A New Approach to Learning Improvement: Starting with an Intervention

Authors:

Laura A. Lambert
James Madison University

Megan R. Good
James Madison University

ABSTRACT

In higher education, institutions routinely assess student learning in degree programs driven by institutional accreditation requirements. Assessment is intended to provide an avenue for improvement with collected data inspiring curricular change and future assessments providing evidence of the efficacy of those changes (i.e., learning improvement). However, evidence of actual learning improvement resulting from assessment is rare, partly because learning improvement projects are resource-intensive, requiring significant faculty time and potentially departmental funding. In this article, we explore a novel approach to learning improvement by using existing data in a STEM undergraduate program. We identified 11 student learning objectives aligned with a previously implemented curricular change. Trained faculty used a common rubric to evaluate student work submitted both before and after the curricular change was implemented. We found evidence of improved student learning. Our findings contribute to the definition of learning improvement generally and in STEM, noting the importance of resource considerations.

Correspondence E-mail: laycocla@jmu.edu

Keywords: Learning Improvement; Curricular Change Assessment; Historical Data Analysis; STEM Education Outcomes; Program-Level Student Learning Objectives

The higher education assessment cycle articulated by Erwin (1991) includes the articulation of student learning objectives, formulating an assessment design, collecting and analyzing data, and using results. The primary goal of assessment is to use results to improve student learning. In the mid-1990s and early 2000s, states and institutional accreditors began requiring assessment of student learning as a means of accountability (Ewell, 2009). These dual purposes (i.e., improvement and accountability) seemed to be in tension, even though the accreditation standards required *improvement* (Smith et al., 2015).

Fast-forward to today and assessment for accountability is evident. For example, among institutions accredited by the Southern Association of Colleges and Schools – Commission on Colleges (SACSCOC) in 2013, 21% of institutions were found non-compliant with the degree program assessment standard; in 2023, this figure fell to 6% (Matveev, 2014, 2024). In addition to institutional accreditation requirements, disciplinary accreditors require assessment. As an example, ABET, a disciplinary accreditor for engineering and applied natural science programs, has Criterion 4: Continuous Improvement (ABET, 2021b).

Evidence of improved student learning resulting from assessment, however, is lacking. Indeed, Banta and Blaich (2011) found very few examples of learning improvement – that is where programs assessed student learning, made targeted changes, and monitored learning over time to determine if the change was indeed an improvement. How could this be!? One explanation is that the definition of improvement was unclear.

In 2014, Fulcher et al. presented a clear definition of learning improvement through their seemingly 'simple' model: one must assess student learning at the degree program level (Assess), make a change (Intervene), and then re-assess learning to determine if the change was indeed an improvement (Re-assess). However, learning improvement as defined by Fulcher et al. (2014) is resource intensive.

Specifically, learning improvement projects require energy from multiple faculty members given the focus is on programmatic outcomes. One faculty member – a champion – is necessary to provide momentum and stamina (Fulcher & Prendergast, 2021). Often, assessment practitioners and educational developers from outside the program are needed to support this work (Smith et al., 2018). Last, learning improvement projects often take years to implement. It's a big lift! The resources needed to initiate a learning improvement project are problematic given that faculty capacity following the COVID-19 pandemic is increasingly limited (McClure et al., 2023). And yet, care and attention to increasing educational quality is important work that must be done.

In this paper, we explore an approach to learning improvement that challenges Fulcher et al.'s (2014) model breakdown where a program lacks baseline data. Indeed, if a program has consistently gathered student work (i.e., data) without assessing it, there is an opportunity to retroactively assess the efficacy of a curricular change with significantly fewer resources than beginning a new learning improvement project.

Background

Theoretical Framework – Learning Improvement

As noted, Fulcher et al. (2014) provided a 'simple' model for learning improvement: 1) Assess student learning at the degree program level, 2) Intervene, or do something differently in the curriculum, and 3) Re-assess future cohorts of students in the

program to determine if the changes were indeed improvements. While conceptually simple, it is tough to pull off. Often, programs have one or two of the three components, but rarely all three in coordination.

Highlighting this, Fulcher et al. (2014) describe three ways the simple model can break down:

1. *"~~Assess~~, intervene, re-assess*. For this breakdown, faculty make a coordinated curricular change. Following implementation, they assess student learning. While significant, Fulcher et al. note that this is not true learning improvement given the lack of baseline learning.
2. *Assess, ~~intervene~~, re-assess*. Here, a program dutifully assesses student learning annually without making curricular changes. Essentially, the focus is on the measurement of learning rather than changes to the curriculum.
3. *Assess, intervene, ~~re-assess~~*. In this scenario, a program assesses student learning and makes intentional curricular changes. However, they do not assess cohorts of students who experienced the curricular change to determine if the change was an improvement. Given the years it often takes to implement changes that span a curriculum, the assessment component can get lost" (p.6).

Completing all the steps of the simple model is challenging on multiple fronts, not only to initiate, but to sustain and see to completion. Indeed, successful projects require significant human resources: multiple faculty members from the program, a faculty champion who leads the charge, assessment practitioners to refine measures (e.g., ensure measures are sensitive enough to capture change) and educational developers to collaborate on scaffolded curricular change. Specifically, the faculty champion is one who not only 'lead the charge', but often contributes sustaining energy to the effort as well as coordinating other needed experts (Fulcher & Prendergast, 2021). Additionally, because creation and implementation of changes to a curriculum can take years to implement, getting results to determine if the changes were improvements is a long-term time investment.

Since the publication of the simple model in 2014, a few programs have successfully engaged in a learning improvement project. A Learning Improvement Community was formed in 2017 (*Learning Improvement Community: About Us*, 2024) and includes nine learning improvement stories (*Learning Improvement Community: Stories About Learning Improvement*, 2024). One of the first examples of learning improvement was James Madison University's Computer Information Systems, where the program dramatically improved students' ability to elicit requirements during consultations with clients (Lending et al., 2018). The project took several years to complete, included a core team of four faculty members, an assessment practitioner, an educational developer, and was supported by the program's administrative team (i.e., department head). This example illustrates the amount of time and resources needed for a successful learning improvement project. Considering that many faculty are still feeling the effects of the disruptions of the COVID-19 pandemic (i.e., rapid and sudden shifts in teaching modality, general uncertainty, etc.), faculty capacity for such a resource-intensive effort is often very low.

Given resource limitations, we challenge Fulcher et al (2014)'s first model breakdown (~~*Assess*~~*, intervene, re-assess*). In this scenario, a program has no baseline assessment data, but has successfully implemented a major curricular change. This is the situation in which the lead author found herself. However, this was not considered a dead-end for learning improvement, given a host of positive situational factors. First, capstone student work had been retained from before, during, and after the curricular change. Second, the academic unit head expressed interest in engaging in learning improvement. Third, the faculty noted they were curious about the impact of the prior curricular change, and also indicated they had limited capacity to engage with anything new. The purpose of this paper is to explore learning improvement with existing data to determine if a prior curricular change was effective. If successful, this approach could provide evidence of learning improvement with minimal faculty resources.

Program of Interest: Integrated Science and Technology

The Integrated Science and Technology (ISAT), BS program is situated in an R2 institution in the mid-Atlantic region. The unit head expressed interest in engaging in a learning improvement project. Before initiating the project, the authors used a *Readiness for Learning Improvement* tool – a resource created by the aforementioned Learning Improvement Community (Lambert & Good, 2024). This tool was used to gather perspectives from faculty members about their readiness to engage in a resource-intensive project by considering the capacity, culture, and commitment of the program.

Ultimately, the tool highlighted that the program had a strong culture for learning improvement and a high level of commitment from department leadership – two of the three key ingredients for learning improvement. Unfortunately, faculty clearly expressed that they did **not** have the capacity to engage in a project. At the same time, they expressed an interest in better understanding the impact of a prior curricular change that was implemented over the past several years. This interest inspired the current project, where we consider a new avenue to learning improvement: the use of existing data. This approach could embrace the positive culture and supportive leadership while limiting resource needs.

Prior Curricular Change

In 2010, ISAT faculty initiated a review of their upper-level curriculum to define and implement elements that would make ISAT graduates uniquely valuable. This process led to the development of 'habits of mind' that would characterize ISAT graduates working in the field years after graduation. The result was a list of objectives and a proposed series of courses dubbed the "Holistic Problem-Solving Spine" (hereafter "Spine"), which focused on complex problem solving and systems thinking. In 2017, this work was integrated into student outcomes, prompting the creation of new courses. During the readiness evaluation process, faculty expressed interest in understanding the impact of this curriculum change, which had been based on prior assessment data.

The Spine was designed to allow ISAT majors to practice a holistic approach to problem solving – a valued way of thinking in the discipline. Importantly, the Spine was designed to encourage students to consider systems as the underlying object of study, to evaluate culture, institutions, nature, and technology as intertwined systems, and to allow for collaborative learning communities. It accomplishes this by giving a structured

exposure to complex problems as well as an introduction to a broader interdisciplinary community. As students advance through the Spine courses, they encounter progressively more open-ended and ill-defined problems, learning methodologies and skills to understand these challenges and define specific areas of inquiry within them.

During the Spine's creation, the faculty created new program-level student learning objectives and aligned the scaffolded courses to these outcomes. Given the intensity of work needed from the faculty to redesign the curriculum, shift credit hours, and design new courses, the process took years. Courses were designed and implemented sequentially, starting with the first-year level classes (100-level courses). The following year, the first-year level courses were revised as needed and the sophomore level classes were implemented, and so on. The first students to have experienced the entire Spine series of courses graduated in 2019.

Faculty noted that while it had been nearly five years since the first graduates experienced the full Spine series of courses, there had been no evaluation of the objectives. Was student learning better after implementation of the Spine? They further noted that in the intervening time, one of the courses had been altered from how it was originally imagined. While acknowledging the major impact the COVID-19 pandemic had on the ability to gather accurate representations of student work, faculty still felt as though this major curricular change should be assessed.

Data situation

Since its inception, the ISAT program has had students research, write, and present a senior capstone project. Students work independently or in groups during their senior year to complete these projects, and they have been a much-celebrated hallmark of the program. We were able to determine that the written capstone papers have been saved electronically since the program's start (1996). While examples of learning improvement projects are rare in the literature, what examples do exist (e.g., Good, 2015; Learning Improvement Community, 2022; Lending et al., 2018; Smith, 2017) have used data gathered during the course of the project. These stories follow the Fulcher et al. (2014) format for documenting learning improvement (i.e., Assess-Intervene-Re-assess).

To evaluate the efficacy of an intervention, it is important that consistent student products are available both before and after the intervention. In our case, we had just that: historical data (i.e., student capstone projects) from before, during, and after the creation and implementation of the Spine. The purpose of this paper is to engage in a learning improvement project using the available historical data. To date, learning improvement projects have been future-oriented, and the authors are unaware of existing literature on learning improvement projects using historical data. Focus on the impact of a prior curricular change – if successful – can save time and provide meaningful evidence to stakeholders.

Method

Student work

Since the beginning of the program in 1996, capstone research has been designed such that students in the ISAT program work with a faculty advisor to complete their research activities. Given the integrated nature of the program, as well as the different concentration options, student projects are varied in scope and topic. Additionally, the

faculty advisors ultimately grade the capstone papers, each coming from diverse content backgrounds. This diversity is reflected in the papers: topics can range from social science to bench science, with formats ranging from more typical prose to a paper that could be published in a scientific journal.

We gathered capstone papers from both before and after the creation of the Spine, focusing on the time periods that were definitively "pre" and "post". We defined pre-Spine as 2010 and earlier, since the creation and implementation of the Spine was a multi-year process, with courses being added in a staggered fashion. Additionally, we defined post-Spine as 2019 and later, with 2020 being excluded from the analysis due to the major interruptions of the pandemic. Capstone papers from 2019 represent the first class to have fully participated in the Spine sequence of classes, and all future cohorts will have fully participated. In addition to papers from 2019 being included in the analysis, papers from 2021 and 2022 were also included. Papers from 2023 were not yet available at the time of rating.

Papers were deidentified to minimize any indication of whether the paper was pre-Spine or post-Spine. Specifically, student names were redacted. Additionally, if an "Acknowledgements" section was present in the paper, that was also redacted to remove references to advisors. That said, citation dates, tools, and software referenced throughout the papers likely provided hints as to when a paper may have been written. Raters were instructed to ignore any such information and to strictly rate the paper on the objectives being examined and the rubric being used.

Measure

A rubric was created to evaluate a subset of the Spine learning objectives (see Appendix A); only objectives aligned with the capstone paper were included in the rubric. A previous rubric used for program assessment served as the baseline for the rubric used in this study, which has tighter alignment with the Spine objectives. We acknowledge that not all Spine objectives can be assessed from capstone papers. When training raters, we asked that they only evaluate the elements on the rubric (e.g., do not evaluate writing quality). Several suggestions from the faculty raters were incorporated into the final version of the rubric used. Given that our interest was in the mean difference between pre-Spine vs. post-Spine scores, exact rater agreement on each item was less of a concern than was rater consistency.

Rater recruitment

Faculty from the department were invited to serve as capstone paper raters during a summer rating session. Raters were in-person for one full day of rater training followed by a time of asynchronous rating. A total of five faculty members agreed to serve as raters and participated in rater training. During the rater training, raters had the opportunity to examine the rubric and the objectives being examined, to have any questions clarified, and to suggest changes or improvements to the rubric. After faculty were acquainted with the rubric and clear on the interpretation of each prompt, they used it to rate two different capstone papers, with group discussion happening between each rating session.

After training, the capstone rating sessions were asynchronous, and raters managed their time within the agreed-upon rating window. Faculty were paid an hourly rate for their time in recognition of work performed over the summer, outside of the

Table 1. *Objectives Evaluated During the Capstone Rating Process*

	Objective
A	Describe the systems out of which a defined complex problem emerges, including interactions between the social, cultural, natural, and technological forces.
B	Evaluate the structure of a system to explain how the system and its components affect the complex problem.
C	Chronicle the history, evolution, and current manifestations of a complex problem.
D	Relate historical events and illustrate how these influence the trajectory of the current system.
E	Use an appropriate temporal scale to evaluate solutions.
F	Account for the governance contexts and constraints that impact a complex problem's stakeholders.
G	Identify avenues and processes for change within a given governance context.
H	Account for stakeholder perspectives, including cultural and institutional power differentials, to inform decision-making.
I	Apply relevant and credible sources of information to address a complex problem and its dimensions.
J	Select an appropriate method to define, analyze, and solve a complex problem.
K	Evaluate and integrate multidisciplinary sources of information to analyze a complex problem.

Note. Each objective starts with "ISAT graduates will…"; this was removed for the sake of brevity.

typically contracted academic year. Of note, some raters agreed to rate a total of 40 papers while others agreed to rate 80 papers – this was due to personal time constraints of each individual. Capstone papers were delivered to the faculty via a shared folder on the institutional OneDrive network along with a rating spreadsheet for data entry.

Rating

Overall, the five raters evaluated 193 capstone papers in total. All post-Spine papers were rated, given that they only spanned three years. For the pre-Spine papers, a random number generator within R (R Core Team, 2025) was used to select which pre-Spine papers to rate. There were three common papers embedded within each of the faculty members' folders to permit checking of inter-rater reliability if desired in the future.

Each paper was examined with the eleven Spine objectives in mind. Faculty raters rated each objective as a 1 (low), 2 (medium), or 3 (high), resulting in a paper having a possible maximum score of 33 and a minimum score of 11. In addition to a numerical rating, raters had the option to leave qualitative feedback either for the overall process or for specific capstone papers. This qualitative feedback was not analyzed explicitly for this

study, though it was discussed in the rater debrief session, and will likely lead to future projects.

Statistical Methods

Our main research question was whether historical data could be used to provide evidence of learning improvement. Our data was generated from the ratings assigned on the rubric by raters. Given that there were two groups being compared - capstone papers written prior to and after the implementation of the Spine curricular change - an independent samples t-test was used within R (R Core Team, 2025). This was not a traditional pre/post repeated measures design, as the papers written in the pre-Spine group did not share authors with papers written in the post-Spine group. After evaluating the assumption of homogeneity of variance via a Levene's test, it was determined that equal variances were not present, and the Welch two-sample t-test was used to compare the means. Additionally, Cohen's d was calculated to determine the practical significance of any differences.

In addition to the overall difference of pre-Spine vs. post-Spine, we had information broken down by objective. We further examined the effect of the Spine broken down by objective by using an independent samples t-test with a Bonferroni adjustment to account for the multiple tests being performed. This allowed us to examine which objectives changed, and which, if any, did not change after the implementation of the Spine. Effect sizes were also calculated to determine the practical significance of any differences.

Reliability

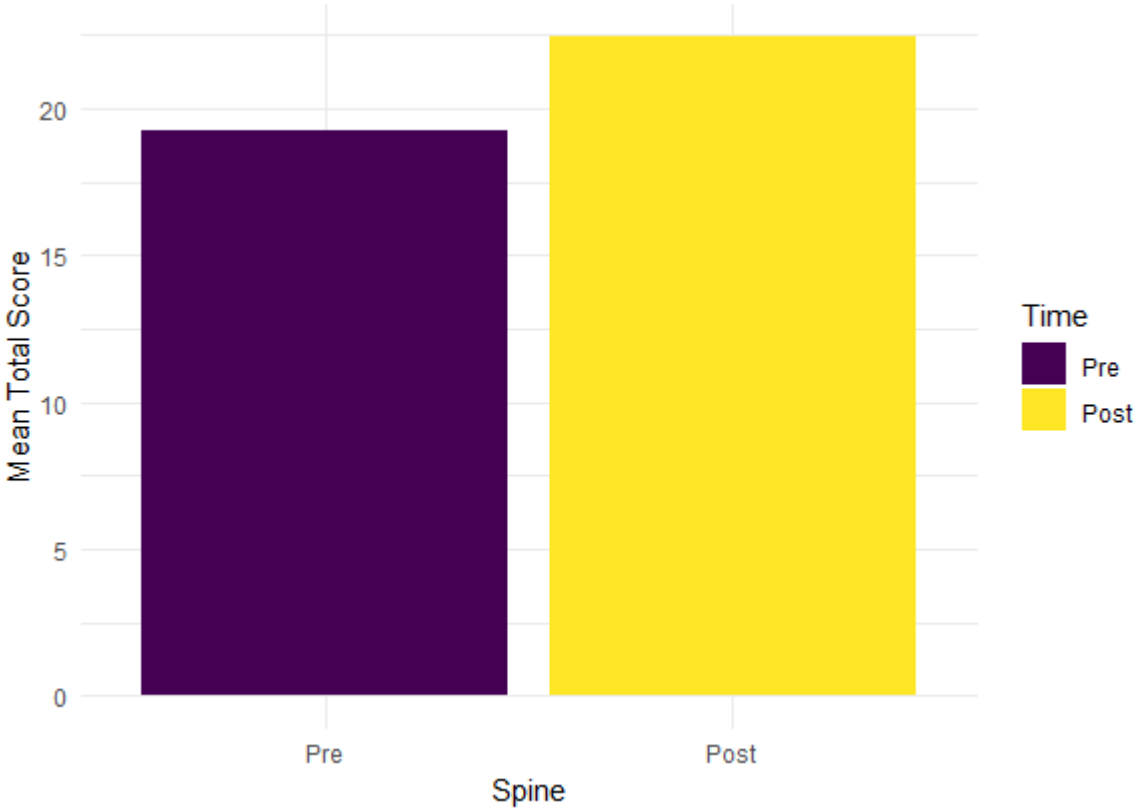
We elected to have a design where each capstone paper was only read by one rater, rather than by two raters. While this did not allow for a calculation of inter-rater reliability, this did allow us to rate twice as many papers. Further, given our parameter of interest was the difference between group means and each rater was scoring an even mix of pre-Spine and post-Spine papers, having one rater per capstone paper would result in a smaller error variance. This design resulted in raters being fully crossed with group, with each rater scoring an equal ratio of pre and post capstone papers. Therefore, the mean square error (MSE) was equal to $\frac{\sigma^2(s:g)}{n_s} + \frac{\sigma^2(rg)}{n_t} + \frac{\sigma^2((s:g)*r)}{n_s n_r}$, where n_t is the total number of raters, n_s is the harmonic mean, and n_r is the number of raters who score each student.

Ultimately, we decided it was better to have more capstone papers rated with a smaller error variance but larger individual student error variance than fewer papers rated and smaller individual student error variance. Another way to consider the MSE is that the *(s:g)*r* component will be twice as large with one rater per paper, but when computing the group mean there will be twice as many papers to average. Therefore, *(s:g)*r* will contribute the same error to the group mean regardless of whether we had 1X papers each rated by 2 raters or 2X papers each rated by 1 rater (C. DeMars, personal communication, 8 May 2023).

Results

Overall

The average total score for papers completed prior to the Spine (N = 110) was a 19.2 while the average total score for capstone papers completed after the Spine (N = 83) was 22.4, resulting in a significant difference ($t(157.87) = 4.66$, $p < .001$), with a 95% confidence interval for the mean difference of (1.84, 4.55). This indicated that for the objectives evaluated using our rubric on student capstone papers, students performed significantly better after the introduction of the Spine series of courses as compared to before. From a learning improvement lens, this indicated that student learning improvement was able to be captured, and the curricular change of the Spine resulted in improved student performance on the objectives we evaluated.



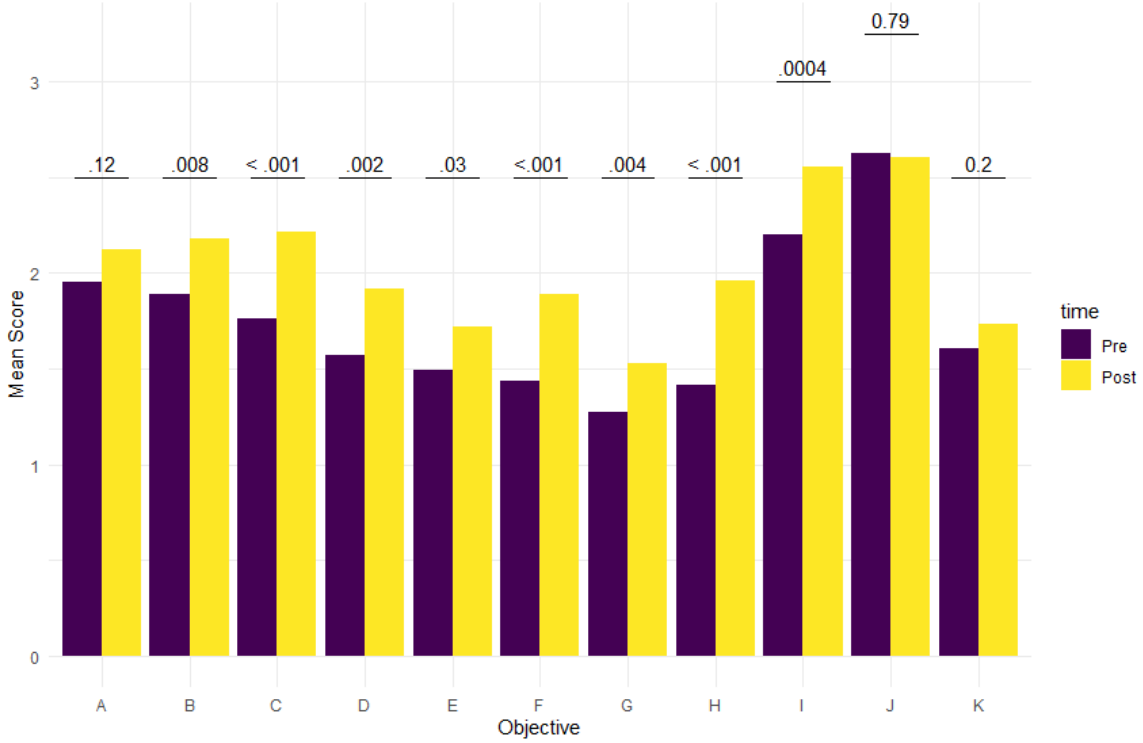
Note. Pre = pre-Spine; Post = post-Spine. The mean difference was statistically significant (t(157.87) = 4.66, p < .001).

Figure 1. *Group Mean Comparison of Student Performance*

We also determined if this was a practically significant effect. Cohen's d was calculated to be 0.69, which, by established rules of thumb, is a moderate to large practical effect size (Cohen, 1988). This can be interpreted to mean that not only is there a statistically significant difference in overall student performance on the assessed learning objectives after the curricular change, but there is also a practical, 'real-world' difference.

Objective breakdown

In addition to examining overall results, we also examined results by objective. Given that multiple pairwise tests were being run, a Bonferroni adjustment to the alpha level was used, resulting in an alpha of 0.005.



Note. Pre = pre-Spine; Post = post-Spine. P-values are indicated above each objective.
Figure 2. Group Mean Differences Broken Down by Objective

Examining each objective (See Table 1 for a complete list), it was found that 7 out of 11 tested showed significant improvement. Of those that did not show improvement, one was likely due to a ceiling effect: objective J was already at 2.6/3, leaving little room for improvement. This objective addressed student ability to select an appropriate method to solve a complex problem. Objective K also did not show significant improvement. This objective was evaluating student ability to integrate multidisciplinary sources of information. Unlike objective J, this lack of difference is not likely due to a ceiling effect (mean of 1.7/3 in the post-Spine group), and instead pointing out an opportunity for further student improvement in the future. Objectives A and E were also not significant and are again likely opportunities for student improvement in the future, and perhaps a potential area of focus for future learning improvement projects. These objectives deal with a student's ability to describe the systems out of which a defined complex problem emerges (A) and use an appropriate temporal scale to evaluate solutions (E). Objective E specifically was referenced in the faculty debrief session as one that may need to be revisited in the future, as its interpretation seemed to be vague.

While objective G did show significant improvement from pre-Spine to post-Spine, overall mean scores were the lowest of all the objectives (1.53/3 in post-Spine). Objective G had been pointed out during the faculty debrief session as one of two objectives that

seemed redundant, dealing with identifying avenues for change within governance contexts. It is possible that this objective and objective F, which dealt with accounting for governance contexts that may impact stakeholders, overlapped too much. Alternatively, it may be that faculty struggled to differentiate the two within the capstone paper, resulting in lower scores on objective G. Regardless of the reason for the low scores, faculty have indicated a desire to revisit objectives F and G specifically to better determine what is unique between them and if both are needed.

Discussion

Learning improvement – that is, true evidence that curricular changes were indeed improvements – is worthwhile. However, learning improvement projects are complex and resource-intensive. Often, programs engage in robust curricular change without assessment. This 'breakdown' in the simple model (Fulcher et al., 2014) can also be seen as an opportunity when student work has been retained from both before and after a significant change.

Undertaking a post-hoc evaluation of learning improvement following a curricular change can be viewed as an easier approach to documenting real learning improvement, as well as contributing to more rigorous evidence to regional and disciplinary accreditors (e.g., ABET). This approach still takes the approach of the 'simple model' as proposed by Fulcher et al. (2014) while adding a viable alternative approach practitioners can take if parts of the model have broken down. In particular, this approach offers an alternative when programs have an intervention, but no assess or re-assess (e.g., when it is ~~assess~~ – intervene – ~~re-assess~~) (Fulcher et al., 2014). Given that programmatic curricular changes happen regularly, yet assessment of these changes is not common, we propose that this approach will be of value to many programs. For those programs that have a disciplinary accreditor, in particular ABET, this approach can aid them in capturing improved student learning and continuous improvement (ABET, 2021a).

Additionally, in this situation, the curricular change has already occurred – what is potentially the hardest part of the simple model is done. Recognizing the intensity and heavy lift of that work, assessing how the curricular change impacted student learning allows faculty to see if all that work 'paid off'. Further, assessing after the intervention is markedly easier than starting a new learning improvement project from scratch – there is much less new work to be performed. This approach honors the hard work already done while still capturing any change in student learning.

Further, learning improvement projects require careful consideration of cultural norms, available resources, and the willingness to adapt to change within academic programs. Namely, it is important to consider the human resources alongside the more commonly considered financial resources. In the case of ISAT, the faculty were exhibiting a culture and commitment to assessment, but were lacking in capacity. A unique aspect of our study was the use of historical data to assess the impact of a prior curricular change—the introduction of the "Spine" series of courses. Most learning improvement projects focus on future-oriented assessments, making our approach distinctive. Historical data spanning a curricular change provided a rare opportunity to evaluate the long-term effects of a programmatic shift. The use of historical data allowed us to examine the tangible effects of the curricular change, offering valuable insights, while respecting the lack of

capacity expressed by faculty with respect to starting a learning improvement project "from scratch", as described by Fulcher et al. (2014).

Generalizability

This study captures the ability of a program to 'look back' at curricular changes – this could be a series of courses, as illustrated here, or other changes such as the addition/removal of single courses, capstone or research requirements, or a reworking of Student Learning Outcomes. Programs are dynamic, with courses, outcomes, and requirements being evaluated in light of changing student needs. Additionally, it has been well documented that Learning Improvement projects, while providing valuable insights into student learning, are very resource-intensive (Fulcher & Prendergast 2021; Fulcher et al. 2014). A backwards-looking approach to learning improvement makes it possible for programs to evaluate prior curricular changes, as well as alleviating some of the resource intensity of building a learning improvement project from the ground up.

While this project was in a STEM program accredited by ABET, the process is not exclusive to such a program. Any program that has continuity of student artifacts – portfolios, final exams, course projects, etc. – over a curricular change could take this approach. The key piece was the availability of student work. For this study, it was programmatic data collection (i.e., the program collected and stored student capstone reports). However, if a history program implemented a curricular change and Professor Jones had a record of student papers she assigned each semester, those papers could potentially serve as artifacts spanning the curricular change. Each program will have to consider the context of their change, and then discuss with faculty what student work is available to 'look back' over the change. From there, they will be able to determine if they have adequate coverage, both before and after, of the curricular change to be able to assess the impact on student learning. The defining feature of student work in this application is continuity and relevance to the outcome(s) being examined rather than a specific assignment type.

Additionally, many institutions consider the assessment timeline to be on an annual cycle – programs are often asked to submit a report annually. Given this short institutional focus, a longer learning improvement project may not be prioritized. It will take longer than a single year to conceive, design, implement, and assess. This approach can be beneficial for programs accredited by ABET. The Continuous Improvement criteria of ABET (*ABET*, 2021a) offers a longer view of the program's improvement history, and this type of work could be productive in showcasing the efficacy of changes made. Recognizing that many programs implement curricular change on more anecdotal evidence (e.g., "Our students seem to be struggling with xyz. Let's implement these curricular changes to support them."), taking the approach detailed in this paper will allow programs to capture the results of their efforts. This approach also allows programs to capture the effects of curricular changes for external accreditors.

Faculty Benefits

An unanticipated, yet positive, secondary outcome of this study was that the rating experience itself was a form of professional development for the faculty engaging in the rating. During the debrief session held after faculty had completed rating the capstone projects, it was found that they had deeper engagement with the Spine student learning

objectives as well as with the capstone project. Faculty observed that there were objectives that may need to be revisited, and others that seemed redundant. This has spurred further faculty discussion around these objectives, feeding into the paradigm of gathering data and using that data for curricular change.

Additionally, faculty were able to identify a seeming deficiency in student writing ability and a perceived stylistic change from capstone papers written pre-Spine to those written post-Spine. This has also spurred further conversation around the topic, including where in the curriculum students are exposed to writing tasks, where in the General Education curriculum they are exposed to writing, and if the program needs to revisit how writing is taught. The stylistic change has also prompted conversation, jointly with the writing skills, surrounding the creation of a style guide or manual that both advisors and students could reference. The acknowledged challenge surrounding the style and quality of writing in the capstone paper stems from academic freedom of the capstone advisors; these advisors are the ones who grade and provide feedback on the paper. Understandably, faculty have different views as to length, content, and what constitutes quality. Regardless, this experience has spurred a desire for further conversation around this topic.

Perhaps an important lesson from the debrief is that engagement should be kept going so that when faculty have the capacity to engage with larger learning improvement projects, the culture and commitment will be well maintained to support such an effort. It is likely that further curricular adjustment will occur as a result of this project, as well as the potential for a larger learning improvement project.

Limitations

One major limitation in this study is the variability of the capstone papers themselves. The exact contents, length, and rigor are determined by the individual capstone advisors, leading to variability in final products. Additionally, each concentration area varies in the approach to scientific writing. Some lean much more towards the approach taken in primary literature of biology and chemistry, while others approach writing more like the social sciences. This variety in approaches also leads to a variety in final products. We are hopeful that the impact of this variety can be minimized, as all concentrations are represented in both the pre-Spine and post-Spine groups.

Conclusion and Future Work

Our study emphasizes the multifaceted nature of learning improvement projects in higher education. It underscores the importance of considering situational factors, such as capacity and historical data, when embarking on such endeavors. By using historical data to assess the impact of a curricular change, we both illustrate the feasibility and additionally provide a pathway for other programs interested in documenting learning improvement to examine existing data in a new light.

The implications of this study are potentially very impactful – programs wishing to engage in learning improvement projects yet that lack the resources to do so 'from the ground up' may find themselves able to make use of historical departmental data to evidence learning improvement. A program interested in taking this approach needs only to look for student work that has spanned curricular changes that can map to objectives impacted by those changes. Perhaps that comes in the form of a senior capstone report, as presented in this study. However, it may also come in the form of a final project for a

course, a homework assignment that has remained largely unchanged for a number of years, or presentations given during courses. While it is recognized that historical data will not exist for every situation, there are situations where it may exist. This will not only allow programs to evaluate curricular changes through a learning improvement lens, but it will also reduce the resource intensity of a learning improvement project, as well as looking at improvement over a larger time scale than the yearly assessment cycle seen in most institutions.

Appendices

Appendix A: Spine learning objectives rubric

Acknowledgements: We would like to thank the Integrated Science and Technology department for allowing this project to take place and for giving us access to the historical capstone reports. Additionally, we would like to thank the faculty raters who took time out of the start of their summer to rate student reports. We would also like to thank Dr. Christine DeMars for her expertise in reliability in aiding study design.

References

ABET. (2021a). abet.org

ABET. (2021b). *Assessment Planning*. https://www.abet.org/accreditation/get-accredited/assessment-planning/

Banta, T. W., & Blaich, C. (2011). Closing the assessment loop. *Change (New Rochelle, N.Y.)*, *43*(1), 22–27. https://doi.org/10.1080/00091383.2011.538642

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). L. Erlbaum Associates.

Erwin, T. D. (1991). *Assessing student learning and development: A guide to the principles, goals and methods of determining college outcomes*. Jossey-Bass Inc.

Ewell, P. T. (2009). Assessment, accountability, and improvement: Revisiting the tension. *National Institute for Learning Outcomes Assessment*, *Occasional Paper #1*. https://www.learningoutcomeassessment.org/documents/PeterEwell_005.pdf

Fulcher, K. H., Good, M. R., Coleman, C. M., & Smith, K. L. (2014). *A simple model for learning improvement: Weigh pig, feed pig, weigh pig*. https://in.ewu.edu/facultycommons/wp-content/uploads/sites/129/2016/12/A-Simple-Model-for-Learning-Improvement_Weigh-Pig-Feed-Pig-Re-Weigh-Pig.pdf

Fulcher, K. H., & Prendergast, C. (2021). *Improving student learning at scale: a how-to guide for higher education* (1st ed.). Stylus Publishing, LLC.

Good, M. R. (2015). *Improving student learning in higher education: a mixed methods study* [ProQuest Dissertations Publishing]. https://commons.lib.jmu.edu/diss201019/18

Lambert, L., & Good, M. R. (2024). An intentional approach to starting a learning improvement project. *Intersection: A Journal at the Intersection of Assessment and Learning*, *5*(4), 25–34.

Learning Improvement Community. (2022, March). [Online document]. *Readiness for learning improvement tool*. https://learning-improvement.org

Learning improvement community: About us. (2024). [Organization website]. https://learning-improvement.org/about-us

Lending, D., Fulcher, K., Ezell, J. D., May, J. L., & Dillon, T. W. (2018). Example of a program-level learning improvement report. *Research & Practice in Assessment*, *13*(2), 34.

Matveev, A. (2014). *Top 10 most frequently cited principles in reaffirmation reviews: 2013 reaffirmation class institutions*. SACSCOC. https://sacscoc.org/app/uploads/2019/09/Most-Frequently-Cited-Principles_2013_Track_A_B-rev.pdf

Matveev, A. (2024). *Most frequently cited principles of accreditation in decennial reaffirmation reviews: Class of 2023*. SACSCOC. https://sacscoc.org/app/uploads/2024/03/2024_Most-Frequently-Cited-Principles_Class-of-2023_web.pdf

McClure, K. R., Sallee, M. W., Ford, J., Gonzales, L. D., Griffin, K., Jenkins, T. S., Kiyama, J. M., Martinez, M., Martinez-Aleman, A. M., Miao, S., Perez, R. J., & Porter, C. J. (2023). *The impact of covid-19 on faculty, staff, and students: Using research to help higher education heal through the pandemic and beyond*. Association for the Study of Higher Education. https://www.ashe.ws/positiontaking

R Core Team. (2025). *R: a language and environment for statistical computing* (Version 4.5.1) [Computer software]. R Foundation for Statistical Computing. www.R-project.org

Smith, K. L. (2017). *Integrating implementation fidelity and learning improvement to enhance students' ethical reasoning abilities* [James Madison University]. https://commons.lib.jmu.edu/diss201019/153

Smith, K. L., Good, M. R., Sanchez, E. H., & Fulcher, K. H. (2015). Communication is key: "Unpacking" use of assessment results to improve student learning. *Research & Practice in Assessment*, *10*, 15–29.

Smith, K. L., Good, M. R., & Jankowski, N. (2018). Considerations and resources for the learning improvement facilitator. *Research & Practice in Assessment*, *13*, 20-26.

Appendices:

Appendix A: Spine learning objectives rubric