Physics Quantitative Literacy: Assessment and Interaction with Student Characteristics

Benjamin B. Kauffman*
California Polytechnic State University, San Luis Obispo

Suzanne Brahmia
University of Washington
(Dated: September 7, 2019)

The Physics Inventory of Quantitative Literacy (PIQL) is a tool made to assess students' level of mathematical reasoning. Quantitative literacy is increasingly important in the modern world, and the assessment can help change the way we teach to serve students in improving it. In this research we look at data from the PIQL and use hierarchical linear modeling to answer questions about how student characteristics like grade in a physics course, SAT math score and gender interact with PIQL score. We also compare the method of hierarchical linear modeling to a traditional multiple linear regression, to determine whether we would obtain different results.

I. INTRODUCTION

The use of concept inventories in PER in the 1990s greatly improved physics instruction and curricula. Researchers showed that the traditional physics classroom was not serving students well in helping them learn the concepts of physics, and new ways of conducting a class were developed and tested using these instruments. However, in a physics classroom we want students to learn more than physical concepts; we also hope to improve their mathematical reasoning ability, and the Physics Inventory of Quantitative Literacy (PIQL) provides a tool to assist in achieving this goal. It is intended to be used in a similar way to concept inventories: to measure quantitative literacy and to test the effectiveness of interventions to improve quantitative literacy. The instrument is still under development, but a version has been administered to students in the introductory physics sequence: 121-mechanics, 122-electromagnetism, and 123-waves. It was given as a pre-test at the beginning of the quarter, so that the results from the second course in the series could act as a post-test for the first course. In this paper we look at this preliminary data and examine the interaction of student characteristics with their score. Traditionally, in this analysis we would use a linear regression with the characteristics as the independent variables and PIQL score as a dependent variable. With multiple characteristics this is called a multiple linear regression (MLR). However, a recent paper[1] by Van Dusen and Nissen proposed Hierarchical Linear Modeling (HLM) as a better alternative to a multiple linear regression in PER. We have adapted their analysis for this research, using this new modeling technique.

* Correspondence email address: bbkauffm@calpoly.edu

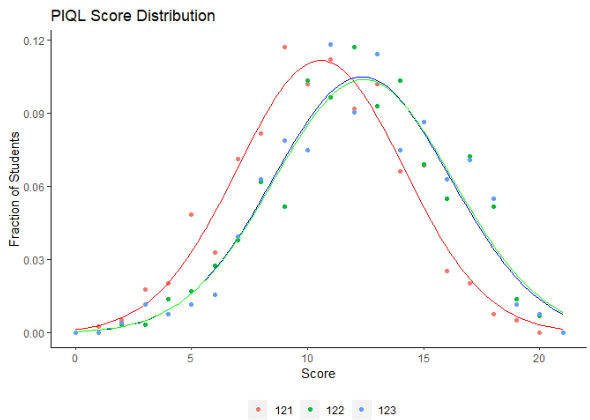

Figure 1. Score distributions for each course. There is some improvement in scores from the 121 pretest to the 122 pretest, however there is no gain at all from the 122 pretest to the 123 pretest.

II. BACKGROUND

A. Quantitative Literacy

Quantitative literacy can be defined as a set of interconnected skills, attitudes, and habits of mind that together support the sophisticated use of elementary mathematics to describe and understand the world. The key idea is that mathematics is not the difficult part, it is rather the interpretation and application of the mathematics. Physics quantitative literacy (PQL) is the use of quantitative literacy in the context of introductory physics. Quantitative literacy is an incredibly important skill in the modern world, one that people in most jobs should have, which is why one of the goals of this project is to increase the level of quantitative literacy in our society.

Table I. HLM Results. Progressive model development with variance and percent variance explained compared to model 1. Factors with high reduced variance are considered to correlate with PIQL score.

Model	Level	Equation	Variance	% Explained
1	1	$score_{ij} = \beta_{0j} + r_{ij}$	12.7	0.0
	2	$\beta_{0j} = \gamma_{00} + u_{0j}$	1.16	0
2	1	$score_{ij} = \beta_{0j} + \beta_{1j} * CourseGrade_{ij}$	9.48	25.8
	2	$\beta_{0j} = \gamma_{00} + u_{0j}$; $\beta_{1j} = \gamma_{01}$	1.07	7.5
3	1	$score_{ij} = \beta_{0j} + \beta_{1j} * Gender_{ij}$	12.6	1.6
	2	$\beta_{0j} = \gamma_{00} + u_{0j}$; $\beta_{1j} = \gamma_{01}$	1.03	11.1
4	1	$score_{ij} = \beta_{0j} + \beta_{1j} * SATmath_{ij}$	10.2	19.8
	2	$\beta_{0j} = \gamma_{00} + u_{0j}$; $\beta_{1j} = \gamma_{01}$	1.03	11.1
5	1	$score_{ij} = \beta_{0j} + \beta_{1j} * CourseGrade_{ij} + \beta_{2j} * SATmath_{ij}$	8.76	31.5
	2	$\beta_{0j} = \gamma_{00} + u_{0j}$; $\beta_{1j} = \gamma_{01}$; $\beta_{2j} = \gamma_{20}$	1.00	13.8

B. PIQL

The Physics Inventory of Quantitative Literacy (PIQL) has been under development by Suzanne Brahmia and her group at the University of Washington. The instrument uses three categories of quantitative literacy in its questions: proportional reasoning, covariational reasoning, and reasoning with sign, which I will define here.

a. Proportional reasoning. The use of ratios and products to describe systems and characterize phenomena.

b. Covariational reasoning. Holding in mind invariant relationships among quantities' values as those quantities vary in a dynamic situation. Covariational reasoning may be closely related to proportional reasoning.

c. Reasoning with sign. The use of sign to describe systems and characterize phenomena. This is a particular challenge to many physics students.

Score distributions for each introductory course are shown in figure 1. This shows a slight increase in scores after the first introductory physics course and no increase at all after the second. This could suggest that we are not helping students develop their mathematical reasoning abilities and could do much more in that area.

III. RESEARCH QUESTIONS

There are many interesting questions that can be explored with the data set we obtained. As one of the goals of the PIQL is to increase representation of underrepresented groups in STEM majors, we wanted to look at student characteristics and demographics. However, when we looked at the data we concluded that there were too few students in underrepresented ethnic groups to have significant findings. We could still look at gender and other information we have, and decided to focus on course grade and SAT math score. We are still developing our understanding of HLM, so we wanted to also look at the difference between results from both analyses. Based on these considerations, we came up with the following research questions:

- How does PIQL score interact with gender, course grade, and SAT math score?
- How do the HLM and MLR analyses differ?

IV. METHOD

A. Motivation

Physics education researchers often use linear regression to investigate phenomena. However, it has recently been pointed out by Van Dusen and Nissen[1] that this might not be the best method for data gathered in education research. The data usually rests in a hierarchical structure which will not be accounted for by the regression. The linear regression assumes independence, and that assumption is violated by the way the data was taken. For instance, there are many differences between sections of a course. They could start at different times, they have different instructors, and they could occur during a semester where a significant event happened and affected the student population. HLM does not need the assumption of independence, and controls for unexpected differences. In Van Dusen

Table II. MLR Results. Progressive model development with variance and percent variance explained compared to model 1. Factors with high reduced variance are considered to correlate with PIQL score. In this case we have no level 2, we are only looking at students.

Model	Equation	Variance	% Explained
1	$score = a$	13.9	0.0
2	$score = a + b * CourseGrade$	10.6	24.1
3	$score = a + b * Gender$	13.6	2.4
4	$score = a + b * SATmath$	11.3	19.1
5	$score = a + b * CourseGrade + c * SATmath$	9.76	29.9

and Nissen's article[1], they demonstrate how using the proper analysis can drastically change the findings of a study. They argue for HLM analysis, and show that a difference that appeared statistically significant with a multiple linear regression is not significant with hierarchical linear modeling. In comparing Force Concept Inventory scores of courses that used collaborative learning, collaborative learning with Learning Assistants or traditional lecture, there were outlier courses that used only collaborative learning, had a high number of students and had a much higher average score. We should attribute some of the increase in score to an unknown variable in those courses, not the variable we are looking at. The hierarchical linear modeling controlled for these differences, where the MLR did not.

B. Model

A linear regression would fit the data to an equation that might look like this:
$Outcome = a * factor_1 + b * factor_2 + c$
We would find the coefficients a, b and c that produced the lowest variance of the outcome residuals. However, this is assuming every student is independent and we know they are not. Students in the same class have the same instructor and have very similar experiences, which are different from students in other classes. These are variables that we are not accounting for and are not randomly varying in our population.
Our data was from 3 courses in 1 quarter, with approximately 900 students total. According to Van Dusen and Nissen, 10 is the minimum number of groups in a level 2 variable required to do the HLM analysis. The only way to have this many was to set up Tutorial section as the level 2 variable to control for. There are around twenty tutorial sections in each course. Students in a tutorial section are from the same lecture class, and they go to tutorial once a week to work on the tutorials with TA's. So our structure looks like this:

Level 1
Student
Level 2
Tutorial section
Our model is a two-level equation that looks like this:
$score_{ij} = \beta_{0j} + \beta_{1j} * factor_1 + beta_{2j} * factor_2$
$\beta_{0j} = \gamma_{00} + u_{0j}$; $\beta_{1j} = \gamma_{01}$; $\beta_{2j} = \gamma_{02}$
Here we have a score for student i in group j, which is given by an intercept that varies depending on the group ($\mathrm{u_{0j}}$) and constant coefficients multiplied by each factor. The fit finds the value of the coefficients β, γ and u that minimize the variance. We start with a base model that has only an intercept varying with the group. If we add a factor such as course grade and it reduces the variance of the fit, we have a correlation between PIQL score and that factor. We progressively added the factors and found the results in table 1.

V. RESULTS

As shown in table 1, course grade and SAT math reduced the variance significantly, indicating a correlation between each of them and PIQL score. The two together decreased the variance more than either individually, so they are not the same predictor. Gender had a very small correlation, only reducing the variance by 1.6 percent at level 1. While there was a 11.1 percent reduction of variance at the section level, we are not sure what that might mean. We hypothesize that it could be a result of highly varying gender distributions in each section, but further research could explore this result in more detail. We also compared the results of the two methods of analysis: hierarchical linear modeling and multiple linear regression. Table 2 shows the results of MLR. The results were very similar, and we would draw the same conclusions with either one. However, we can trust the HLM analysis more because we are not breaking the same assumptions and we are correctly controlling for the hierarchical structure of the data.

VI. CONCLUSION AND FURTHER RESEARCH

The results discussed indicate PQL is potentially a predictor of course grade, before the course starts. This could have huge implications in the way we can teach physics, namely it says that we should focus on developing students' quantitative literacy in the classroom. If PQL is an underlying skill that is a large barrier for many students, helping them develop it could drastically improve their performance in physics. SAT math has some correlation with the PIQL score, and we could explore this more. How strong is the correlation, and what might it mean? We might have expected a low correlation because we are expecting that the two tests assess different things. The most straight-forward further research would be to gather more data so that the HLM analysis is more appropriate. With enough courses across institutions, using the level two variable and allowing for variation across courses will be more necessary. This might show a more significant difference between HLM and MLR. We could also look at differences between ethnic backgrounds as well as gender. We did not have enough data from underrepresented groups in this analysis, but with more data we could explore the scores of these groups and how they could be increased.

ACKNOWLEDGEMENTS

I would like to thank Suzanne Brahmia for helping me enourmously in every step along the way. I would also like to give special thanks to Alexis Olsho for getting the student data together and anonymizing it so that I could work on it. Trevor Smith was instrumental in the statistical analysis and interpreting the results of HLM, so thank you to him. I would also like to thank the entire Physics Education Group at UW for welcoming me this summer and the INT REU program for this opportunity.

[1] B. Van Dusen and J. Nissen, *Modernizing use of regression models in physics education research: A review of hierarchical linear modeling* (Physical Review Physics Education, 2019).