Creating Statistically Equivalent Versions of a Test of Quantitative Literacy in Physics Contexts

Trevor I. Smith
Rowan University

Zachary Bischoff
Rowan University

Brett Boyle
Rowan University

Jack Sayers
Rowan University

Charlotte Zimmerman
University of Washington

Philip Eaton
Stockton University

Alexis Olsho
United States Air Force Academy

Suzanne White Brahmia
University of Washington

Keywords: quantitative literacy, assessment, physics

The Physics Inventory of Quantitative Literacy (PIQL) is a 20-item multiple-choice test designed to measure the development of students' physics quantitative literacy (PQL) across multiple physics courses (Olsho et al., 2023; Smith et al., 2020; White Brahmia et al., 2021). Repeated testing, coupled with requiring up to 40 minutes for students to complete the test, could lead to testing fatigue and unreliable results. We seek to create two shorter versions of the PIQL (a.k.a. piqlets) that are statistically equivalent to each other in terms of student performance on three facets of PQL (ratios and proportions, covariation, and signs and negativity).

Han et al. (2015) used a large data set of student responses to a 30-item conceptual physics test to identify combinations of items that produced testlets with the most similar average student scores. They demonstrated the equivalence of the testlets using item response theory (IRT) and correlating individual students' scores across versions. We follow their example by creating 12-item piqlets that each contain four overlapping anchor items (to facilitate comparisons between versions) and eight distinct items. Our work was guided by these research questions:

1) Which combination of items produces piqlets with the smallest score differences?
2) How does the reliability of these piqlets compare to each other and to the PIQL?
3) How similar are the psychometric parameters for the anchor items across piqlets?

Data were collected using the full PIQL in three introductory physics courses at a large public university in the western US (2100–3200 students in each data set). We considered 240 combinations of items for the piqlet versions, subject to constraints that both the content and the format of items were equivalent across the two versions. We calculated a total test score for each piqlet based on the average percentage of items answered correctly, as well as subscores for the three facets of PQL. For each of our three data sets we determined the average score difference between the piqlets, calculated Cronbach's α for each piqlet, and applied IRT analyses to each.

The combinations of items that we identified as being the most similar had overall average score differences from 0.6–1.3%. Cronbach's α values ranged from 0.67 to 0.75, with differences less than 0.01 between versions. We see strong correlations between individual student scores on the two piqlets, with $0.79 \leq r \leq 0.85$. The IRT parameters support the statistical equivalence of the piqlets with parameters of overlapping items agreeing to within 0.1. These preliminary results suggest a strong potential for identifying piqlets that are statistically equivalent for the broader population of mathematics and physics students based on a larger, more diverse, data set.

Acknowledgments

This work is supported by the NSF under Award Nos. DUE-1832470, DUE-2214283, and DUE-2214765; and an NRC Research Associateship award at Air Force Research Laboratory.

References

Han, J., Bao, L., Chen, L., Cai, T., Pi, Y., Zhou, S., Tu, Y., & Koenig, K. (2015). Dividing the Force Concept Inventory into two equivalent half-length tests. *Physical Review Special Topics – Physics Education Research, 11*(1), 010112. https://doi.org/10.1103/PhysRevSTPER.11.010112

Olsho, A., Smith, T. I., Eaton, P., Zimmerman, C., Boudreaux, A., & White Brahmia, S. (2023). Online test administration results in students selecting more responses to multiple-choice-multiple-response items. *Physical Review Physics Education Research*, *19*(1), 013101. https://doi.org/10.1103/PhysRevPhysEducRes.19.013101

Smith, T. I., White Brahmia, S., Olsho, A., & Boudreaux, A. (2020). Developing a reasoning inventory for measuring physics quantitative literacy. In S. Smith Karunakaran, Z. Reed, & A. Higgins (Eds.), *Proceedings of the 23rd Annual Conference on Research in Undergraduate Mathematics Education.* (pp. 949–955). http://sigmaa.maa.org/rume/RUME23.pdf

White Brahmia, S., Olsho, A., Smith, T. I., Boudreaux, A., Eaton, P., & Zimmerman, C. (2021). Physics Inventory of Quantitative Literacy: A tool for assessing mathematical reasoning in introductory physics. *Physical Review Physics Education Research*, *17*(2), 020129. https://doi.org/10.1103/PhysRevPhysEducRes.17.020129