Assessment of General Physical Science Education
Spring 2023

Phillip Vargas

Unit Level Liaison for Dept. of Physical Science

For the past 3 years the Department of Physical Science has been administering the Lawson Classroom Test of Scientific Reasoning (CTSR) assessment to every introductory physical science class. While there may be potential benefits from assessing a different learning outcome or modifying the approach of how we are assessing scientific reasoning, the department still considers administering a consistent assessment the best course of action. The pandemic has brought with it significant disruptions, and it would be difficult to state these disruptions are not impacting the department's learning assessment results. Introducing additional variables would only increase the dimensionality of the data and hinder out ability to interpret results. Consistent assessments should eventually revert to the mean values and provide baselines that are more statistically reliable as the disruptions abate.

As we transition out from the pandemic and as we hopefully return to a more consistent instructional environment, having this data will be invaluable. There are many confounding variables already present in attempting to assess student learning and making methodology changes too frequently can exacerbate those. Having a longitudinal study will provide the department with control data for our future experiments (both intentional and natural). This will allow the department to better evaluate changes in the future while addressing challenges the results currently present.

The Assessment Tool

The most prevalent tool utilized for general education physical science courses in higher education is the Lawson Classroom Test of Scientific Reasoning (CTSR) developed by Anton Lawson in 1978 and subsequently revised until 2000. While prevalence does not necessary correspond with efficacy, it has also been extensively validated and administered to multiple institutions. This data is warehoused in the PhysPort website and is constantly being updated by participating faculty members and institutions. Administering this tool allows the department to compare our assessment results with similar courses at other universities, in addition to

benchmarking them against our program-level outcomes. Capturing this data provides an excellent starting point to building a strong general education assessment foundation.

This tool measures scientific reasoning across six domains 1) conservation of matter and volume, 2) proportional thinking, 3) probabilistic thinking, 4) correlational thinking, 5) control of variables, and 6) hypothetical-deductive reasoning. These skills are essential components to science courses and are typically included when defining scientific reasoning. These sills also align with the department's program-level learning outcomes: 3) Analyze and interpret data using mathematics and computational thinking and 4) Construct explanations and engage in arguments from evidence.

Deployment System

Student participation in learning assessments have been a significant challenge. This challenge only increased during the pandemic. As many of the department's general education courses were already available online, this provided the department with an opportunity to implement and evaluate a variety of processes for online assessment. These efforts could continue to bear fruit after the pandemic as they would be inculcated in the department's culture and could then be deployed in course sections with synchronous and asynchronous modalities.

In fall 2020 the Learning About STEM Student Outcomes (LASSO) platform was piloted as a potential candidate for a large-scale assessment data acquisition tool. This system was specifically designed for this type of deployment in administering, analyzing, and reporting assessment outcomes. Unfortunately, the pilot was unsuccessful. At most, 1 student completed the assessment in each of the piloted sections. This platform is part of The University of Colorado's Learning Assistant Alliance which had questions that were not pertinent to our student body and the assessment was being ignored due to coming from an external institution.

In fall 2021 a similar platform was developed within the Office 365 application suite. Utilizing Forms, Excel, Outlook, and Power Automate an instrument could be created, emailed to all the students, and results stored in central location. Additionally, the form and email message could be customized for our students and be sent from the department liaison who could respond to questions or concerns. This system was developed and piloted in Fall 2021 and deployed to all students since. While the technology of this platform was sound, it was clear the participation from sections was not uniform. It is suspected that participation was significantly influenced by

select instructors promoting the assessment or providing extra credit for completing it. While this affect has been seen and confirmed many times in college-wide assessment in requires reiterating its importance and its scope encompasses individual departments as well.

In Spring 2023, with several sections being offered in-person a hybrid administration model was deployed where participating faculty members administered a paper copy to their courses, and the remaining sections had the learning assessment administered though Microsoft Forms. Three of the six faculty members teaching in-person classes responded and agreed to administer a paper version of the learning assessment in their class. While this model added a layer of complexity in terms of coordination and merging data sets, the benefits in student participation outweighed the effort. As there was no opposition to this request, only a nonresponses, there is encouragement that faculty participation will increase in upcoming semesters, and additional effort will be made to communicate and streamline the process.

Results

Participation Results

In fall 2022 there were 25 sections of Physical Science classes with 502 students initially enrolled. At the time of the assessment 396 students were enrolled across these sections. In spring 2023 there were 24 sections of Physical Science classes with 500 students initially enrolled. At the time of this assessment 396 students were enrolled across these sections. While the decline in general education students continued this year for the fifth year in a role there was a significant increase in respondents. This true for both the fall and spring semesters where 69 (17.4%) and 103 (26.0%) students completed the survey compared to the previous spring where only 4.2% completed the survey. The improvement in number of respondents from previous semesters can only in part be contributed to the hybrid model as it was only administered in spring 23. This could at account for at most a 9.4% increase in participation. It is possible that having more faculty on campus and having face-to-face conversations about value of the assessment is improving faculty participation as well as reminders to promote the assessment may have contributed to the increase in the online deployment. Additionally, as the severity of the pandemic diminishes, faculty members and students simply may have more time to pursue and complete ancillary tasks.

This increased participation improved the reliability of the data, giving the fall 2022 results a margin of error of 11% at a confidence level 95%, and the spring 2023 assessment results of and 8% margin of error at a confidence level 95%. While a 5% margin of error is the *de facto* gold standard and the goal of the assessment committee, this is challenging goal to meet with a smaller population. However, the department had been making strides toward it. Again, when reviewing student participation, not performance, there were clusters from specific sections. The department is confident that as the processes get more routine a culture will build around them and participation will continue to increase.

Performance Results

Similar to previous assessments there is not sufficient data or knowledge of confounding variables to perform comparative analysis to external institutions or internal institutional dimensions (course, modality, etc. However, this assessment does allow us to compare relative performance across the six scientific reasoning domains. Additionally, as this was the same survey that was administered in the previous year, the department can start to from a historical timeline, and with the inclusion of the fall semester an additional timepoint can be included going forward.

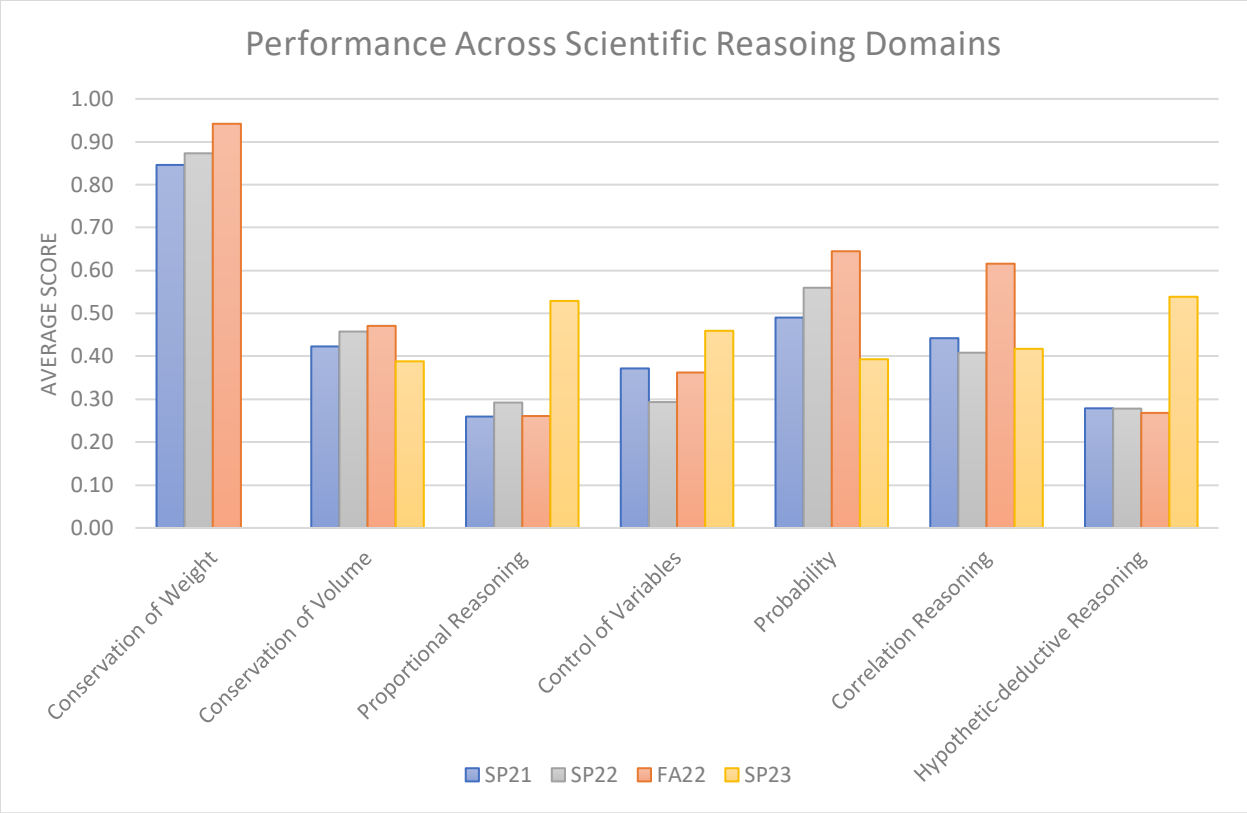
The performance results from the 2033-2023 were the most dissimilar in the last three years. The first discrepancy occurred with an error to the first two questions in the online administration. When reviewing this error, it looks as though it was part of a question bank was dragged and dropped into another question when checking response tallies. Unfortunately, this error was not noticed until many students already completed the assessment. Instead of trying to piece together correct responses, this domain was simply removed for the spring 2023 results. Protocols will be changed to mitigate these errors in the future.

The second notable feature of the analysis was the significant differences in results between the fall and spring semesters for the proportional reasoning (fall ’22: 0.26, spring ’23: 0.53), probability (fall ’22: 0.64, spring ’23: 0.39), correlation reasoning (fall ’22: 0.62, spring ’23: 0.42), and hypothetic-deductive reasoning (fall ’22: 0.27, spring ’23: 0.54). Since performances are deviating in both directions, this reduces the potential for selection bias from the change in administration. It raises the potential possibility that the fall and spring populations may be substantially different. However, there is not enough sample points to conclude this. It does

reinforce the need to administer the assessment in both terms of the academic year, and not assume the spring term results can be generalized.

Despite the variations within the domains of learning, the average scores remain consistent across these four assessments with a range of 0.41 to 0.43. While the latest assessment is at the current maximum, this shift is well within the margin of error and should not be considered a valid increase. It does, however, present the department with ample opportunity to explore initiatives that explicitly improve scientific reasoning skills in additional to scientific content and processes.

These results continue to provide further evidence that appears contrary to anecdotal evidence in the department. It is often discussed that student's scientific reasoning is higher than their mathematical background. However, in these results from all the assessments, students appear to be performing relatively well in conservation and probabilistic thinking, while their understanding of how to control for variables and hypothetic-deductive reasoning is significant lower. This type of reasoning is key in experimental design and to a large part of the scientific method. As mentioned in the previous reports there may be significant difference between these performances in lab and non-lab classes, this point may need some reflection in the department and possibly retooling of laboratory practices. However, the shift to and back in remote teaching may be affecting these some domains. Some possible solutions could be requiring students to perform more experimental setup, develop more open-ended problems, or even potentially even having students design their own experiments.



As this instrument has been validated in several education settings, it was not necessary to analyze its efficacy. However, an analysis of independence of the performance across these domains are conducted with a correlation matrix analysis. This analysis shows any dependencies within these domains. This could illustrate that particular classroom assignments have a larger effect on a subset of these domains. This analysis is consistent across these domains from previous years.

	Conservation of Weight	Conservation of Volume	Proportional Reasoning	Control of Variables	Probability	Correlation Reasoning	Hypothetic-deductive Reasoning
Conservation of Weight	1.00	0.22	-0.12	0.36	0.54	0.18	0.31
Conservation of Volume	0.22	1.00	0.24	0.21	0.32	0.33	0.26
Proportional Reasoning	-0.12	0.24	1.00	0.28	0.08	0.24	-0.14
Control of Variables	0.36	0.21	0.28	1.00	0.40	0.15	0.17
Probability	0.54	0.32	0.08	0.40	1.00	0.52	0.06
Correlation Reasoning	0.18	0.33	0.24	0.15	0.52	1.00	0.21
Hypothetic-deductive Reasoning	0.31	0.26	-0.14	0.17	0.06	0.21	1.00

Additionally, this new process allows disaggregation of the data by course, session, modality, or demographics. While this type of analysis is planned for future reports, additional safeguards need to be addressed first to ensure confidentiality and anonymity.

Conclusion

In spring 2021 the first large scale administration of a single assessment instrument across all of the general education courses offered in the Department of Physical Science was conducted. The acquired results met the 10% margin of error at a 95% confidence level and showed the strengths and weaknesses of our students learning across six domains of scientific reasoning. This assessment was repeated in spring 2022 with a significant decrease in participation, but similar results. In the 2022-2023 academic year this assessment was administered in both the fall and spring semesters. While students initially appear to be performing relatively well in conservation and probabilistic thinking. Their understanding of how to control for variables and hypothetic-deductive reasoning was significantly higher this year than in previous years. While this may be affected in part by having in-person labs or several discussions around previous assessment results, a causal relationship has not been established and further data is needed to confirm these results are stable. The mean performance of 0.42 has remained stable across semesters and years for the past three years despite significant disruptions.

Based on the variations across the learning domains in the fall and spring terms the department will continue to administer the assessment in both semesters using electronic and in-person methodologies. This additional data will help to determine if term populations, assessment administration or other unknown variables contributed to the variability in learning domains in the 2022-2023 academic year. Both administration methodologies will continue to be streamlined for faculty members to encourage them to volunteer instructional time or promote these assessments to their students in the future. This should continue to improve participation rates to and ultimate reach the 5% margin of error at a 95% confidence level.