Testing is Teaching Too: Transitioning a Large-Lecture Course from Summative to Formative Exams

Robert Arens
Professor and Associate Department Head, California Polytechnic State University, San Luis Obispo

Brian Osborn
Assistant Professor, California Polytechnic State University, San Luis Obispo

Carmen Trudell
Associate Professor, California Polytechnic State University, San Luis Obispo

Abstract

For students and teachers alike exams can be a dreadful experience with both parties left questioning the value of the exercise. Large-lecture courses tend to employ an exam culture that is more focused on expedience than efficacy as the promise of efficient grading often triumphs over the desire to create meaningful learning experiences. Within the Architectural Technology Fundamentals courses at Cal Poly we have found that machine-readable tests, which use multiple-choice and true-false questions, tend to assess students' understanding of course topics at only the most basic level and are misaligned with our aspiration to foster students who can integrate and apply their knowledge of course topics to their own design work.

In response, we have transitioned away from a mode of summative assessment and toward exams that we consider to be formative teaching tools in themselves. These include vignette-based exams that ask students to apply course topics to architectural scenarios. This paper discusses our use of vignette exams in large-lecture format architectural technology courses and reflects on the advantages and challenges. These insights come from three forms of assessment. First, grading the exams allows for an analysis of student performance. Second, dialogue with students through direct conversation provides input into their personal experiences with the exams. Finally, anonymous surveys assess the effectiveness of exams in supporting student learning.

Our findings indicate that the vignette exams allow for a more revealing assessment of students' understanding of course topics. With machine-readable tests we could see when a student performed poorly in a topic area, however, the nature of their misunderstanding was not always apparent. In contrast, vignette exams reveal specifically where within each problem a student makes a mistake and therefore which aspect of the topic was misunderstood. Further, students report that they experience a holistic and integrated way of thinking through the vignette exams and that they "feel like architects" having completed the test. This sense of working on something meaningful positively impacts students' perception of the relevance of course material to their education and their future lives as professionals.

Keywords: Pedagogy, Exams, Architectural Technology, Large-lecture courses, Course design

The Shift from Summative to Formative

Especially in a large-lecture course, instructors can rely on a small number of exam scores to determine a student's grade in the class. A common exam scenario follows a pattern of students cramming the night before a test by frantically reading the course texts—often for the first time, reviewing lecture notes, and conversing with classmates. Instructors also cram to write machine-readable exam questions that can be efficiently graded. While this has become the normal *testing* ritual, there may not be much *learning* or *teaching* taking place. It became obvious to our teaching team that the way we talked about, wrote about, and administered exams was about generating students' scores for the course. We poured over the numeric data and made judgements about how well our students *understood* and *knew* the content based-on how accurately they would choose between a list of possible answers. Our efficient tests were designed to inspire studying and memorization, which can definitely promote learning, but we realized that we were not designing tests where *learning* was the primary focus. These tests were designed to record *recall*, but did little to further students' *thinking*.

The 1993 publication "Measuring What Counts: A Conceptual Guide for Mathematics Assessment"[1] (MSEB) outlined three principles for assessments. We have found these principles to be useful aspirational goals for own course assessments. The following paraphrase these goals while editing them to remove specific references to mathematics. *The Content Principle*: Assessment should reflect the content that is most important for students to learn. *The Learning Principle*: Assessment should enhance learning and support good instructional practice. *The Equity Principle*: Assessment should support every student's opportunity to learn important content.

It is especially important to note that there is no mention that assessment should be used to assign a grade or score to a student. The language in MSEB is formative in that the assessments are learning focused, rather than summative, in that they allow for a simple culmination of the course instruction.

In fall of 2016 we made a fundamental shift toward exams that are focused on learning. We shifted from machine-readable on-line exams with 50 to 90 questions to human-graded vignette-exams with 3 to 6 questions. Along with this came another change in the resources that we made available during the exam. The multiple-choice tests were administered in a closed-book scenario and required a student to have everything they would need to know accessible by memory. The vignette-exams are open-notes, open-internet, and open-book—encouraging students to know how to navigate the resources available to them (and to any practicing architect). The students now prepare for these tests by revisiting webpages, readings, and course notes. However they do not do this in order to memorize the content but, instead, to ensure that they can find what they might need during the test more quickly and then know how to apply it. The students do not need to know the answer to the fill-in-the-blank, but they do need to know how and where to source sound information to inform their answer. We believe this is a more equitable learning experience, as organization of resources versus memorization of information, is less targeted on a single and particular way of thinking. Students who may not be good at quickly memorizing and recalling are at a disadvantage by the multiple-choice assessment.

In order to have enough multiple-choice questions to fill the testing time, we'd generate a high number of questions that were very narrowly focused and specific. This was misaligned with our broader course goals of educating architects that are able to ask competent and confident questions about the technical aspects of design and practice, and helping students to develop values about the environmental and human impacts of

development. The multiple-choice exams were misaligned with the learning principle, content principle, and the equity principle outlined by MSEB.

From School Work to An Architect's Work

L. Dee Fink is an educational scholar who has been an influential guide to how we are rethinking exams. Fink describes, "…significant learning is learning that makes a difference in how people live – and the kind of life they are capable of living. We want that which students learn to become part of how they think, what they can and want to do, what they believe is true about life, and what they value – and we want it to increase their capability for living live fully and meaningfully."[2] One of the challenges posed by Fink is to get students to think about their education in terms of their life, and not just as something they have to do while they are studying. We approached this by shifting the test away from an assessment that would be perceived as "school work" and moved toward an assessment that would be perceived as "an architect's work." We hoped that this would inspire students to see it as significant toward their chosen profession. We were quite confident our students only saw the multiple-choice as meaningful to their grade in the class, but not to their life. Anecdotally, when students turn in their vignette-exams, we've heard many of them say that "I feel like an architect" which is evidence that they are not in the "school work" mindset. The students perceive this assessment as authentic, and therefore valuable.

At the end of the first year with vignette-exams, we surveyed our class of 140 students about their experience. 110 students responded to the survey. When asked if they thought that the vignette-style exams were preparing them for their future profession (Figure 1), 58 responded either strongly agree or agree. While there is room for improvement here, this number does indicate that the majority of students see the activity of test taking as meaningful beyond the class.

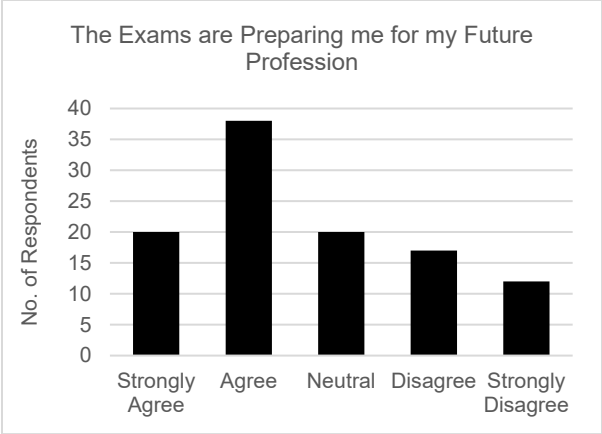

Fig. 1. Student responses to a year-end survey (June 2018) after the pilot year with vignette-exams. 58 of 110 respondents indicate a positive correlation with the exams and their profession after graduation.

Conversely, when asked if students thought that the vignette-exam tested memorization (Figure 2) only 24 students responded that they agree or strongly agree. Compare this to the results when students were asked if they felt challenged to think critically when taking the vignette-exam (Figure 3). 84 Students confirmed that they agree or strongly agree. These three questions taken together can lead to a conclusion that the students do perceive the exams as relevant to their future life beyond school, and also as an assessment that invites them to think critically about architectural issues.

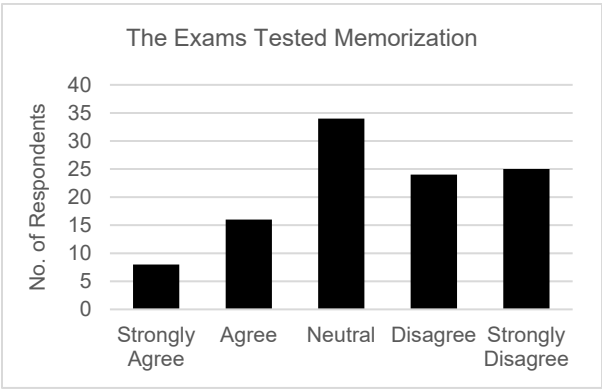

Fig. 2. 48 of 110 students responded that they do not believe the vignette-exams test memorization, compared to 24 students who agree or strongly agree that the exams do test memorization. (June 2018 survey)

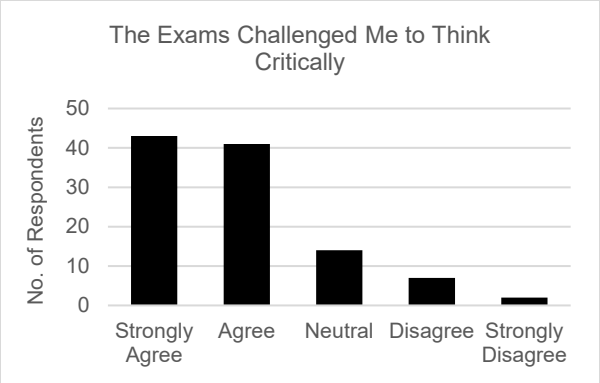

Fig. 3. 84 of 110 students responded that the vignette-exams challenged them to think critically. (June 2018 survey)

From Finished to Feedback

We believe that exams are powerful educational tools and that, if done well, they can be "concrete illustrations of the important goals to which students and teachers can aspire."[3] We will use an example from our 2017-18 course to illustrate how the vignette-exams have increased the quality of communication from student to teacher, and in turn from teacher to student. One of the topics taught in the Architectural Technology Fundamentals class is solar geometry. This foundational knowledge is employed throughout the lessons on daylighting, passive solar heating, solar shading, building orientation and massing. In our class we rely most heavily on polar sun path charts (Figure 4), which is a graph of the sun's positions over a year by latitude drawn in plan (horizontal projection). Understanding how to read the sun path chart is a skill required to be successful in many subsequent topics in the courses.

When assessing students with a machine-readable exam we would present a polar sun path chart and ask students to read it. In general, students did quite well on these questions, whether given in multiple-choice or fill-in-the-blank format. For the example shown in Figure 4, 76% of students answered the question correctly. This result would lead the teaching team to believe that our teaching practices were highly effective.

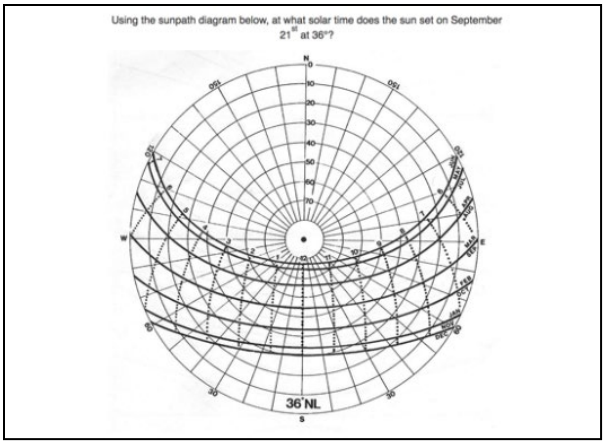

Fig. 4. A multiple-choice exam question assessing ability to read a polar sun path chart. 76% answered correctly (fa 2016).

When assessing the same course content with the vignette-exams (Fig. 5), students first read the sun path chart based on given criteria and then apply that reading to an architectural situation. In the midterm exam for the fall quarter of 2017, the architectural situation given to the students was to locate the best area of a site where a café with rooftop solar photovoltaics should be placed, and to also locate the best location for outdoor seating that would be shaded in the afternoon. To answer this question, students had to use the sun's location to determine shadow lengths and directions and then sketch these shadows on the provided site plan. Grading this question revealed to us that 1/3 of our students were reading the sun path chart incorrectly even though they could answer the first part of the question correctly. Through the three-part vignette question, we found that many students were drawing the shadows inverted from the direction they should have been drawn in. This mistake indicates that students were reversing the position of the sun in relation to the position of the site/body. Without the follow-up questions that required students to do something with the solar information, the instructors previously believed that there was widespread understanding of solar geometry in the class. The reality was that there was a very common misunderstanding

that only came to light when students were asked to apply solar geometry to an architectural problem.

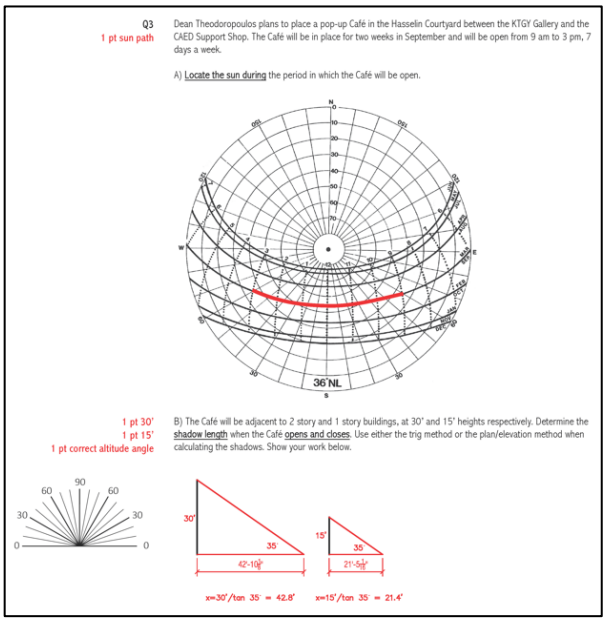

Fig. 5. A three-part vignette question where students first read the sun path chart, then determine the shadow lengths for given sun positions, then sketch these shadows on a site plan to determine the best location for a solar powered café on campus. (fall 2017)

Incidentally, other assignments (not exams) in the class also did not bring this issue to light. The third part of the three-part vignette question asked students to locate two outdoor programs on the site with particular time-based requirements for sun/shade. The question was written such that If students misunderstood the solar geometry they would provide a site design that does not meet the users sun/shade needs. While it may seem like a small misunderstanding initially, the result is an architectural proposal that does not meet the user's needs, which is a significant failure in our eyes. Because of this feedback, and more clear understanding of the student's specific understanding, we have adjusted our teaching practices around this topic. Making visible these learning issues is not just an Architectural Technology Fundamentals problem, but an Architecture problem. We see students making mistakes of a similar nature in their design studio work, and we assume this continues into their early career. Without an assessment tool that provides a concise and clear venue for each of the core learning goals and skills to be expressed, we were not able to fully learn about the quality of the teaching and the learning taking place in the class.

Examining the Exams

There have been many challenges involved in writing and grading vignette exams with many possible correct answers for large numbers of students, often with turn-around times of only a week.

Challenge 1: Generating Questions

After the vignette-exams are graded, our practice is to return exams to students, and provide a detailed rubric showing how to derive correct answers. We see this is an important step in learning-focused exams. Each term and year we then must write new questions to prevent simple copying from last year's rubric. At this point the team is committed to generating new questions, which

entails creating CAD drafted base drawings and continually creating new scenarios. While it is time-consuming, we believe this work is worth the effort.

Challenge 2: Human-Read Exams

We work with a team of 4 instructional student assistants (ISAs) who grade exams based on faculty-generated rubrics. Each ISA grades one question for the entire cohort of students ensuring consistency of grading by question. ISAs spend between 8 and 12 hours each per exam, and it typically takes about 5 days to complete preliminary grading. The team of instructors then randomly checks exams, and if an evaluation issue presents, the instructor will look through all the exams and fix evaluation errors.

Prior to beginning the evaluation period, the ISAs and instructors will meet and look through a number of student exams while also dialing-in the grading rubric. We devise a method of assigning points to particular types of answers. We cannot anticipate the range of answers that will be provided, even when we think we have limited the conditions sufficiently. In some cases, answers are quite clever and clearly demonstrate understanding of the concepts. In other cases answers are bizarre and it is unclear if the student knows what they are doing.

A key to our grading approach is placing an emphasis on the process over the final answer. We allocate points for each step in the process, so that students who demonstrate the right methodology with minor errors are assessed accordingly. In some cases, such as in a question which asks for an answer to be sketched, a student will realize that they made a mistake in the drawing but they won't have time to re-do the work during the exam. We encourage students to explain themselves in the margin if needed. We do not deduct points from a student's score if they provide an explanation that clearly demonstrates understanding, even if there's inaccuracy in the sketch.

Once the exams are returned to the students, the educational experience continues. Because vignette-exams do not necessarily have a single *correct* answer, there is some room for *negotiation*. After the first vignette-exam, students who wanted to know why they were marked-down for their responses inundated our office hours. The discussion quickly degraded to one about scoring which was not the discussion we wanted to have about the course content or about how to learn. In order to reframe these discussions, we introduced an exam wrapper[4]. The exam wrapper is a handout that students completed prior to coming to office hours to discuss their exam. We would give modest credit for completing the wrapper to incentivize those students who didn't do well on the exam to meet with a professor. The exam wrapper asks students three types of questions: How did they prepare for the exam? What kinds of mistakes did they make on the exam? What would they do differently before/during the next exam?

The exam wrapper highlights study practices that are not shown to be effective, such as re-reading class notes, as well as study practices that are highly effective, such as working on sample problems with classmates. Students list the amount of time they spent doing each type of preparation, allowing us to talk about exam study habits rather than points. Another helpful aspect is that the exam wrapper asked students to explain the types of mistakes that were made. This has enabled us to better understand which parts of the exam were confusing to students and write clearer questions with better scaffolding.

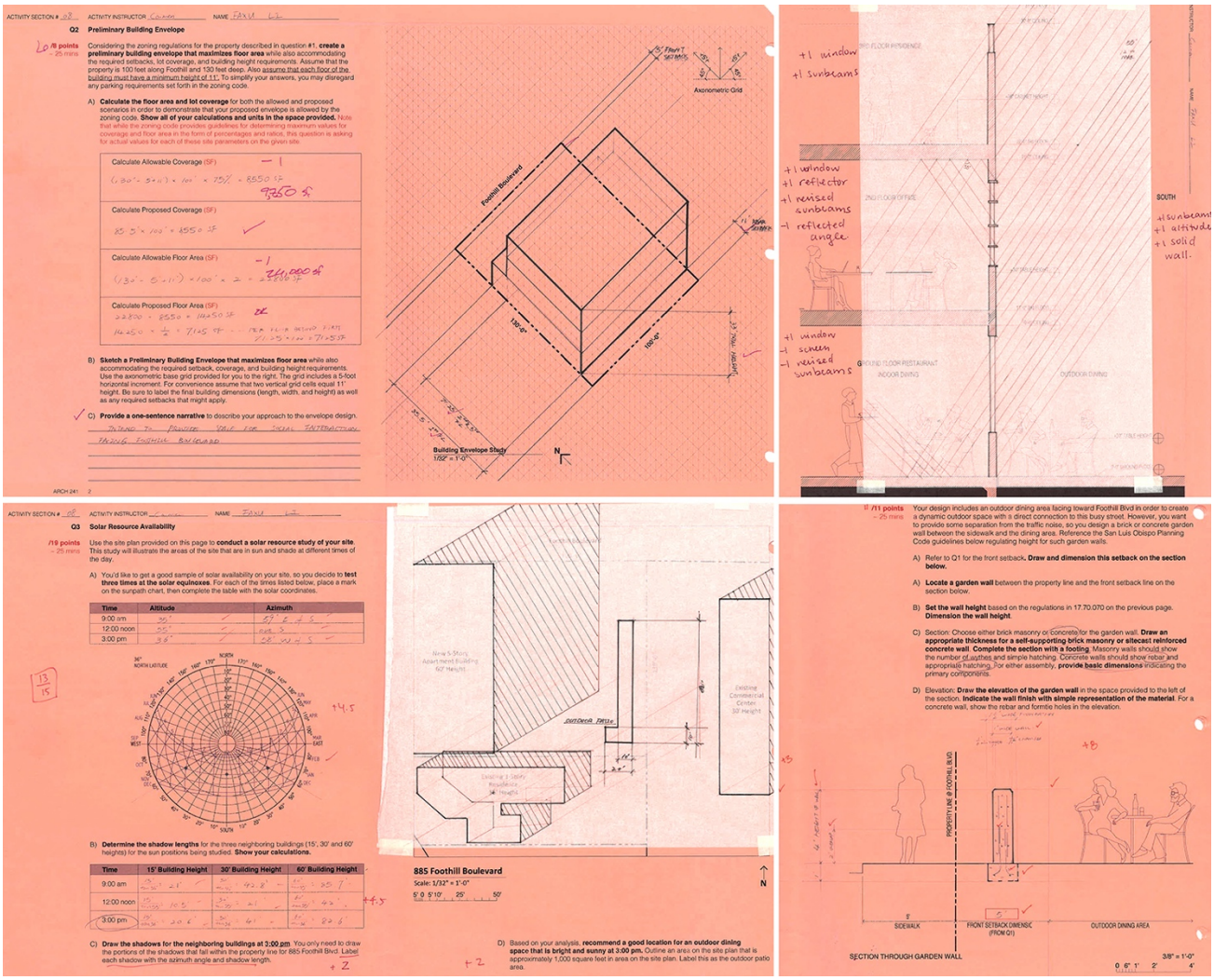

Fig. 6. A final exam where students work through the topics of materials, assemblies, environmental control systems, and site systems sequentially (fall 2018). Actual student answers and grader notes are shown.

Challenge 3: Integrated Topics

The Architectural Technology Fundamentals courses integrate the topics of materials, assemblies, environmental control systems, and site systems, which are taught by three instructors. The first challenge is to write exams that integrate these topics while also not overwhelming students. Our approach has been to write each exam as a single vignette where questions are answered sequentially (Figure 6). In Fall 2018, we provided an urban site in San Luis Obispo, California. Questions 1 and 2 asked students to look up the zoning code online and sketch a diagram of set-backs and lot coverage, then sketch a possible building massing for the given program (site systems). Question 3 analyzed solar geometry, sketched shadows for the adjacent

structures, and determined the best location for an outdoor patio (site systems). In Question 4, students were given a skeleton of a wall section for one wall in their proposed building massing to sketch over in order to design three daylighting schemes (ECS). In Question 5 students chose masonry or concrete to design a code-compliant site wall, documenting their proposals with a section of the wall and its footings (materials & assemblies). In Question 6 students wrote a short essay explaining their material choice in terms of physical properties and human perceptual experience (materials & assemblies).

Fall quarter is the student's first term of Architectural Technology Fundamentals. Great care must be taken when crafting the exams to not overwhelm students, nor to write an exam where a misunderstanding early in the test leads to overall failure in following topics.

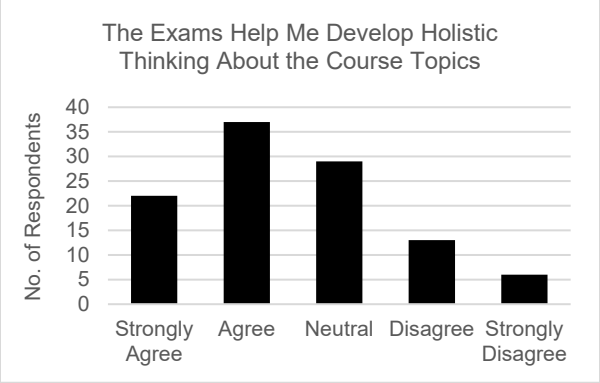

Fig. 7. 59 of 110 students responded positively (June 2018)

When asked to reflect on the vignette-exam and students holistic thinking about course topics, most students reported affirmatively (Figure 7). As vignette-exam designers, this process of writing exams that can successfully integrate the wide-ranging course topics into one coherent scenario, is an excellent litmus test. If the subjects do not work well in a scenario, then perhaps the course content proportions and sequence need to be reassessed.

Challenge 4: Time

The most consistent negative student feedback we receive is that there is insufficient time to complete the exams, and that this time pressure leads to stress and mistakes. We continue to explore solutions to this problem in several ways. We strive to remove repetitive tasks, such as calculating areas of numerous spaces, which are not necessary for assessing student ability. We have also added recommended lengths of time next to each question to help students better manage the 2 to 3 hours allocated to complete exams. Recently we experimented with a take-home final exam. Even with this format, students expressed concern that they spent too much time on it. Apparently when given multiple days to complete the problems, students spent that entire time. We did not see a drastic change in grades for the take-home exam, but we did hear that it was less intimidating and caused less anxiety

Conclusions

Course redesign is a constant for all educators, especially those teaching Architectural Technology who endeavor to present engaging and relevant content while sparking student interest in technical knowledge necessary for bringing their designs to life. Sometimes course redesign is centered on format or delivery methods. Often it is focused on the proportion, sequence, or nature of the content. Most of the time, however, assessment methods tend to remain constant: multiple-choice midterm and final exams.

As part of our course redesign efforts our teaching team questioned the benefits of conventional test-taking, both for students and instructors. Inspired by scholarship from teachers and experts in other disciplines, we considered ways that assessment could advance student learning while at the same time modelling methodologies used by architects and designers in daily practice. The vignette-exams we created emphasize

lifelong learning over memorization (formative vs summative) by asking students to apply an analyze/research/apply methodology to problem solving, a strategy that will serve them well in the rest of their education and throughout their careers.

Feedback we've gained through direct contact with students and anonymous surveys has reinforced our initial assumptions. Figure 8 shows the results from two years of student surveys that indicate a clear majority of students find value in the four stages of the vignette-testing scenario: studying prior to the exam, problem solving during the exam, using rubrics to reflect on the exam, and discussing the exam with peers and instructors. Far from conclusive, this feedback is nevertheless encouraging enough to pursue further refinements and face the challenges outlined in the body above. Our refinements will be guided by further feedback (we're currently surveying upper level student perception of the learning methods discussed here and the impact on their work) and by further research into innovative and best practices in other disciplines.

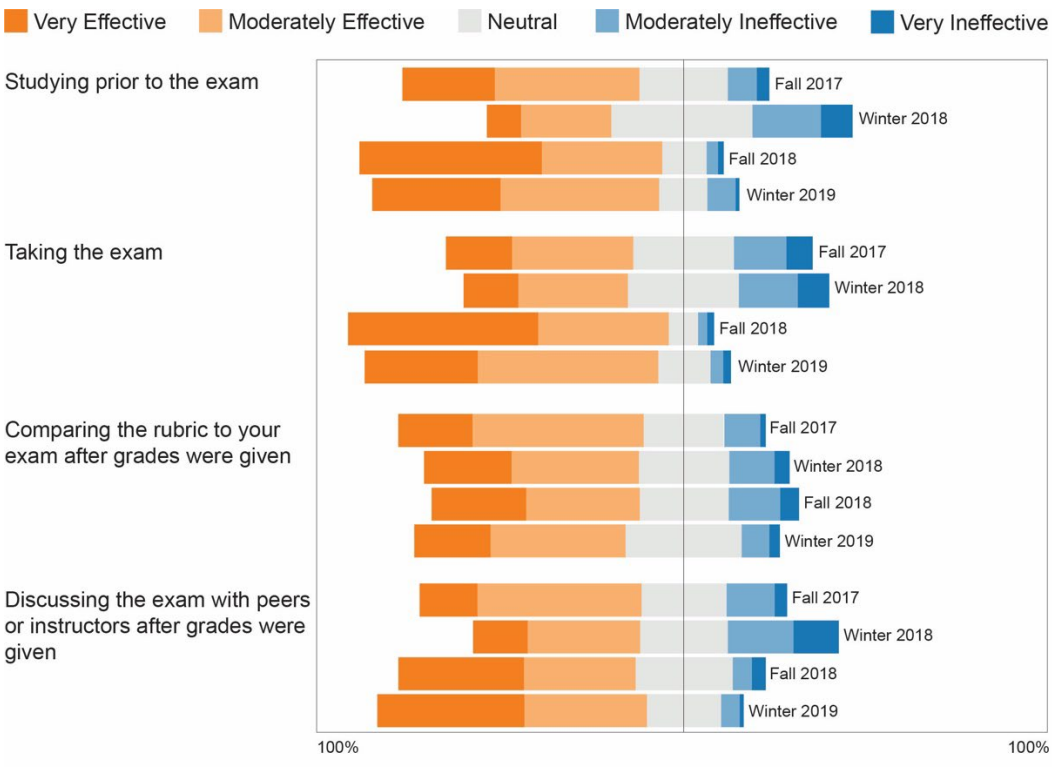

Fig. 8. Four quarters of student survey results showing that the majority of students find studying, taking, and reviewing the vignette-exams as effective in contributing to their learning. We also see improvement from the first year (fall 2017 and winter 2018) to the second year ((fall 2018 and winter 2019) indicating that our approach to exam writing is also improving.

References:

1 Mathematical Science Educational Board (MSEB) “Measuring What Counts: A Conceptual Guide for Mathematics Assessment” National Academy Press: Washington, DC. 1993

2 Fink, L. Dee. *Creating Significant Learning Experiences: An Integrated Approach to Designing College Courses*. Jossey-Bass Publishers: San Francisco, CA. 2013.

3 Fink. p 7

4 From Chapter 2, Make Exams Worth More Than the Grade, by Marsha C. Lovett in the book, *Using Reflection and Metacognition to Improve Student Learning*, edited by Matthew Kaplan, Naomi Silver, Danielle La Vaque-Manty, and Deborah Meizlish. Stylus Publishing, LLC: Sterling, VA. 2013