Wright State University
CORE Scholar

Scholarship in Medicine - All Papers Scholarship in Medicine

2020

Implementing Peer Instruction in Medical Education and the Impact on Student Success

Alexander Chase
Wright State University - Main Campus, chase.21@wright.edu

Follow this and additional works at: https://corescholar.libraries.wright.edu/scholarship_medicine_all

Part of the Medical Education Commons

Repository Citation
Chase, A. (2020). Implementing Peer Instruction in Medical Education and the Impact on Student Success. Wright State University. Dayton, Ohio.

This Article is brought to you for free and open access by the Scholarship in Medicine at CORE Scholar. It has been accepted for inclusion in Scholarship in Medicine - All Papers by an authorized administrator of CORE Scholar. For more information, please contact library-corescholar@wright.edu.

Implementing Peer Instruction in Medical Education and the Impact on Student Success

Scholarly Project Final Report

Alexander Chase

Dr. Dean Parmelee, WSU BSOM Office of Medical Education

Literature Review:

In 1991, a physicist at Harvard University, Dr. Eric Mazur, implemented a new style of classroom learning into his introductory physics courses named Peer Instruction.[1] The objective of Peer Instruction (PI) was to engage students in interactive discussion throughout the time spent in the classroom. In the PI format, students were assigned reading that was then tested at the start of each class through a reading quiz. Dr. Mazur then spent the remainder of class quizzing his students with questions that were designed to highlight the important aspects of the material in question. All students were required to participate by first recording individual responses. After their initial response, discussion regarding the answer was allowed between students to force them to again think critically to defend their answer.

To evaluate the effectiveness of Peer Instruction, Dr. Mazur administered multiple standardized exams in his courses. A test designed to evaluate conceptual knowledge showed an improvement in scoring between the beginning of the class and 2 months after class began when using Peer Instruction. Furthermore, the improvement in scores on this test was greater than the improvement achieved on the same exam using conventional teaching methods, e.g. lecture. A second standardized exam which involved a mix of conceptual and computational questions again showed improvement between the average scores achieved by students instructed with Peer Instruction compared to those in conventional classes. This test, along with results from identical final exams given by Dr. Mazur in 1985 when conventional teaching had been used, and 1991 when he employed Peer Instruction, showed that Peer Instruction not only improved conceptual knowledge related to physics, but that students were better equipped to solve specific problems as well.[1] In a ten year review of Peer Instruction in the introductory physics classes at Harvard, student performance as measured by standardized assessments was not only maintained but improved over that time period. This improvement stemmed from modifications made to increase student pre-class preparation and in-class engagement.[2]

Additional research has recognized the successful integration of Peer Instruction across the department of physics at Harvard University, as well as the spread of the instructional style to other universities such as Appalachian State University in North Carolina. This research notes that the ease of transition to Peer Instruction from conventional teaching methods is seemingly easy, but that intrinsic resistance to change is the largest obstacle.[3] Additionally, a 700-user survey of instructors indicated successful implementation of Peer Instruction into their curriculum. This was a broad questionnaire which included instructors from a variety of educational institutions from undergraduate universities to high schools, and who taught a variety of science disciplines including physics, chemistry, life sciences, engineering, astronomy, mathematics, and other disciplines.[4]

It has been shown that the most effective way to build knowledge and develop critical thinking skills in undergraduate education is through active engagement.[5] Knowing that such active engagement is crucial to learning, the Boonshoft School of Medicine (BSOM) at Wright State University implemented a version of PI in a significant portion of classroom time for a new curriculum designed to use only 'engaged' or 'active' learning in the classroom. This large-scale implementation began in the Fall of 2017 with the group of students who comprise the Class of 2021. In addition to BSOM, other medical schools have also begun a transition to active learning modalities within their curriculum. However, little literature exists demonstrating the impact of PI within medical education. The goal of this project is to show how medical student performance in PI relates to the desired outcome of student learning. Similar to the studies by Mazur, the student learning outcome will be assessed using the scores of a final summative examination. Student performance in PI classroom sessions will be quantified using the first polling data (individual, formative) from those sessions. By comparing these values, it will be shown whether PI performance can be used to accurately predict final summative examination scores. This will build on the work completed by Mazur by looking at the effect of PI on the summative learning of an individual student, while also expanding the knowledge of applying PI in medical education.

Research Question:

Is individual performance on Peer Instruction questions 1st poll predictive of performance on summative evaluations?

The aim of this project will be to demonstrate how the performance of students on the 1st poll, or individual, portion of PI is predictive of their scores on summative assessments. Although it would seem obvious that individual performance on a set of PI sessions would predict summative assessment outcomes, we do not know the predictive value set points or the relative strength of this prediction. Determining how Peer Instruction 1st poll scores predict final performance may benefit students because they will know early-on in a course which direction their daily performance is headed, and if in a negative direction, then they can seek help sooner rather than later. Furthermore, the value of Peer Instruction in the biomedical science domain of medical education has not been explored. Since much of this domain is less concept and formula grounded than math and physics, is it a value-added learning strategy? Does having to learn a bolus of vocabulary and terminology in preparation to solve clinical problems detract from its potential effectiveness? This investigation will begin to address this larger question by just one component of PI data as it relates to medical student summative performance in a biomedical science course.

Methodology:

Student performance data from the Fall 2017 Origins I & II modules was de-identified and coded with a generic number for tracking by the BSOM Office of Medical Education (OME) for the proposed analyses. Origins I & II are the introductory molecular and cell biology courses taught at BSOM. Specifically, the scores from the 1^{st} poll (individual) portion of PI was utilized, along with the individual components of Multiple-Choice Questionnaire (MCQ) exams and the NBME Customized final exam. Only one course module was utilized in order to gain an understanding of the relevancy and structure of this data analysis.

The component of PI data that was utilized is also commonly called 1^{st} poll data as it occurs based on the knowledge a student came to class with based on the assigned readings. The 2^{nd} poll data represents re-entry of answers to the same PI question as the 1^{st} poll after peer discussion occurred. Only 1^{st} poll individual scores for each PI session were utilized in this study.

In the context of this study, the PI data is distinctly formative and the NBME final exam is the summative data marker, but the MCQ data could potentially have been used as both formative and summative data depending on whether it was referenced to the PI or NBME scores. This study used MCQ scores as summative data points compared to PI data since the goal was to understand the predictive value of PI performance.

These data were compared using descriptive statistics. A linear regression was applied to the data to determine if predictions for MCQ and NBME exam scores could be made based on the 1^{st} poll performance of PI. PI labeling occurred in the format of O1-1 which indicates "Origins 1-PI 1." Likewise, O2-4 indicates "Origins 2-PI 4." The linear regression models were built by adding PI scores to the model sequentially. The purpose of this method was to determine the earliest point at which PI scores can accurately predict NBME final score. The linear regression models were named after the latest PI added to the model. Thus, model O1-1 is a linear regression based solely on the first PI of Origins 1. Model O2-4 is a linear regression model based on all the Origins 1 PI scores, and the first four PI scores of Origins 2. The final model is an all-inclusive, or comprehensive, linear regression model incorporating all PI scores. The final model is synonymous with model O2-13.

Exclusion/Inclusion criteria: The 1^{st} poll PI scores, MCQ individual scores, and NBME final scores for all 115 members of the BSOM class of 2021 were utilized. Of note, there were instances in every PI session where the correct answer was not achieved by a majority of the class even after peer-peer discussion and the 2^{nd} poll. In these instances, the question was given as credit and marked as correct for all the individuals within that class. These questions were included within this analysis as correct answers on the 1^{st} poll data. While this is an inflation of 1^{st} poll PI scores, the intent of the project is to provide predictive value based on available scores. Additionally, there were rare instances of missing PI data due to absences throughout the module with 15 missing PI score data points total. All data from the individuals with missing scores was excluded beginning with the grouped analysis of PI scores from which the data was missing.

Results:

In the regression analyses of the models created by chronologically adding PI scores in succession and the comparison to the corresponding individual NBME final score, the model R-square value stopped consistently changing with significance following the model which added PI O1-7 (Table 1). This was determined per the F change significance (p-value) which was equal to 0.047. The R-square value at this point, utilizing the scores from the first 7 PIs, was 0.617. Following the model adding PI O1-7, there were three additional significant changes to the R-square value with the models adding PI O1-9, O1-11, and O2-8. These were interspersed with PI scores that did not significantly add to the model and were thus deemed to be inefficient additions to the model.

In analyzing the final model in this succession of analyses, when all PI scores from Origins 1 and 2 were used in generating the model, the predictive R-square value was equal to 0.75. Two PIs were found to contribute significantly to this final model (p-value < 0.05): O1-5 and O1-9. Both of these PIs were found to add significantly to the R-square value when they were added to the predictive model in succession, respectively. The additional PIs which added significantly to the R-square value in the consecutive succession analyses, O1-11 and O2-8, had p-values of 0.070 and 0.078 as part of the all-inclusive final model. In the model utilizing all PI scores, PI O1-7 missed the cut-off for significance as its p-value was 0.069 (Table 2).

Of note, N decreased at three distinct intervals during the consecutive analysis. This was due to limitations in statistical software which allowed for only nine PI score cohorts to be analyzed at one time, combined with the default of the program to drop any student from the analysis if that person were missing scores for any of the PIs (i.e. an absence). Thus, for the models associated with PIs) O1-1 through O1-9, N equals 113. For models O1-10 through O2-4, N equals 111. For models O2-5 through O2-13, which therefore includes the final, all-inclusive model, N equals 104. The discreet decreases in N occurred beyond the point where the model was determined to stop changing significantly. This decrease in N explains the apparent decrease in consecutive R-square value which occurs between PI O2-4 and O2-5 (Figure 1).

The final model which utilized all PI scores found that PI O1-9 contributed most significantly to the final NBME score with p-value equal to 0.002. Linear regression analysis of PI O1-9 to NBME final score found an R-square value equal to 0.34 (Table 3). Additional models were created in order to compare Origins 1 PI scores to NBME scores as well as Origins 2 PI scores to the same final. It was shown that Origins 1 scores had a R-square value of 0.701 with PIs 5, 9 and 11 contributing significantly (Tables 4 & 5). Origins 2 scores had a R-square value of 0.659 with PIs 4, 10 and 11 contributing significantly (Tables 6 &7).

The ability of PI scores to predict the related MCQ score was varied with final R-square values ranging from 0.336 to as high as 0.547 (Figure 2). There were differing numbers of PIs for the six MCQs administered throughout Origins 1 and 2.

Conclusions:

Individual performance (1st poll) on Peer Instruction questions is an important indicator of performance on summative evaluations. The analysis of sequential PI scores found that with each additional PI score that was added to the model, the ability to more accurately predict the individual's final NBME score was improved. This ability of the model to predict final score ranged from accounting for approximately 18% of an individual's final score after the first PI, to 75% of an individual's final score by PI O2-13, the twenty-sixth and last PI of the module. Furthermore, the increase in ability to predict final score was improved significantly with each additional PI score beginning with the first PI, O1-1, until the seventh PI session, O1-7. The R-square value of the model including all PIs through O1-7 was 0.671. This indicates that after just seven out of twenty-six PIs (27% of all PIs), the overall predictive value of PI scores accounts for 67% of the final NBME grade with a standard error of 5%. This prediction of 67% of an individual's final score represents nearly 90% (67%/75%) of the total predictive value derived from PI scores. After PI O1-7, the change in predictive value on final NBME scores no longer changes significantly with each consecutive additional PI added to the model. Interestingly, certain PI sessions were found to add more significantly to the predictive ability of the model than others. PI O1-9, O1-11, and O2-8 all added significantly to the model when they were included in turn. However, these were interspersed with PI scores that did not significantly add to the model and were thus deemed to be inefficient additions. Overall, when utilizing the final model which included all PI scores, only PIs O1-5 and O1-9 significantly added to the ability of the model to predict final NBME score. This indicates that out of twenty-six PI scores which predict three quarters of an individual's final score, only two (7.7%) are significantly adding to the model. This seems to indicate the obvious, that collectively the ability of individual PI scores to predict final score is significant (75%) but that most PIs do not significantly add to that predictive value alone. Analysis of PI scores collectively from Origins 1 versus 2 did not find any notable differences between the two in predictive impact on final NBME score, accounting for 70% and 66% of final grade, respectively. Nor was a significant trend found among PI predictive value on subsequent MCQ scores. A general positive trend on R-square value was found with addition of PIs to predictive models ranging from 34% to 55%, but there weren't enough PIs per corresponding MCQ to establish a distinct pattern.

This research establishes that individual PI scores are significantly predictive of final NBME score. It begins to define some of the initial data points that further studies can expound on. One limitation of this study is its size with N for the cumulative model equaling 104. The original cohort for this research was 115, but due to absent data ultimately 11 individuals were excluded. This represents a nearly 10% dropout. Further evaluation of an expanded cohort of data now that two additional classes will have gone through the Origins modules would be beneficial. Additionally, this study uncovers some interesting trends regarding individual PI influence on NBME final score. The cumulative model indicated that while most PIs did not contribute significantly to the model as would be expected, two PIs did: PI O1-5 and O1-9. O1-9 contributed most significantly and further analysis utilizing that PI alone resulted in an R-square value of 0.340 with a standard error of 6.5%. This means that approximately over a third of the NBME final score is predicted by individual performance on a single PI. Upon investigation of

Origins 1 and 2 grouped PI scores, it was revealed that in each respective module, there are three out of the thirteen PI scores which significantly contribute to the predictive ability of the model. Further investigation could seek to define why there are these disparities in weight of PI, and seek to understand if they represent truly significant differences or merely coincidence.

References:

1. Mazur E. *Peer instruction: A user's manual.* 1st ed. Upper Saddle River, NJ: Prentice Hall; 1997.

2. Crouch CH, Mazur E. Peer instruction: Ten years of experience and results. *American journal of physics*. 2001;69(9):970-977.

3. Tobias S. *Revitalizing undergraduate science: Why some things work and most don't. an occasional paper on neglected problems in science education.* ERIC; 1992

4. Fagen AP, Crouch CH, Yang T, Mazur E. Factors that make peer instruction work: A 700-user survey. . 2000.

5. Johnson DW, Johnson RT, Smith KA. *Active learning: Cooperation in the college classroom.* ERIC; 1998.

Index:

Table 1: R-squared Values for Consecutively Constructed PI Predictive Models

Model Number (Latest PI Added)	R Square	Std. Error of the Estimate	Sig. F Change
O1-1	18%	7%	0.000
O1-2	35%	6%	0.000
O1-3	45%	6%	0.000
O1-4	51%	6%	0.001
O1-5	59%	5%	0.000
O1-6	60%	5%	0.035
O1-7	62%	5%	0.047
O1-8	63%	5%	0.101
O1-9	67%	5%	0.000
O1-10	68%	5%	0.408
O1-11	69%	5%	0.029
O1-12	70%	4%	0.075
O1-13	70%	4%	0.561
O2-1	70%	5%	0.582
O2-2	71%	5%	0.328
O2-3	71%	4%	0.109
O2-4	72%	4%	0.101
O2-5	71%	4%	0.206
O2-6	71%	5%	0.704
O2-7	71%	5%	0.975
O2-8	73%	4%	0.017
O2-9	73%	4%	0.199
O2-10	74%	4%	0.068
O2-11	75%	4%	0.151
O2-12	75%	4%	0.657
O2-13	75%	4%	0.672

Table 2: Relative Contributions from Individual PI Scores to Final All-Inclusive Model

PI	Unstandardized Coefficients	95.0% Confidence Interval for B		Sig.
		Lower Bound	Upper Bound	
O1-1	-0.022	-0.158	0.114	0.747
O1-2	-0.021	-0.127	0.084	0.685
O1-3	0.035	-0.090	0.159	0.584
O1-4	0.012	-0.095	0.119	0.822
O1-5	0.101	0.006	0.197	0.038
O1-6	0.007	-0.091	0.105	0.885
O1-7	0.096	-0.008	0.200	0.069
O1-8	-0.035	-0.140	0.069	0.499
O1-9	0.151	0.056	0.245	0.002
O1-10	-0.029	-0.121	0.062	0.527
O1-11	0.087	-0.007	0.182	0.070
O1-12	0.032	-0.067	0.132	0.518
O1-13	0.026	-0.079	0.132	0.621
O2-1	-0.012	-0.111	0.088	0.812
O2-2	0.024	-0.072	0.120	0.614
O2-3	0.039	-0.051	0.129	0.393
O2-4	0.037	-0.050	0.124	0.397
O2-5	0.068	-0.055	0.191	0.274
O2-6	0.008	-0.081	0.097	0.853
O2-7	-0.035	-0.128	0.058	0.456
O2-8	0.066	-0.008	0.140	0.078
O2-9	0.029	-0.060	0.119	0.517
O2-10	0.080	-0.005	0.165	0.066
O2-11	0.063	-0.034	0.160	0.201
O2-12	0.017	-0.089	0.123	0.755
O2-13	0.019	-0.068	0.105	0.672

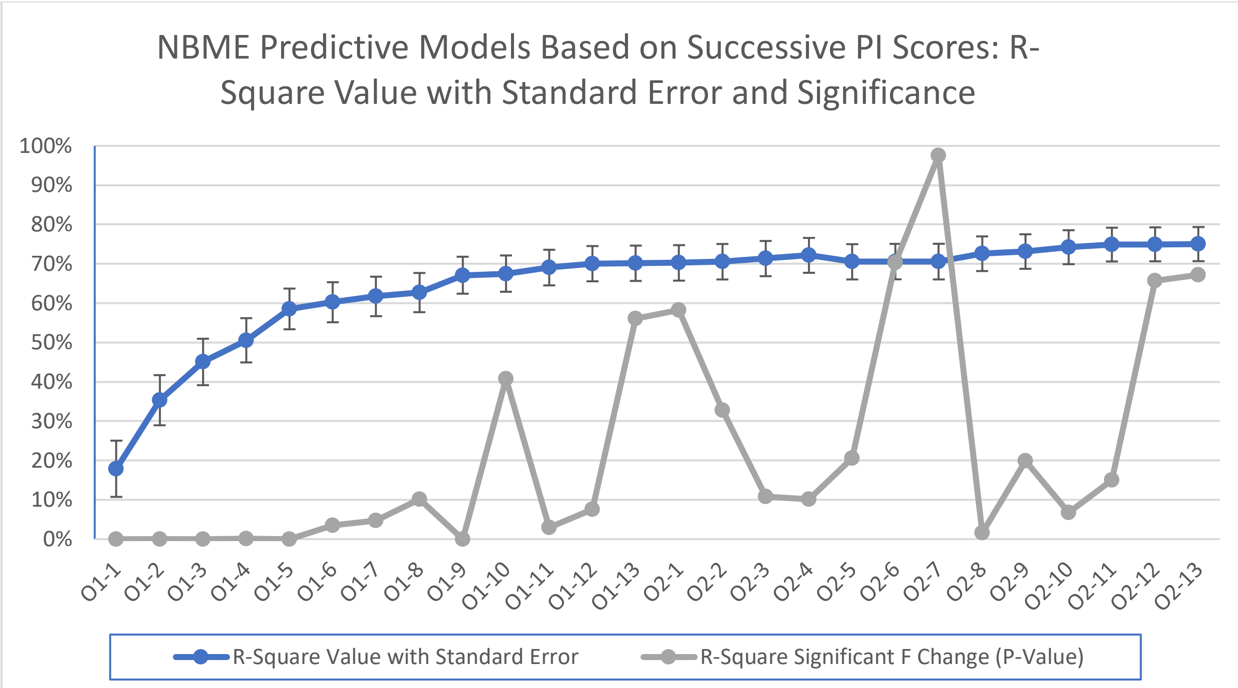

Figure 1: R-squared Values for Consecutive PI Models with Standard Error and Significance

Table 3: PI O1-9

Model	R Square	Std. Error of the Estimate	Sig. F Change
O1-9	0.340	6.46871%	0.000

Table 4: Model Summary for Origins 1

Model	R Square	Std. Error of the Estimate	Sig. F Change
Origins 1	0.701	4.47767%	0.000

Table 5: Individual PI Relevance to Origins 1 Model

Model	Unstandardized Coefficients	95.0% Confidence Interval for B		Sig.
		Lower Bound	Upper Bound	
O1-1	0.032	-0.091	0.155	0.609
O1-2	0.033	-0.053	0.120	0.448
O1-3	0.057	-0.057	0.172	0.323
O1-4	0.067	-0.028	0.163	0.166
O1-5	0.152	0.064	0.241	0.001
O1-6	0.041	-0.051	0.133	0.380
O1-7	0.066	-0.032	0.163	0.184
O1-8	1.391E-06	-0.093	0.093	1.000
O1-9	0.140	0.056	0.225	0.001
O1-10	0.000	-0.080	0.080	0.998
O1-11	0.092	0.010	0.174	0.028
O1-12	0.073	-0.010	0.157	0.083
O1-13	0.027	-0.068	0.122	0.569

Table 6: Model Summary for Origins 2

Model	R Square	Std. Error of the Estimate	Sig. F Change
Origins 2	0.659	4.86521%	0.000

Table 7: Individual PI Relevance to Origins 2 Model

Model	Unstandardized Coefficients	95.0% Confidence Interval for B		Sig.
		Lower Bound	Upper Bound	
O2-1	0.055	-0.040	0.150	0.251
O2-2	0.062	-0.031	0.155	0.187
O2-3	0.033	-0.059	0.126	0.477
O2-4	0.116	0.032	0.200	0.007
O2-5	0.058	-0.063	0.179	0.342
O2-6	0.066	-0.026	0.159	0.158
O2-7	-0.031	-0.127	0.065	0.522
O2-8	0.073	-0.001	0.148	0.054
O2-9	0.072	-0.017	0.162	0.112
O2-10	0.096	0.003	0.188	0.042
O2-11	0.099	0.003	0.195	0.044
O2-12	0.037	-0.067	0.140	0.480
O2-13	0.039	-0.047	0.126	0.372

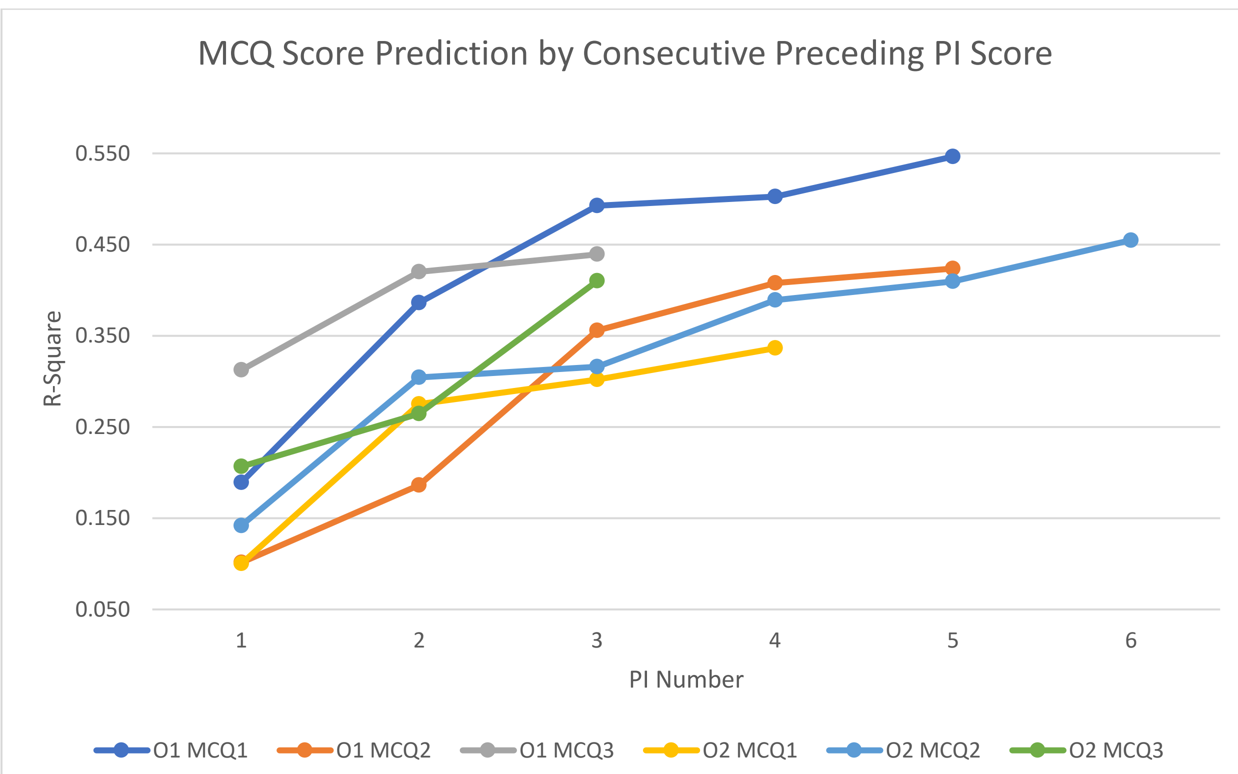

Figure 2: R-Square Values for Individual PIs Relative to Corresponding MCQ