education.* (pp. 1181–1182). Oklahoma City, OK.

Thompson, P. W. (2010). Quantitative reasoning and mathematical modeling. In L. L. Hatfield, S. Chamberlain & S. Belbase (Eds.) *New perspectives and directions for collaborative research in mathematics education* (Vol. 1, pp. 33- 57). Laramie, WY: University of Wyoming.

Vlassis, J. (2004). Making sense of the minus sign or becoming flexible in "negativity". *Learning and Instruction*, *14*(5), 469–484. doi: 10.1016/j.learninstruc.2004.06.012

Wells, J., Henderson, R., Stewart, J., Stewart, G., Yang, J., & Traxler, A. (2019). Exploring the structure of misconceptions in the Force Concept Inventory with modified module analysis. *Phys. Rev. Phys. Educ. Res. 15*(2), 020122. doi: 10.1103/PhysRevPhysEducRes.15.020122

White Brahmia, S., Olsho, A., Smith, T. I., & Boudreaux, A. (2018). NoNIP: Natures of negativity in introductory physics. In A. Traxler, Y. Cao, & S. Wolf (Eds.), *Physics Education Research Conference 2018.* Washington, DC. doi: 10.1119/perc.2018.pr.Brahmia

White Brahmia, S., Olsho, A., Smith, T. I., & Boudreaux, A. (2019). A framework for the natures of negativity in introductory physics. In A. Weinberg, D. Moore-Russo, H. Soto, & M. Wawro (Eds.), *Proceedings of the 22nd Annual Conference on Research in Undergraduate Mathematics Education.* (pp. 68–75). Oklahoma City, OK.

White Brahmia, S., Olsho, A., Smith, T. I., & Boudreaux, A. (in press). A framework for the natures of negativity in introductory physics. *Physical Review Physics Education Resesearch.*, Retrieved from https://arxiv.org/abs/1903.03806