International Journal of Engineering Education Vol. 37, No. 3, pp. 650–662, 2021
Printed in Great Britain

0949-149X/91 $3.00+0.00

Impact of Decision-Making in Capstone Design Courses on Students' Ability to Solve Authentic Problems*

ERIC BURKHOLDER
Department of Physics, Stanford University, Stanford CA USA 94305. E-mail: eburkhol@stanford.edu

CARL WIEMAN
Department of Physics & Graduate School of Education, Stanford University, Stanford CA USA 94305. E-mail: cwieman@stanford.edu

Previous work in problem-solving has been limited by lack of a framework to describe how experts solve authentic problems, and a lack of assessments to measure authentic problem-solving. We have developed an assessment of expert problem-solving in the context of chemical process design to measure how well undergraduate engineering students are learning to solve authentic problems. We measured changes in students' problem-solving over the course of two different capstone design courses to see (1) how much students learned and (2) whether problem-solving outcomes varied between the two courses. We find that students are learning some problem-solving in capstone design courses, but not as much as one might hope: scores on most metrics of expertise range from 30–70%. Variations in what students learned between the two courses can be explained by what decisions students were given an opportunity to practice making during the course. These results suggest that undergraduate students need more deliberate practice making the decisions that expert engineers do as they solve authentic problems.

Keywords: problem-solving; decision-making; capstone design

1. Introduction

The ability to solve complex problems is central to engineering expertise. Indeed, recent graduates of engineering programs cite problem-solving as one of the most important technical skills required of them in their work [1]. While it is generally held that we train undergraduates to be good problem-solvers, there is little research to support this belief. Some work actually suggests that graduates are not prepared to solve the problems they encounter in the workplace [2].

Traditionally, students' exposure to the kind of authentic problems they might encounter as practicing engineers is limited to their experiences in the capstone design course. Capstone design courses provide "an experiential learning activity in which the analytical knowledge gained from previous courses is joined with the practice of engineering in a final, hands-on project [3–5]." These courses play a central role in accreditation of engineering programs. ABET requires that "students must be prepared for engineering practice through a curriculum culminating in a major design experience based on the knowledge and skills acquired in earlier course work and incorporating appropriate engineering standards and multiple realistic constraints." [6] The actual implementation of the capstone design experience varies widely [7].

As described by [3], the evaluation of capstone design courses is typically subjective and provides little or no hard evidence of benefits, with some exceptions [8, 9]. For example, [10] states that he is convinced of the value of capstone design courses only from his experiences with these courses. Generally, the continued corporate sponsorship of many capstone design courses and positive student surveys regarding their experience are taken as evidence for the courses achieving their desired outcomes [11]. Most engineering educators involved describe the courses as successful and worthwhile [3], though some have varying opinions on how design should be incorporated into the curriculum [12–14],

One difficulty in objectively measuring the outcomes of the capstone design courses is that the desired outcomes, e.g., problem-solving, are difficult to measure. We have been carrying out extensive research on how experts solve authentic problems and have used this to create a framework by which to capture many elements of problem-solving in terms of the decisions experts make. Using this framework, we have created an assessment that can be used in capstone design courses to test how much they teach students to make decisions like experts on numerous aspects of chemical process design. In the following study, we describe the use of an assessment of problem-solving in chemical process design to measure changes in problem-solving in two capstone design courses for chemical engineers. We sought to answer two research questions:

1. Does students' ability to solve authentic problems increase over the course of the capstone design experience?

* Accepted 4 December 2020.

2. Does the format of the capstone design course affect how much authentic problem-solving that students learn?

Though this is a case study and the applicability of the results is somewhat limited, this study reveals particular shortcomings and strengths of capstone design experiences. Importantly for instructors, it informs them of things they can do to improve students' problem-solving skills that are often overlooked.

2. Analytical Framework

Researchers in physics [15–21] and engineering [22–26] education research have studied problem-solving. This has provided essential insights into differences in how experts and novices approach problems. For example, [27] found that novices focus on surface features of physics problems, while experts focus on the conceptual structure of problems. Based on empirical and prescriptive models of problem-solving, researchers have developed methods of teaching problem-solving to undergraduate students [28–33].

A central limitation to the existing work in problem-solving is that most studies focus on experts and students as they are solving structured, textbook-style problems. While these pose a challenge to the students, experts are often able to determine the pathway to a solution in advance, and these "problems" are thus simple exercises to the expert [34–37]. Recently, researchers have begun to study experts as they solve authentic problems, such as the problems a researcher or practicing engineer would encounter on the job. These problems have conflicting goals, multiple solution methods, and multiple forms of representation [2]. Solving these problems involves making far more decisions and requires more extensive skills than solving well-structured problems typically encountered in courses [26, 38].

Another limitation of the research on problem-solving is that there are few, if any, agreed-upon assessments of authentic problem-solving. Researchers typically study problem-solving with think-aloud interviews, which are time intensive and impractical to implement in a classroom setting. A few assessments of problem-solving that do exist are typically done with "knowledge-lean" tasks such as the Tower of Hanoi problem [39]. These are able to capture general problem-solving heuristics, but do not adequately capture the role that deep disciplinary knowledge plays in expert problem-solving. Indeed, other studies have shown that experts draw heavily on their knowledge base to define the problem at hand [40–42]. Some researchers in engineering education have developed rubrics that measure problem-solving [33, 43], but they are not widely used, and the theories upon which they are based are not empirically validated.

Our research group is involved in a large-scale project that addresses both identified limitations of the problem-solving literature. First, we conducted an empirical study of expert problem-solving to determine how experts solve authentic problems [44]. This entailed conducting semi-structured interviews with experts from a range of STEM fields in which we asked the experts to describe how they solved a particular problem in their work. Interviews were based on a modification of the critical decision method of cognitive task analysis [45]. Experts were asked to particularly focus on the decisions they made as they solved the problem in question. From these interviews arose a set of 29 decisions and five additional themes that were consistent across all of the various fields. While the decisions were consistent across fields, how the experts made those decisions was field-specific and guided by a mental construct called a predictive framework which explicitly incorporates the relevant disciplinary knowledge. Predictive frameworks are mental models of a problem's key features and the relationships between them; these allow the expert scientist or engineer to explain observations, reason mechanistically about a situation, and conduct mental simulations of the problem. Harlim & Belski [46] identified some of the decisions that we identified in our work, but their list is not complete.

From this empirical model of expert problem-solving, we set out to develop assessments that capture these decision-making processes. Textbook problems are not suitable for these assessments because many of the expert decisions are made for the students. For example, physics textbooks always explicitly state assumptions for the student instead of allowing the student to make their own assumptions or simplifications. We have developed these assessments for many STEM disciplines, but will focus on the chemical process design assessment in this work

2.1 Assessment Design

The basic structure of our assessments is to present the problem-solver with a non-functioning system, such as a flawed product design schematic or a medical patient history. We then ask the problem-solver general questions which require them to make some of the expert decisions we identified and ask increasingly more detailed questions about the system as the assessment progresses [47]. This allows us to capture a wider range of problem-solving skills: experts and more expert-like students will notice important features and failures of the system in the beginning in response to general questions, whereas less expert-like students may not until specifically prompted and weak students will never notice. A detailed schematic of the assessment design is illustrated in Fig. 1.

This assessment is carefully structured to probe expert decision-making. The specific questions posed in the chemical process design assessment may be found in Fig. 1. In the assessment, the problem-solver is situated as a practicing engineering who has assigned an intern to design a process to produce tetrachloroethylene, a dry-cleaning solvent. The basic chemistry underlying the process and a table of physical properties is given to the problem-solver. The situation we present the problem-solver with is reasonably authentic: an intern has developed a block flow diagram which you, a practicing engineer, must read over to check for errors, etc. This requires a fair amount of disciplinary knowledge, but also requires the engineer to make decisions such as how well the given solution holds, what are particular areas of difficulty, etc. The engineer has to worry about high-level features such as whether mass and energy balances are obeyed, whether the process accomplishes its goals (and whether it does so efficiently), and importantly, what changes are needed to improve or fix the process.

2.2 Assessment Validation & Scoring

We have conducted several pilot studies to validate and refine the assessment so that it could be used in chemical engineering courses. In the first study, we conducted think-aloud interviews with 42 students and 3 experts (professors of practice with extensive industry and teaching experience) from two highly-ranked chemical engineering programs [47]. In that study we found that the students who had completed a capstone design course noticed 3 more errors and improvements students on average than students who had only taken an introductory course in design. Students with industry design experience noticed 7 more errors and improvements

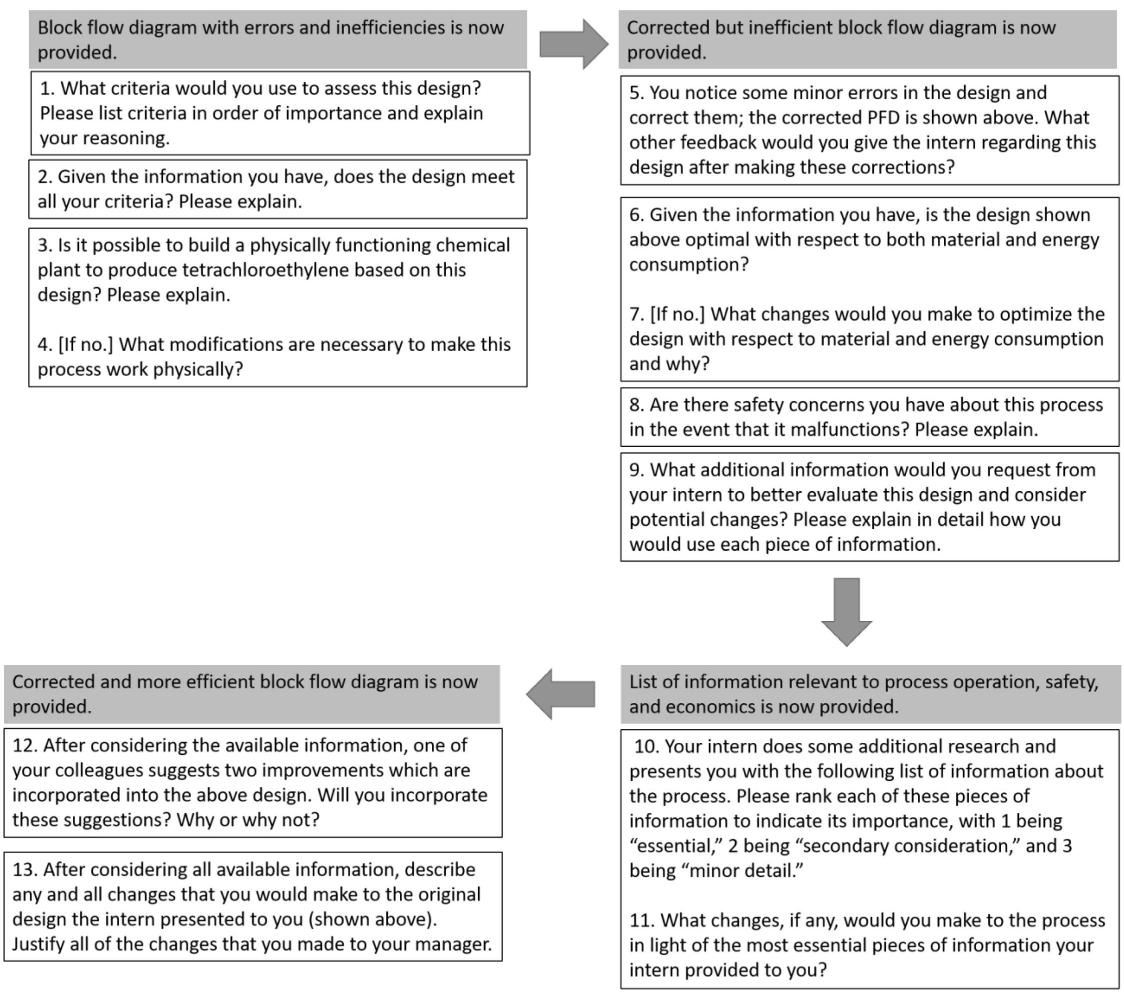

Fig. 1. Assessment design and questions.

Table 1. Summary of various sub-scores for the chemical process design assessment. All scores range from 0 to 100%.

Score	Description
Evaluation Criteria	Number of student-identified criteria that were also cited by experts divided by the total number of student-identified criteria.
Information Request	Number of student-requested pieces of information that were also requested by experts divided by the total number of student-requested pieces of information.
Information Ranking	Experts cited three pieces of information as most important. This score is the fraction of those pieces that students rank as most important.
Design Errors	Fraction of errors that students notice in the design.
Design Improvements	Fraction of improvements that students suggest to the design.
Safety	Number of student-identified safety considerations that were also cited by experts divided by the total number of student-identified safety considerations.

on average than the introductory students. We used the expert responses to establish grading rubrics for the questions regarding criteria, what information the students requested, and how the students ranked the various pieces of information.

We next conducted a study in the introductory process design course at two highly selective research universities to determine best practices as to how to administer the assessment. We found that making the assessment zero-stakes resulted in students not putting an earnest effort into their responses. As a result, we recommend making the assessment a required assignment for the course in question. The assignment is graded only for completion, but this makes it worth enough of a student's grade to get more reliable responses. We find that students take about 45-60 minutes to complete the assessment if putting in a reasonable effort.

As we have conducted further validation studies of this assessment, we have developed a grading rubric that reflects the different dimensions of expert problem-solving that it probes. Because of the inherent multi-dimensionality of the assessment, it does not have a single score, but rather 6 different sub-scores, summarized in Table 1. The rubric for scoring the assessment may be found in Ref. [48]. The first is the evaluation criteria score. This score is the number of criteria students cite that were also cited by at least one of our three experts, divided by the total number of criteria the students cite – this penalizes students who select unimportant criteria. The second score is the information request score, which is derived in a similar manner. We divide the number of pieces of information the students request that at least one of our experts requested by the total number of pieces of information the students' request. The third score is the information ranking score. When we administered the assessment to our experts, we found that 2 of 3 experts agreed on how important all pieces of information were, but all three experts cited three specific pieces of information that were essential. The information ranking score is the fraction of these three pieces of information that students rank as essential. The fourth score, design errors, is the fraction of errors the students notice in the original design – note that this is derived from the students' answers to the first four questions, before they are shown the corrected design. The fifth score, design improvements, is the fraction of material and energy inefficiencies that students notice and correct in their answers to questions 5–7. The final score is the safety score, which is derived in the same way as the evaluation criteria and information ranking scores. All scores range from 0–100%.

We note that this scoring scheme does not capture all of the student decision-making. For example, it does not allow us to identify which criteria or information students cite most often and how that compares with the evaluation criteria and information. It does not capture the results of question 11, where students have the opportunity to make further changes to the design based on the information provided to them. It also does not capture their reasoning as to whether or not they accept the optimized design presented to them. Capturing these dimensions of student reasoning requires an analysis of students' open-ended responses. While such an analysis is highly informative, and we would recommend that individual instructors do so, this is not the focus of this study. This scoring scheme captures how expert-like student reasoning is along several dimensions and is a reasonable proxy of how expert like their thinking is in general, and thus provides valuable information to instructors as to areas where instruction may be improved. This scheme is discriminating in that we see a wide range of scores on the assessment from a variety of students [49]. Furthermore, the scores that we collect are valid in that they represent how expert-like the students are in their reasoning about the process.

3. Methods

We administered the chemical process design assessment as a pre- and post-test (via Qualtrics) in two capstone design courses in the chemical engineering department at an elite private research university in the northeastern United States. We chose this particular department for our study because the students choose between two capstone design courses to take in their final semester, but the core chemical engineering curriculum up to that point is otherwise exactly the same for all students enrolled in either course. Thus, differences between students can be attributed to differences between the two capstone design courses with greater confidence.

All participants were senior undergraduate students majoring in chemical engineering. The participants were 66% female, 8% URM and 8% international students. At pre-test, we also collected data on students' undergraduate research experience, teaching (TA) experience, the number of project-based courses they had taken, whether they had industry design experience, whether they had participated in a project team, and whether they had any leadership experience. Descriptive statistics for these additional measures are presented in Table 2. The participants were divided between two courses: product design and plant design. 22 of the 24 students enrolled in product design completed the pre- and post-test compared with 30 of the 38 students enrolled in plant design. We did not analyze responses from students who did not complete both the pre – and post-test.

We find that students in plant design have significantly more research experience ($p = 0.02$) and TA experience ($p = 0.05$) than students in product design. We tested for significance using the Mann-Whitney U-test. A larger proportion of students in product design reported industry design experience (Fisher's exact test, $p = 0.10$) when compared with plant design. There were no significant differences on any of the other metrics.

3.1 Course Context

These capstone design courses are both taught by experienced engineers and share many educational objectives. These courses are the first opportunity students have to think holistically about a chemical product or plant. A primary goal of these courses is to introduce students to ambiguous and open-ended problems where there is no single correct solution. The problems have a clear goal but are otherwise vaguely defined so students need to do some investigation to determine how to achieve the goal in question. Another primary goal is to encourage students to reflect on their answers – e.g., check whether the size of a distillation column is possible or makes sense – something that is typically not emphasized in other courses in the curriculum.

There are other shared goals between the two courses. Students should become more comfortable making estimates of quantities that they need, asking questions, consulting with outside experts and many other sources to obtain information needed, using simulation software, and should develop teamwork and presentation skills. Consistent with industry standards, students are expected to think about the financial implications of all the decisions they make regarding the product or plant design.

3.2 Instruction in the Product Design Course

Product design is a 5-unit, project-based capstone design experience that simulates an industry work environment. Students work in teams of 4 to apply chemical product design principles, combined with hands-on prototyping, to advance a product concept through a Technology Readiness Level phase. The design teams apply stage-gate chemical product design principles including market and economic analysis, patent search, environmental, regulatory, and safety issues. Students provide weekly reports covering economic analysis, product specifications, experimental design, process modeling, and regulatory analysis.

The course has two one-hour lectures per week in which instructors review concepts related to the weekly deliverables. The design teams meet with their "managers" (instructor and teaching assistant) weekly to present key milestones in product development. At the mid-point and end-point of the

Table 2. Background data on the study participants. For project-based courses, research experience, TA experience, internships, problem-based interviews, and project team experience we report the median value of the distribution. For design experience and leadership experience, we report the number of students who reported these experiences

Course	Project-Based Courses	Research (months)	TA (semesters)	Internships	Problem-Based Interviews	Design Experience (students)	Project Team (semesters)	Leadership (students)
Product Design (N = 22)	1	6	0	1.5	1	10	0	19
Plant Design (N = 30)	1	10.5	1	1	1	9	0	24

semester, students present a "fresh-eyes review" in the form of a presentation and report to industrial collaborators and sponsors. All deliverables for this course are presented as a group, there were no individual assignments outside of this assessment.

The deliverables for this course cover both the technical and economic aspects of product design. Students first identify customer needs, market/data trends, and information about the competitive landscape to inform the feasibility; they write memos and give presentations summarizing this information. The students evaluate what the most important customer attributes for the design that they will be addressing and address these during prototyping. Students produce a product design that, in addition to customer attributes, accounts for engineering constraints, product packaging, supply chain, and regulatory issues. Students conduct a series of experiments to help them arrive at a prototype product design. Students conduct an economic analysis which includes market demand, raw materials cost, product costs, product price, investment and cash flow analyses, and determination of the payback period. After these studies, students design a process from scratch for how they will produce their product at scale, highlighting any unique operations/procedures and basic equipment sizing. They develop a quality control plan, production schedule, and do some preliminary modeling of their pilot-scale model to make recommendations for what needs to be done for large scale production. The students summarize all of this work with a recommendation as to the viability of their product and its large-scale production.

3.3 Instruction in the Plant Design Course

Plant design is a 5-unit, project based capstone design course that simulates an industry work environment. Students work in teams of 4 to prepare a full-scale feasibility study of a chemical process. This includes product supply and demand forecasts, development of mass and energy balances, and a process flow sheet sufficient for estimating the capital and operating costs of the facilities. Students also define all off plot support facilities and estimate associated costs. Ultimately students develop an economic analysis and provide a recommendation as to the viability of the project.

The course has two one-hour lectures per week in which instructors review concepts related to the weekly deliverables. The design teams meet with their "managers" (instructor and teaching assistant) weekly to present key milestones in process development. At the mid-point and end-point of the semester, students present a "fresh-eyes review" in the form of a presentation and report to industry experts. All deliverables for this course are presented as a group, there were no individual assignments outside of this assessment.

Deliverables in plant design address technical, logistical, and economic aspects of designing a chemical plant based on a pilot scale process design. Students start by developing a project execution plan, design basis memorandum, quality plan, and initial market analysis. They then analyze the current condition process flow and initial process capacity and constraints based on the pilot-scale process they are working with (they do not design the *process* from scratch). They analyze the economics, utility energy balances, and develop a preliminary plot plan for their chemical plant. They then work on preliminary design and sizing of equipment based on physical constraints (e.g., kinetic and thermodynamics data). They then conduct a full simulation of their process based on the equipment designs. At this stage students consider regulatory constraints (Safety, environmental, security, etc.) and optimize their simulation to minimize energy use and waste. They develop a product packaging scheme and material storage/ supply chain plan, before producing a final plant layout, quality systems analysis, start-up and shutdown plan, and final economic analysis.

3.4 Quantitative Analysis

The first author scored all of the pre- and post-tests according to the scoring scheme described in the Analytical Framework. To simultaneously examine differences between courses and changes in scores between pre and post, we used a robust linear mixed-effects model. The basis of the model is:

$$Score = \beta_0 + \beta_1\ Time + \beta_2\ Course + \beta_3\ Course \times Time + Student, \tag{1}$$

where *Score* is the score of interest, *Time* is a binary variable that is 0 at pre-test and 1 at post-test, *Course* is a binary variable that is 0 for plant design and 1 for product design, and *Student* is the random effect of student. The pre-scores for students in plant design are β_0, and the pre-scores for students in product design are $\beta_0 + \beta_2$. The post-scores for students in plant design are $\beta_0 + \beta_1$, and the post-scores for students in product design are $\beta_0 + \beta_1 + \beta_2 + \beta_3$. There is a significant change in scores for plant design students if β_1 is significant. There is a significant change in scores for product design students if $\beta_1 + \beta_3$ is significant. There is a significant difference in pre-scores between the two courses if β_2 is significant.

In addition to this base model, we also controlled for the effects of undergraduate research experience, TA experience, and industry design experience on pre-scores. We selected these variables because there were significant differences between the two courses, and thus wanted to see if any of the difference between pre-scores in the two courses could be explained by these differences. The final model is:

$$Score = \beta_0 + \beta_1\ Time + \beta_2\ Course + \beta_3\ Course \times Times + \beta_4\ Research + \beta_5 TA + \beta_6\ Industry + Student, \quad (2)$$

where *Research* is the number of months of research experience, *TA* is the number of semesters of TA experience, and *Industry* is a binary variable that is 1 when a student reports industry design experience and 0 otherwise. Using the robust lmm package in R, we estimated the regression coefficients in Equation 2 above, as well as the size of the random effect of student and the correlations between all of the fixed effects. We used robust estimation methods due to our small sample size of 52 students.

For the purposes of this paper, we define an effect to be significant if the coefficient β_i is larger than its associated standard error. There are two reasons for this definition instead of looking at p-values. First, whether p-values for fixed effects coefficients in mixed-effects models are appropriate is a subject of debate. Some suggest that instead of p-values, one should use likelihood ratio tests or look at changes in the Akaike Information Criterion (AIC) as fixed effects are added to or removed from the model [48]. Second, our statistical power is limited here by our small sample size and the relatively large number of variables in the model. Thus, it is difficult to determine whether our effect sizes are different from zero using traditional t-tests, particularly if the effect size is small. In any case, the American Statistical Association recommends against using p-values as go/no-go tests of significance [50]. By using this modified definition of significance, we believe we will identify the most educationally significant effects.

4. Results

The mean and standard error for each score in each course and pre-test and post-test are reported in Table 3. The coefficients for the mixed-effects model may be found in Table 4. We first discuss the changes in scores and differences between courses, before discussing the correlations between pre-scores and research experience, TA experience, and industry experience at the end.

4.1 Evaluation Criteria Scores

Students in product design have lower evaluation criteria scores than students in plant design at both pre- and post-test. At pre-test, the difference is 0.52 standard deviations. Students in plant design see a 0.46 standard deviation (15 percent) decrease in evaluation criteria scores from pre- to post-test. Students in product design see a comparable decrease of 0.41 standard deviations (11%) from pre- to post-test.

Students' criteria scores are decreasing not because they are mentioning fewer expert-identified evaluation criteria (pre/post difference is -0.0031 ± 0.21), but because they are mentioning more criteria overall (0.81 ± 0.28 more criteria). Because scores reflect the fraction of criteria mentioned which were

Table 3. Descriptive statistics for results of the assessment for both courses. Numbers reported are averages plus or minus the standard error

Course/Time	Evaluation Criteria	Information Request	Information Ranking	Design Errors	Design Improvements	Safety
Product Design (Pre-test)	42±5.8%	51±8.9%	61±7.2%	24±5.0%	24±4.5%	67±9.5%
Product Design (Post-test)	31±4.3%	58±8.3%	65±4.1%	26±6.6%	31±3.3%	92±5.0%
Plant Design (Pre-test)	55±4.0%	58±7.9%	64±5.0%	32±6.1%	39±3.7%	61±7.9%
Plant Design (Post-test)	40±3.9%	71±6.1%	66±4.4%	40±5.2%	42±3.2%	87±5.8%

Table 4. Table of coefficients and standard errors (in parentheses) for the linear mixed-effects models of assessment scores. The scores are scaled such that coefficients are in units of standard deviations of scores

Variable	Evaluation Criteria	Information Request	Information Ranking	Design Errors	Design Improvements	Safety
Intercept	0.55 (0.24)	–0.35 (0.29)	–0.011 (0.27)	–0.44 (0.25)	0.20 (0.28)	–0.46 (0.21)
Time (1 = Post)	–0.46 (0.24)	0.33 (0.30)	0.064 (0.28)	0.39 (0.26)	0.14 (0.19)	0.60 (0.22)
Course (1 = Product Design)	–0.52 (0.27)	–0.12 (0.34)	–0.057 (0.31)	–0.15 (0.29)	–0.87 (0.30)	0.26 (0.24)
Time × Course	0.046 (0.38)	–0.11 (0.46)	–0.016 (0.43)	–0.35 (0.40)	0.23 (0.29)	–0.11 (0.33)
Research Experience (months)	–0.0040 (0.0088)	0.013 (0.011)	–0.0077 (0.010)	0.0075 (0.0093)	–0.0097 (0.011)	0.012 (0.0078)
Teaching Experience (semesters)	–0.043 (0.056)	0.066 (0.068)	0.078 (0.063)	0.045 (0.059)	0.039 (0.071)	0.025 (0.049)
Industry Design Experience	–0.045 (0.21)	0.10 (0.24)	0.10 (0.24)	0.49 (0.22)	0.28 (0.27)	0.14 (0.19)

also mentioned by experts, this causes the scores to decrease from pre to post-test. Essentially, we are penalizing them for knowing more criteria but without acquiring judgement as to the relevance of those criteria.

4.2 Information Request Scores

Students in plant design see a significant increase in information request scores from pre to post-test (0.33 standard deviations, 13%). Students in Product design see a somewhat smaller increase in scores (0.22 standard deviations, 7%).

4.3 Information Ranking Scores

Information ranking scores at both pre-test and post-test are comparable in both the plant design and product design courses. There is no significant change in information ranking scores from pre to post-test in either plant design (0.064 standard deviations, 1%) or product design (0.048 standard deviations, 3%). The scores range from 61–65%, so this lack of change is not due to a ceiling or floor effect.

4.4 Design Errors

Students in plant design see a 0.39 standard deviation (8%) increase in the number of design errors they notice from pre- to post-test, while students in product design do not change from pre- to post-test.

4.5 Design Improvements

Students in product design suggested fewer improvements to the design than students in plant design (0.87 standard deviations at pre-test, 12%). Students in plant design did not change from pre- to post-test. Students in product design saw a 0.37 standard deviation (8%) increase in scores, which is significant.

4.6 Safety

Students in product design were more expert-like in their safety concerns about the process than product design students at pre-test (0.26 standard deviations). Plant design students saw a 0.60 standard deviation increase in safety scores, while product design students say a 0.49 standard deviation increase in scores.

4.7 Correlation between Prior Experience and Pre-score

Research experience, TA experience, and industry experience are not significantly correlated with students' pre-scores for evaluation criteria or information request. There is a small correlation between students' pre-scores and their undergraduate research experience (0.013 standard deviations per month of research experience). There is a small correlation between teaching experience and information ranking score: 0.078 standard deviation increase in scores per semester of TA experience. Students with industry design experience have significantly higher Design Errors pre-scores (0.49 standard deviations). We see that design experience is correlated with higher pre-scores on design improvements as well (0.28 standard deviations). Finally, there is a small correlation between research experience and safety scores: students see a 0.012 standard deviation increase in pre-scores per month of research experience.

5. Discussion

The most notable results indicated by Tables 3 and 4 is that students are generally receiving low scores on the assessment, and see only small improvements from pre- to post-test. Indeed the design errors and design improvements scores range from 24–42%, despite the students having significant practice in analyzing designs during the course. This suggests that students are not getting sufficient practice with reflecting on their solutions in this course, as deciding how well the chosen solution holds is a decision common to all questions regarding the design. One reason for these small gains is that this is team-based course and students frequently use a "divide and conquer" approach. This means that not all students get practice making all of the decisions necessary to complete the project. For example, a student may become an expert in safety issues, but not learn about the economic modeling aspects because a teammate was handling those issues.

Furthermore, effect sizes measuring the change over time were, with the exception of Safety, small. Though the students are improving in areas where they receive significant practice (see below), these changes are smaller than one might hope. As we discuss below, changes in students' scores and differences between the two courses are largely explained by what decisions the students were able to practice making in their respective capstone design courses. Students saw gains when given the opportunity to practice making decisions and saw no improvement when they did not get the chance to make certain decisions.

The small effect sizes suggest that students are not developing expert-level problem-solving from just one capstone design course and are far from expert problem-solvers at the end of the undergraduate curriculum. While disheartening, the results also point to a solution: for students to become better problem-solvers, they simply need more practice making the set of decisions that comprise expert problem-solving. One way to achieve this would be to restructure capstone design courses such that students are able to make more expert decisions with feedback and opportunities to revisit and improve decisions. For example, instead of being given the specific goals for their project, students should have to spend time choosing their own goals. Instead of being given a weekly schedule that chooses students' priorities for them, the students may need a more open-ended experience where they decide what to work on in a given week.

We note that we are not criticizing the instruction in these particular courses. Indeed with so few objective measures of capstone design courses, it is not immediately obvious what the problem-solving outcomes after a single authentic course *should* be. Developing expert problem-solving is not the only goal of the capstone design experience. Other outcomes of these courses include teamwork, being able to synthesize technical knowledge to solve a real-world problem and being able to deal with ambiguity. Thus, while the gains we see are small, they may be quite reasonable for a single course. However, significant changes to the undergraduate experience may be necessary if students are to become more expert-like by the end of the curriculum.

We suggest that, to improve students' problem-solving further, a different philosophical approach must be taken toward teaching problem-solving in the undergraduate curriculum. Rather than worrying about specific forms of pedagogy in the capstone design course, or whether the projects are simulated or industry sponsored, focus should be placed on whether students are being allowed deliberate practice in making the same decisions as an expert engineer would. Such practice would ideally be distributed throughout the curriculum, such that students more closely resemble expert engineers by the end of the undergraduate curriculum. This requires a fundamental change to the way we teach engineering science courses like thermodynamics and fluid mechanics. Rather than focusing solely on content, students must be given the opportunity to use that content knowledge to make expert decisions so that they can become more expert-like by the end of their undergraduate careers. Indeed other researchers have seen such results in courses such as heat transfer [51].

5.1 Population Differences Reflected in Pre-Scores

From inspection of Table 3, one can see that product design pre-scores (and thus, post-scores) are lower than plant design pre-scores across all metrics. From the mixed-effects modeling in Table 4, we see that this difference is not explained by differences in industry, research, or teaching experience. We posit that this difference in pre-scores may be explained by a difference in GPA. Students in plant design have higher GPAs on average (Cohen's $d = 0.40$, $p = 0.06$). While GPA is not expected to be a perfect predictor of how well students perform on this assessment of real-world problem-solving, it seems plausible that there would be a correlation between GPA at this level, as it reflects many engineering courses, most of which are common to both groups, and problem-solving performance.

In analyzing the data we also noticed two weak, but significant, correlations: a correlation between information request scores and research experience, and a correlation between information ranking scores and teaching experience. We cannot say with certainty what the cause of these correlations is, but it seems plausible that students engaged in research and teaching get more practice making the respective decisions required to answer the two questions. For example, ranking information requires students to decide upon priorities, decide whether information is valid/reliable, and decide whether information matches their expectations. These are all decisions involved in the grading process. A TA must make a rubric for a given problem (decide upon priorities through points allocation), decide whether information matches their expectations (whether the student solutions match their solutions), and decide upon the quality of the information presented to them (how good a student's solution is). Meanwhile, students with substantial research experience may get more practice deciding what information is needed to solve the problem, whether they have enough information, and whether their problem-solving approach is working. An analysis of decision making by students engaged in undergraduate research would be needed to confirm this explanation. Indeed, [52] find that undergraduate research experiences require students to engage in decision making regarding criteria for suitable experimental evidence, designing of experiments and testing the experimental design, and analyzing and presenting data. Deciding upon criteria for suitable evidence requires students to identify what information is needed to solve the problem and whether they need more information, and troubleshooting an experimental design requires students to decide how well their problem-solving approach is working.

The correlations between undergraduate research experience and TA experience also suggest ways instructors might improve problem-solving. For example, having students create rubrics to assess each other's deliverables and grade those deliverables would involve them in the same kind of decision making as a TA. The results suggest that their information ranking scores would improve if they practiced making those decisions.

5.2 Changes that are Consistent across the Two Courses

There are several notable consistencies across the two courses. The first is that we see evaluation criteria scores decrease over time. This decrease occurs not because students are mentioning fewer evaluation criteria (1.5 on average at pre- and post-test), but because they are mentioning more criteria overall (0.81 more at post-test), making the fraction of evaluation criteria decrease. This is a reflection of the students' uneven progression toward expertise. During the capstone design course they learn about all the elements that go into doing a feasibility study for a chemical process or product, and all the criteria on which feasibility is determined. However, they have not learned to judge the importance of appropriate criteria. Similar results have been seen in the medical clinical reasoning assessment our group has developed and applied to medical students. When given a list of tests to order for a patient, experienced physicians choose tests sparingly and deliberately, while students list everything they know is possible [53].

The second notable finding is that there was no change in information ranking scores for either course over time: scores range from 61–65% so this is not due to a ceiling or floor effect on scores. We expect that this is because students are not engaged in deliberate practice in making the expert decisions necessary to decide which information is most important. As outlined in the methods section, deciding which information is most important relies upon three decisions: (1) deciding upon priorities, (2) deciding if information is valid, reliable, etc., and (3) deciding how information collected compares with expectations derived from the predictive framework. In both of these courses, students do not have to spend much time deciding upon priorities. Which aspects of the product or process design they work on from week-to-week is determined by the instructors, thus students do not get to make that decision.

It is interesting that, despite the fact that students in both courses need to seek out substantial amounts of information from online and text sources, students do not seem to get practice deciding whether information is valid/reliable and comparing information to expectations. We suspect that this reflects deficiencies in students' predictive frameworks. Their predictive frameworks are not sophisticated enough for them to make predictions about what the information they gather will look like. Because they do not have well-developed expectations, they also don't spend much time thinking about whether the information they gather is valid or reliable. Most students will be able to do some superficial evaluation of whether a specific *source* of information is reliable, but not the specific pieces of information. This is consistent with their broad failure to reflect on their solutions during the assessment.

Another notable result is the substantial increase in safety scores for students in both courses. The capstone design course is the first experience students have considering issues of safety in engineering design; thus this result is not unexpected. What is surprising is that the safety scores are so high at pre-test (67% and 60% in product and plant design, respectively) given the fact that students typically don't spend much time considering issues of safety before the capstone design experience. This may be because the range of possible answers as to safety concerns is more limited than the range of possible criteria or information. Generally respondents (including experts) were concerned only with issues related to toxicity and environmental impacts of reactants and high temperatures or runaway reactions.

5.3 Changes that are Different in the Two Courses

Students in plant design become more expert like in the information that they request during the assessment, whereas product design students do not change. Students in both courses are given the opportunity to practice deciding what information is needed and whether they have enough information to solve the problem. One plausible explanation is that how these decisions are made is highly context-dependent – i.e. dependent on having the appropriate predictive frameworks. Students in plant design spend time developing predictive frameworks that allow them to assess and design processes, whereas student in product design spend time developing frameworks that allow them to think about the feasibility of chemical products. Students in plant design thus get more specific practice in deciding what information is relevant to the design of a process, whereas students in product design get less practice in this area.

On a related note, students in plant design get significantly better at noticing errors in the original design presented to them, whereas students in product design do not. We expect that this is because students in plant design spend significantly more time working with process flow diagrams compared with the product design students. But curiously, this does not also result in plant design students becoming better at suggesting *improvements* to the design, while product design students do. We expect that this is an idiosyncrasy associated with the projects the students complete. In the plant design course, while students have to develop a full process flow diagram, the design of the process itself is often known in advance – indeed, in industry processes are often developed and then licensed, rather than being redesigned by each company which wants to make a certain product. Thus, students do not spend much time thinking about conceptual ways to improve the design, they are only concerned with numerical optimization. Product design students, on the other hand, are conceiving of a novel chemical product, for which a design may not exist. They may thus get more practice thinking about conceptual building blocks of the process and focus less on the numerical optimization.

5.4 Limitations and Future Work

There are several notable limitations to this work. The first concerns the reliability of students' written answers to the assessment. It is known that think-aloud interviews provide far more detail into student and expert reasoning than what is written down. We may therefore be missing certain aspects of student reasoning that would affect the response coding and assessment scoring. However, from the think aloud interviews that we conducted, students' thoughts about the process were generally not any more revealing than their written responses [47].

The second limitation is that we had a small, relatively select sample population. Whether patterns in decision-making are similar across capstone design courses at other institutions is largely unknown. We are currently analyzing student responses from a public, teaching-intensive institution to compare with the results from this private, research-intensive institution. Preliminary results suggest only minor differences between the institutions. Furthermore, the small number of students (22 in product design and 30 in plant design), limit our statistical power. However, we have enough students to show that these courses are not making a large impact on students' preparation to be expert problem solvers, despite this being the central focus of those courses. Furthermore, we have enough data to show specifically where the two courses can be improved.

The primary focus of this work was to investigate (1) whether students' problem-solving improved during the capstone design courses and (2) whether this improvement was different for different types of capstone design courses. To do this, we converted qualitative data (student responses) into quantitative variables so that we could determine the significance of various differences between courses and over time. We specifically investigate how students' criteria, information requested, etc. differs from an expert engineer in order to better assess the predictive frameworks that students graduating from an undergraduate chemical engineering program have.

One area related to both limitations and future work concerns the open-response nature of the assessment. In past work, we found that students needed significant motivation to put an earnest effort into completing our assessment, particularly at the end of the course. This resulted in our decision to make the assessment a required assignment in the capstone design courses. It is possible that we did not see greater increases in students' problem-solving skills because students did not put their full effort into the post-assessment. That being said, from the written responses, it does not seem as though students put any less effort into the post-test compared to the pre-test. Students wrote approximately the same amount for their responses at pre – and post-test. Furthermore, we found no significant correlation between time spent on the assessment and any of the assessment sub-scores.

We are currently consulting a wider range of students and experts in order to develop a closed-response assessment. A closed-response version would be simpler to score and more easily used in courses where time is already a limited resources. Once the closed response version is developed and widely distributed, further analysis such as item response modeling can be conducted to empirically verify the central decisions probed by each question.

6. Conclusions

Our research questions concerned whether students' problem-solving improved during the capstone design course, and whether the format of the course affected how much students did or did not learn. We found that students saw gains in problem-solving when they were given the opportunity to practice making key decisions and saw no improvement when they did not get the chance to make a certain decisions. Indeed, the differences between courses also stemmed from students getting practice making different decisions in the two courses.

References

1. H. J. Passow, Which abet competencies do engineering graduates find most important in their work? *Journal of Engineering Education*, **101**(1), pp. 95–118, 2012.
2. D. H. Jonassen, J. Strobel and C. B. Lee, Everyday problem solving in engineering: Lessons for engineering educators, *Journal of Engineering Education*, **95**(2), pp. 1–14, 2006.

3. A. J. Dutson, R. H. Todd, S. P. Magleby and C. D. Sorensen, A review of literature on teaching engineering Design through project-oriented capstone courses, *Journal of Engineering Education*, **86**(1), pp. 17–28, 1997.
4. G. Pomberger, Software engineering education – Adjusting our sails, *Education and Computing*, **8**(4), pp. 287–94, 1993.
5. J. F. Thorpe, Design of mechanical systems: A capstone design course in mechanical engineering design, *Proceedings, 1984 ASEE Annual Conference*, pp. 803–7. ASEE, 1984.
6. ABET, Criteria for Accrediting Engineering Programs 2019–2020, 2019.
7. J. Pembridge and M. C. Paretti, Characterizing capstone design teaching: A functional taxonomy, *Journal of Engineering Education*, **108**(2), pp. 197–219, 2016.
8. E. W. Banios, An engineering practices course, *IEEE Transactions on Education*, **35**, pp. 286–93, 1992.
9. K. Worthington, The evaluation of past industrial engineering projects by undergraduate engineering economy students, *Proceedings, 1979 Asee Annual Conference*, pp. 130–32, ASEE, 1979.
10. R. C. Born, A capstone design experience for electrical engineers, *IEEE Transactions on Education*, **35**, pp. 240–42, 1992.
11. J. R. Phillips and M. M. Gilkeson, Reflections on a clinical approach to engineering design, *Design Theory and Methodology*, **31**, pp. 1–5, 1991.
12. R. S. Culver, D. Woods and P. Fitch, Gaining professional expertise through design activities, *Engineering Education*, no. July/August, pp. 533–36, 1990.
13. D. L. Evans, B. W. McNeill and G. C. Beakley, Design in engineering education: Past views of future directions, *Engineering Education*, no. July/August, pp. 517–22, 1990.
14. B. V. Koen, Toward a strategy for teaching engineering design, *Engineering Education*, **83**(3), pp. 193–201, 1994.
15. J. I. Heller and F. Reif, Prescribing effective human problem-solving processes: Problem description in physics, *Cognition and Instruction*, **1**, p. 177, 1984.
16. P. Heller, R. Keith and S. Anderson, Teaching problem solving through cooperative grouping part 1: Group versus individual problem solving, *American Journal of Physics*, **60**, p. 627, 1992.
17. P. Heller and M. Hollabaugh, Teaching problem solving through cooperative grouping part 2: Designing problems and structuring groups, *American Journal of Physics*, **60**(7), pp. 637–44, 1992.
18. J. Larkin, J. McDermoott, D. P. Simon and H. A. Simon, Expert and novice performance in solving physics problems, *Science*, **208**(4450), pp. 1335–42, 1980.
19. W. K. Adams and C. E. Wieman, Analyzing the many skills involved in solving complex physics problems, *American Journal of Physics*, **83**, p. 459, 2015.
20. C. E. Wieman, Comparative cognitive task analyses of experimental science and instructional laboratory courses, *The Physics Teacher*, **53**, p. 349, 2015.
21. L. Hsu, E. Brewe, T. M. Foster and K. A. Harper, Resource letter rps-1: Research in problem solving, *American Journal of Physics*, **72**, p. 1147, 2004.
22. D. H. Jonassen, Engineers as Problem Solvers, *Cambridge Handbook of Engineering Education Research*, Cambridge University Press, Cambridge, pp. 103–18, 2014.
23. D. Jonassen, Design for decision making, *Educational Technology: Research and Development*, **60**, pp. 341–59, 2012.
24. C. Atman, R. S. Adams, M. E. Cardella, J. Turns, S. Mosborg and J. Saleem, Engineering design processes: A comparison of students and expert practitioners, *Journal of Engineering Education*, **96**(4), pp. 359–79, 2007.
25. D. Woods, An evidence-based strategy for problem solving, *Journal of Engineering Education*, pp. 443–59, 2000.
26. N. S. Hong, D. H. Jonassen and S. McGee, Predictors of well-structured and ill-structured problem solving in an astronomy simulation, *Journal of Research in Science Teaching*, **40**(1), pp. 6–33, 2003.
27. M. T. H. Chi, P. J. Feltovich and R. Glaser, Categorization and representation of physics problems by experts and novices, *Cognitive Science* **5**, pp. 121–52, 1981.
28. K. Heller and P. J. Heller, *Competent Problem Solver – Calculus Version*, McGraw-Hill, New York, 2000.
29. J. L. Docktor, J. Dornfeld, E. Frodermannm, K. Heller, L. Hsu, K. A. Jackson, A. Mason, Q. X. Ryan and J. Yang, Assessing student written problem solutions: A problem-solving rubric with application to introductory physics, *Physical Review Physics Education Research*, **4**, 010111, 2016.
30. E. W. Burkholder, J. K. Miles, T. J. Layden, K. D. Wang, A. V. Fritz and C. E. Wieman, Template for teaching and assessment of problem-solving in introductory physics, *Physical Review Physics Education Research*, **16**(1), 010123, 2020.
31. G. Cabodi, P. Camurati, P. Pasini, D. Patti and D. Vendraminetto, Schema-based instruction with enumerative combinatorics and recursion to develop computer engineering students' problem-solving skills, *Int. J. Eng. Educ.*, **36**(5), pp. 1505–1528, 2020.
32. S. J. Kimmel, J. S. Kimmel and F. P. Deek, The common skills of problem solving: From program development to engineering design, *Int. J. Eng. Educ*, **19**(6), pp. 810–817, 2003.
33. N. J. Mourtos, Challenges students face in solving open-ended problems, *Int. J. Eng. Educ.*, **26**(4), pp. 846–859, 2010.
34. K. Duncker and L. S. Lees, On problem-solving, *Psychological Monogaphs*, **58**(5), pp. *i*–113, 1945.
35. A. Newell and H. A. Simon, *Human Problem Solving*, **104**(9), Prentice-Hall, New Jersey, 1972.
36. G. Polya, *How to Solve It: A New Aspect of Mathematical Method*, Princeton University Press, New Jersey, 2004.
37. F. Reif, *Understanding Basic Mechanics*, Wiley-VCH, Weinheim, Germany,1995.
38. B. Means, E. Salas, B. Crandall and T. O. Jacobs, Training decision makers for the real world, *Decision Making in Action: Models and Methods*, Ablex, Norwood, NJ, pp. 306–26, 1993.
39. D. E. Egan and J. G. Greeno. *Knowledge and Cognition*, Lawrence Earlbaum, Oxford, England, 1974.
40. S. Wineburg, Reading Abraham Lincoln: An expert/expert study in the interpretation of historical texts, *Cognitive Science*, **22**, pp. 319–46, 1998.
41. C. Singh, When physical intuition fails, *American Journal of Physics*, **70**, pp. 1103–9, 2002.
42. H. A. Schoenfeld, *Mathematical Problem Solving*, Elsevier, 1985.
43. A. Mendez, M. Florensa, C. Molleda, C. Fernandez, A. Alazcar, J. B. Ramiro-Diaz, E. Sardonil, E. Manrique and T. Montoro, Development of a method of assessment of the problem-solving competency at the Technical University of Madrid, *Int. J. Eng. Educ.*, **30**(6), pp. 1749–1758, 2014.

44. A. M. Price, C. Kim, E. W. Burkholder, A. V. Fritz and C. E. Wieman, A universal structure of science and engineering problem solving, arXiv:2005.11463 [physics.ed-ph], 2020
45. B. Crandall, G. A. Klein, and R. Hoffman, *Working Minds: A Practioner's Guide to Cognitive Task Analysis*, MIT Press, Cambridge, Massachusetts, 2006.
46. J. Harlim and I. Belski, Long term innovative problem solving skills: Redefining problem solving, *Int. J. Eng. Educ*., **29**(2), pp. 280–290, 2013.
47. E. W. Burkholder, L. Y. Hwang and C. E. Wieman, Evaluating the problem-solving skills of graduating chemical engineering students, *Ed. For Chem. Eng.*, **34**, pp. 68–77, 2020.
48. E. W. Burkholder, A. M. Price, M. P. Flynn and C. E. Wieman, Assessing Problem Solving in Science and Engineering Programs, *2019 Perc Proceedings*, Provo, Utah, 2019.
49. S. G. Luke, Evaluating significance in linear mixed-effects models in R, *Behavior Research Methods*, **49**, pp. 1494–1502, 2017.
50. R. L. Wasserstein and N. A. Lazar, 2016, The Asa's statement on p-values: Context, process, and purpose, *The American Statistician*, **70**, pp. 129–33, 2016.
51. K. E. Cook, Y. L. Han, T. R. Shuman and G. Mason, Effects of integrating authentic engineering problem centered learning on student problem-solving, *Int. J. Eng. Educ.*, **33**(1), pp. 272–282, 2017.
52. N. G. Holmes and C. E. Wieman, Examining and contrasting the cognitive activities engaged in undergraduate research experiences and lab courses, *Physical Review Physics Education Research*, **12**(2), 020103, 2016.
53. C. J. Kim, A. M. Price, C. E. Wieman and S. F. Chen, Evaluating the diagnostic process using a decision-making framework, 2021.

Eric Burkholder is a postdoctoral scholar and lecturer in the Department of Physics at Stanford University. He studies a range of issues in engineering and science education, including problem-solving in chemical engineering curricula. He has a PhD in chemical engineering from the California Institute of Technology, were he studied the physics of soft active matter.

Carl Wieman is a professor in the Department of Physics and Graduate School of Education at Stanford University. He has studied issues of teaching and learning in physics and other disciplines for several decades, including the founding of PhET and the founding of the Science Education Initiative at the University of British Columbia.