Formative feedback and scaffolding for developing complex problem solving and modelling outcomes

Frank, Brian; Simper, Natalie; Kaupp, Jake

Published in:
European Journal of Engineering Education

DOI:
10.1080/03043797.2017.1299692

Formative feedback and scaffolding for developing complex problem solving and modeling outcomes

Brian Frank[1], Natalie Simper[2], Jake Kaupp[1]

[1]*Faculty of Engineering and Applied Science*

[2]*Office of the Provost*

Queen's University

Kingston, ON K7L 3N6 Canada

Corresponding author:

Brian Frank

Faculty of Engineering and Applied Science

Queen's University

Kingston, ON K7L 3N6 Canada

brian.frank@queensu.ca

Word count: 6768

Formative feedback and scaffolding for developing complex problem solving and modeling outcomes

This paper discusses the use and impact of formative feedback and scaffolding to develop outcomes for complex problem solving in a required first-year course in engineering design and practice at a medium-sized research-intensive Canadian university. In 2010, the course began to use team-based, complex, open-ended contextualized problems to develop problem-solving, communications, teamwork, modeling, and professional skills. Since then, formative feedback has been incorporated into: task and process level feedback on scaffolded tasks in-class, formative assignments, and post-assignment review. Development in complex problem solving and modeling has been assessed through analysis of responses from student surveys, direct criterion-referenced assessment of course outcomes from 2013-2015, and an external longitudinal study. The findings suggest that students are improving in outcomes related to complex problem solving over the duration of the course. Most notably, the addition of new feedback and scaffolding coincided with improved student performance.

Keywords: formative feedback; scaffolding; complex problem solving; modelling; assessment

1. Introduction

Complex problem solving, involving addressing ill-structured problems where necessary information is not known a priori and the solution path is not visible at the start (Jonassen 2010), is fundamental to higher education and in engineering specifically. It is an essential requirement for engineers outlined in the Washington Accord ("Graduate Attributes and Professional Competencies" 2013), and it is also a fundamental expectation in engineering education (Jonassen 2014). Society and the workforce have shifted such that ill-structured problem solving, and interactive problem solving skills are considered core requirements, and emphasized in frameworks of fundamental competencies ("Employability Skills Framework - Source Matrix" 2016; Griffin and Care 2015). These types of problems are a key element in modern tests of critical thinking and problem solving and are designed to go beyond the tests of formal logic that were the norm several decades ago (Pascarella and Terenzini 2005; Greiff, Holt, and Funke 2013; Zahner 2013; Stein and Haynes 2011).

The vast majority of professions, including engineering, involve non-routine cognitive and interpersonal tasks (Autor and Price 2013; Casner-Lotto and Barrington 2006;

Engineers Canada and Canadian Council of Technicians and Technologists 2009); successful task completion requires complex problem solving skills. Practice of engineering case studies show that collaboration, complexity, ambiguity, and contradictions are the present norm (Trevelyan 2014; Williams 2002), rather than the more traditionally valued closed-ended problem solving.

One of the distinguishing characteristics of engineering activities is the extensive use of models within problem solving processes (Jonassen 2014). A significant fraction of engineering instruction is spent on the effective use of mathematical models (with, generally, less emphasis on the development and evaluation of those models).

In order to develop the ability to solve complex problems, students need to be exposed to them repeatedly (Sweller, Ayres, and Kalyuga 2011). When assigned ill-structured problems students are often uncomfortable with a "lack of guidance"; the necessary information and process is not provided from the beginning, as it is with the more traditional closed-ended problems. Developing competence in complex interconnected activities requires early and frequent practice, as argued by Sheppard (Sheppard et al. 2009):

> Developing the expertise of professional practice is an iterative process. Thus the ideal learning trajectory is a spiral, with all components revisited at increasing levels of sophistication and interconnection. Learning in one area supports learning in another.

Studies on the development of expertise demonstrate the importance of a period of relevant deliberate practice with feedback and self-reflection (Davidson and Sternberg 2003). Educational studies and cognitive psychology have also highlighted the importance of practice, self-testing, and feedback (Dunlosky et al. 2013; Hake 1998; Hattie 2008). Formative feedback, the information communicated to the learner that is intended to modify his or her thinking or behavior for the purpose of improving learning (Shute 2008), can have significant positive impact on learning. Specific approaches that use formative feedback include peer instruction (Hake 1998; Crouch and Mazur 2001) and computer-based formative feedback tools (Jiao 2015; Kleij, Feskens, and Eggen 2015).

Not all feedback is formative. Large-scale syntheses and meta-analyses (Kluger and DeNisi 1996; Shute 2008; Hattie 2008) highlight that feedback can actually have negative effects on learning, and that it is most effective when directed at specific tasks that focus on specific goals with low task complexity and when perceived threats to self-esteem are low (Kluger and DeNisi 1996). Hattie's Visible Learning model emphasizes the importance of both feedback to the instructor about student performance and formative feedback to students from the instructor to improve their performance (Hattie 2008). Hattie's model presents four levels of feedback, although he argues only the first three are usually effective:

- Task level feedback (how well tasks are understood)
- Process level feedback (the process needed to understand and perform tasks)
- Self-regulation level feedback (self-monitoring, directing, and regulating actions)
- Self level feedback (personal evaluations and effect (usually positive) on the learner)

A useful approach to ensure that feedback is provided on challenging goals is to scaffold complex tasks. The instructor (or another student) helps a student bridge the gap between current abilities and intended goal using either tools or techniques (Hammer and Green 2011; O'donnell, Dansereau, and Hall 2002; Rosenshine and Meister 1992; Beed, Hawkins, and Roller 1991). The scaffolding should be in what Vygotsky called the zone of proximal development (ZPD), where students can only proceed with appropriate scaffolding (Rosenshine and Meister 1992). As student competency increases, the scaffolding is gradually removed, allowing them to independently complete tasks that previously required assistance. Formative feedback itself can act as a scaffold for complex tasks, facilitating learning and improved performance (Shute 2008).

In the context of complex problem solving, scaffolding can involve instructors or fellow students modeling the process of breaking down large problems into smaller tasks or steps and applying general principles. This guided practice helps students to organize and connect the knowledge (Ambrose et al. 2010), to reduce the cognitive load and complexity of the problem, and to receive feedback on the small tasks or steps prior to addressing the larger complex problem. First hand experience of guided practice is critical for the development of expertise (Davidson and Sternberg 2003). Scaffolding effectively develops skills such as argumentation that are critical to solving complex problems (Cho and Jonassen 2002), and develops cognitive procedures for collaborative problem solving in engineering (Zheng et al. 2014; Zheng and Cao 2015). As students develop competence, their need for scaffolding is reduced, and they can address problems independently.

In large classes, providing rich and individualized feedback is is particularly difficult for ill-structured tasks that require open-text responses. Scaffolded computer-supported tasks can be quite useful (Cho and Jonassen 2002; Zheng and Cao 2015) but the initial set up is often time-consuming. The limited available time for grading and feedback in large university courses forces instructors to allocate their resources to tasks where the feedback is most effective. Identifying the most effective ways in which to apply feedback and what types of scaffolding work best in a constrained environment is often an iterative process; e.g. (Hammer and Green 2011). This paper focuses on the use of deliberate practice and feedback for developing complex problem solving skills.

2. **Constructs and Instructional Context**

Complex problem solving (CPS) has been the subject of much interest in psychology, general higher education, and engineering education specifically. Jonassen argues that problems vary in five ways: structuredness, context, complexity, dynamicity, and domain specificity (Jonassen 2010). Problem solvers have to acquire and apply knowledge about the complex scenario's structure in order to reach their goal. This may require information reduction, causal learning via interaction, hypothesis testing, dynamic decision making, and self- and task-monitoring (Greiff et al. 2015).

In the engineering context CPS often includes the application of engineering knowledge, abstract thinking, and the solving of many component sub-problems. Particularly when they involve wide-ranging conflicting issues and diverse groups of stakeholders, there is often no obvious solution (International Engineering Alliance 2013). Table 1 shows some of the dimensions associated with some measures of complex problem solving that will be discussed in more detail below.

Table 1: Dimensions in problem solving in recent assessments and frameworks.

Course approach	**Stages of the creative process (Sawyer 2011)**	**Creative problem solving (Isaksen, Dorval, and Treffinger 2010)**	**Valid Assessment of Learning in Undergraduate Education: Problem solving (Rhodes and Finley 2013)**	**Critical thinking assessment test (Stein and Haynes 2011)**
Problem definition	Find the problem	Framing problems	defining problem	separating relevant from irrelevant information
Information	Acquire the knowledge	Exploring data	identify strategies	integrate information to solve problems
Model and analyze	Gather related information		propose solutions/hypot hesis	learning and apply new information
Generate and compare ideas	Incubation	Constructing opportunities		
	Generate ideas	Generating ideas		
	Combine ideas	Developing solutions	evaluate potential solutions	
Implement	Select the best ideas		implement solution	use mathematical skills to solve real-world problems
Evaluate			evaluate outcomes	
Communicate	Externalize ideas	Building acceptance		
Review and learn				

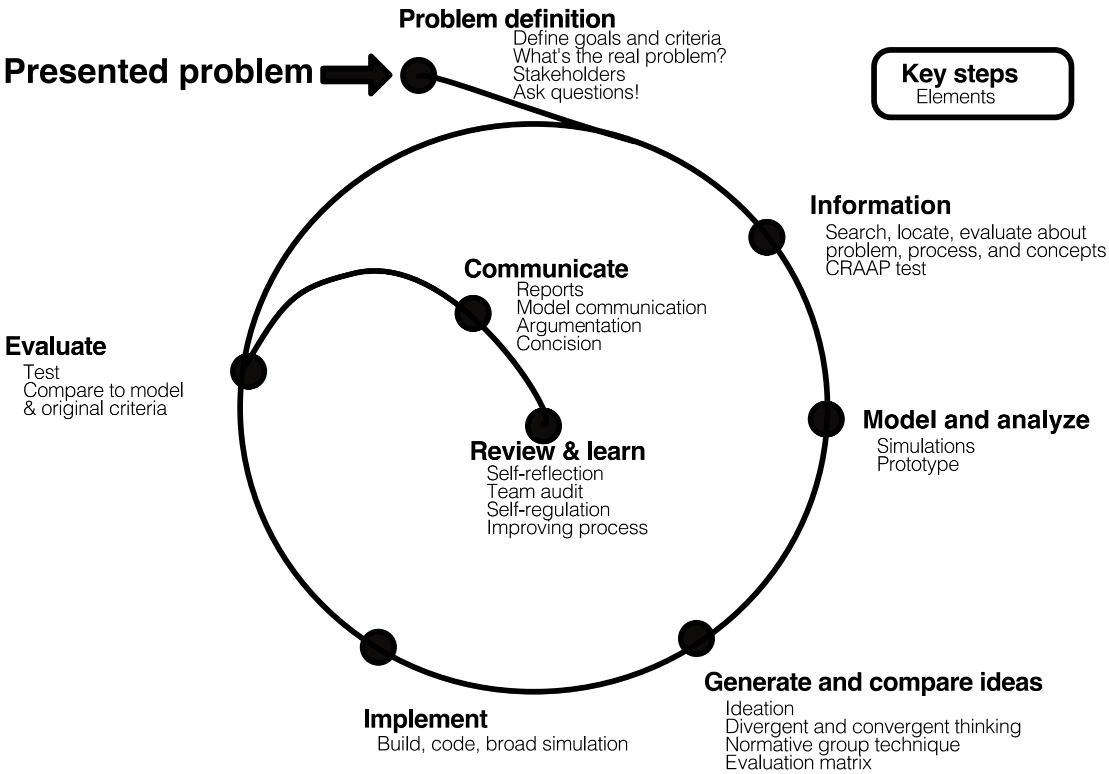
This study is set in a compulsory first-year course at Queen's University, a mid-sized, research-intensive, Canadian institution. The course is the first in a four-year sequence of engineering design and practice courses that use engineering projects to develop complex problem solving, design, and professional (Frank et al. 2013) skills based on the rationale discussed above, particularly Sheppard et al's recommendation of a professional spine (Sheppard, Macatangay, Colby, & Sullivan, 2009). The full-year course has approximately 750 students and provides an introduction to engineering practice, including complex problem solving, engineering design, and laboratory experimentation. The course is broken into three modules, the first of which runs from September to December and focuses on complex problem solving in an engineering context.

Modeling and analysis are often key elements of engineering problem solving (Zawojewski, Diefes-Dux, and Bowman 2008); they are a key focus of this module. The module uses multi-week authentic tasks known as *model-eliciting activities* (MEAs) (Zawojewski, Diefes-Dux, and Bowman 2008; Diefes-Dux et al. 2012) that require students to develop models of real engineering systems while solving a contextualized problem. The expectations are a challenge to student teams but problems are heavily scaffolded to push students to develop higher-order thinking (Kaupp and Frank 2013).

This module prepares students for real client-driven projects in the second semester, where they work in teams of five with a client from the local community. Open-ended design problems are integrated into the curriculum each year as part of the faculty wide Engineering Design and Practice Sequence.

Returning to the first module of the first year course, each week addresses a particular topic relevant to the course process for complex problem solving illustrated in Figure 1. This process has been customized to suit course needs from commonly used complex problem solving and design processes (Sawyer 2011; Dym et al. 2005; Isaksen, Dorval, and Treffinger 2010). Before class, students are assigned a reading related to that week's concept (e.g. problem definition, information evaluation, etc.). In class, students work in teams of three or four on a task relevant to the week's topic, often using a cloud-based documentation collaboration system like Google Drive (Frank and Hamgini 2014).

[Figure 1 here]



For example: in one week, the student teams work in class to analyze elements of argumentation in an article that they can use as an information source for their MEA. In another week, they evaluate a range of information sources that will be useful for their MEA. The instructor observes the students working on the small constrained tasks that will eventually contribute to the MEAs, and provides feedback both to individual teams and to the class as a whole. Over the first six weeks of the course, that are set in the context of the first of two MEAs, MEA1, students teams work on:

- forming teams defining roles, creating team contracts, and resolving conflict
- defining problems (asking questions, identifying needed information, identifying required functions and attributes, identifying stakeholders)
- evaluating a range of information sources needed for MEA1
- developing and critiquing models of physical systems
- analyzing the elements of argumentation in an article used for MEA1
- writing an engineering report

Initially, most students in the program struggle with the assigned open-ended complex problems, not due to fundamental inability to perform the necessary calculations, but rather due to unfamiliarity with complex problem solving. The stages and cognitive processes used to solve these problems are fundamentally different from the highly structured and closed-ended word problems that have been prominent in their previous educational experience. The MEAs are carefully constructed such that the assigned tasks are in Vygotsky's zone of proximal development (ZPD); students would have difficulty completing them on their own but can with the provided scaffolding.

The MEAs require students to define the presented problem, identify the information needs, evaluate the quality of the available information sources, model the problem, draw supported conclusions using effective arguments, and communicate the work. Here is an example of the problem presented in an MEA:

> In this scenario your team works for *[organization name]*, specializing in energy distribution. The Toronto power failure in April 2014, caused by arcing between transmission lines and the local utility lines, made many municipalities concerned about transmission line sagging. Consequently, the local city council put out a request for proposal (RFP) for an analysis of transmission line sag over a range of expected conditions for the weather norms and extremes of the selected place of interest. *[[organization name]* responded to the RFP, and won the contract with the city. Specific requests from the municipality are in the letter from the city council, document 1 *(not shown here)*. *[Organization name]*'s engineering director, Reginald Grover, has assigned the project to your team as described in document 2 *(not shown here)*. Some background information is also attached below, and presented as part of the in-studio activities.

2.1 Feedback and scaffolding to develop complex problem solving ability

In the module, students are guided through the mathematical expressions required to model cable sag, but need to make assumptions and approximations to identify the values for the equations. For the first couple of years, the course was delivered in large lecture halls, and primary feedback came in response to the MEAs and was directed at multiple outcomes in a relatively large report. This meant that students did not have an opportunity to practice and receive feedback on individual components (e.g. problem definition, information evaluation) before tackling the main assignment. Students improved, but student feedback, instructor and TA observations, and analysis of student performance led to course changes. Since 2013 two key elements have been used to support improved development of complex problem solving: (a) more formative feedback, and (b) more scaffolding that target specific challenging concepts and procedures. The course development over time is summarized in Table 2.

Table 2: Assessment and scaffolding in the course from 2010-2015

Year	**Summative assessments**	**Feedback & formative assessments**	**Scaffolding**
2010	Three MEAs	Written feedback on MEAs	Overall process
2011	Three MEAs	Written feedback on MEAs	Overall process
2012	Three MEAs	Same as 2011, plus technical document writing assignment, in class web-based response system	Information evaluation, argumentation
2013	Two MEAs	Same as 2012 plus pre-MEA1 mini-assignment and pre-class quiz	Overall process, problem solving, information evaluation, modeling, argumentation
2014	Two MEAs	Feedback on teaming activities in active learning classroom, pre-class quizzes, short team deliverable, post-MEA1 feedback	Same as 2013, but more structure, examples
2015	Two MEAs	Same as 2014 but adding individual scaffolded deliverable before MEA1, more detailed post-MEA1 feedback	New overall process illustrated in Figure 1, more explicit focus on communicating a model

Each year the amount of feedback and scaffolding increased or was clarified based on student performance the year before. In the 2015 academic year, the tasks were as described in Table 3, aligned with Hattie's feedback framework:

Table 3: Feedback mode and level.

Week	**Task**	**Feedback mode**	**Feedback Level**
1-12	**Weekly in-class tasks** in active learning room of 140 students	Oral to groups and class	Task, process
2-5,7-11	**Pre-class quiz** on weekly reading	Written	Process
1	**Technical report writing assignment**, focusing on technical report formatting and data analysis	Written, oral upon request	Process
5	**Individual progress report** focusing on modeling, information evaluation, and argumentation	Written, oral upon request	Task, process

Week	Task	Feedback mode	Feedback Level
7	**MEA1** - first integrative summative report by teams	Written, oral upon request	Task, process, self-regulation
9	**Idea generation and decision making** report by teams	Written, oral upon request	Task, process
12	**MEA2 -** last integrative summative report by teams	Written, oral upon request	Task, process, self-regulation

The in-class feedback targets both task-level (how well tasks are understood) and process-level (the process needed to understand and perform tasks). Written comments, written at both the task and process levels, are provided on deliverables; students are encouraged to discuss the feedback with the instructor or grader. Furthermore, the MEAs require students to include a self-assessment section that describes their process for addressing the task, critically self-evaluates their report, recommends a grade, and describes how their performance could be improved. Students receive written feedback by graduate teaching assistants directed at task and process levels on the main body of the report, and at the self-regulation level on their self-evaluation section.

2.2 Feedback and scaffolding in the most recent course delivery

The annual changes to feedback and scaffolding, shown in Table 2, have led to the most recent course delivery structure (used in 2015), shown in Table 4. The steps in the complex problem solving process shown in Table 1 and Figure 1 are broken into small steps adapted from a range of sources.

Table 4: Scaffolding structure used to develop key elements of complex problem solving.

Element	**Scaffolded structure**
Problem definition	Students required to *identify stakeholders, functional requirements, questions to be asked, information needs*
Information evaluation (Meriam Library, California State University, Chico 2016)	Students required to evaluate information on five criteria: *Currency, Relevance, Accuracy, Authority, Purpose*
Developing and communicating a mathematical model, using elements from (Quarteroni 2009)	Students required to include 5 elements: 1. Description of real world system 2. Assumptions/approximations in making mathematical model 3. Input and output of the model 4. Relate mathematical output to the real system 5. Evaluate the model compared to other models or real data
Argumentation, adapted from (Toulmin 2003)	Students required to address 5 elements of an argument: *Claim, justification, evidence, qualifier, rebuttal*

In 2015 a highly scaffolded formative assignment - creating an individual progress report on MEA1 – was added as preparation for the broader MEA1 problem. Students are required to explicitly use the provided structure. These are the scaffolded formative assignment instructions:

This assignment is set in the context of MEA1. Your team works for *[Organization name]* Consulting, specializing in energy distribution, which responded to the RFP from a local municipality.*[Organization name]*'s engineering director, Reginald Grover, has assigned the project to your team, and has asked for a short 1-page progress report on the modeling and analysis from each team member.

In your progress report, ensure you do the following:

- Select as the topic something you have started modelling (e.g. cable tension, sag, etc.).
- Model that, using all five elements of a model described in week 4 (*the real system, model description, model output, relate model output back to real system, and evaluate model*)
- In your report cite 1-2 sources underlying the equations or data. Support your use of these sources by five elements of the CRAAP test discussed in week 3 (*currency, relevance, authority, accuracy, and purpose*)
- In your final evaluation of the model, ensure you make a strong supported argument discussed in week 4 (recall this includes *claims, evidence, justification, qualifier, and rebuttal*)

Once you have finished, *indicate clearly in the left margin of the document where to find the key elements of models, the CRAAP test, and elements of argumentation*.

As can be seen in the last instruction, students are required to explicitly identify *where* their response meets all required elements. This provides a simple approach for self-evaluation, and allows graduate teaching assistants to more easily identify required elements and thus provide feedback faster. This is a low-weight formative assignment, so the teaching personnel emphasize the benefit of putting significant effort into the assignment to allow quality feedback before they start the large summative assignment (MEA1).

3. Impact

The question to be answered is: does the addition of new feedback and scaffolding make a difference in student learning? Showing impact of any initiative in one course on the development of overall complex problem solving skills is challenging, as students are engaging with these concepts in multiple courses throughout their program. However, the approaches described here do seem to support skill development, as indicated by evidence from three sources:

- student self-reported learning
- criterion-referenced scores on course learning outcomes assessed within the course

- criterion-referenced scores on program-wide learning outcomes assessed outside the course

If formative feedback and scaffolding help develop the ability to solve complex problems, then student self-reports and changes in course scores should reflect that. Two questions will be evaluated:

Q1: Do students self-report learning gains resulting from the course and from feedback?
Q2: Is there a significant change in student course performance after new formative feedback and scaffolding are added?
Q3: Does a student cohort demonstrate improved performance over time?

The sections below address these questions. Because students have multiple courses and extracurricular activities, the impact of other courses on complex problem solving cannot be discounted. However, this is the only course that both specifically targets, and devotes significant time to, developing complex problem solving ability. Furthermore, in the 2013-2014 academic year students were asked to rank the first year courses that most helped develop critical thinking and problem solving, and they overwhelmingly listed this course as the most significant - more than all other first year courses combined.

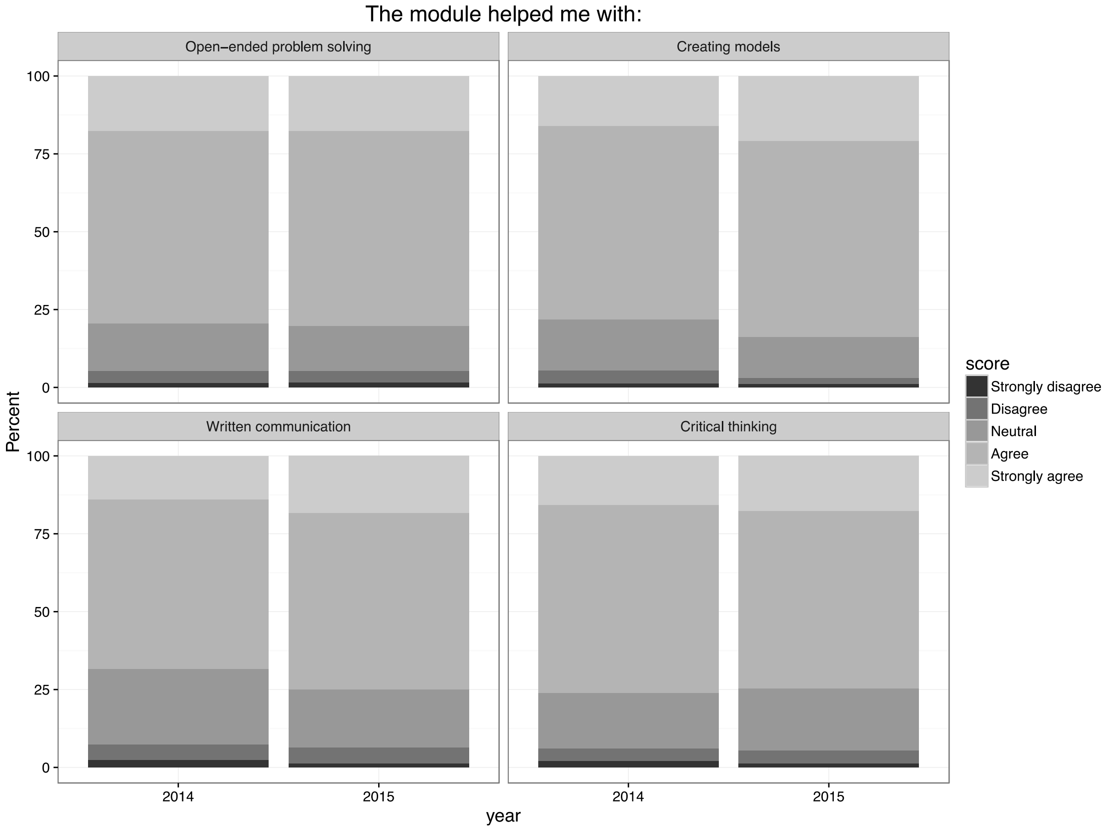
Question 1: Do students self-report learning gains resulting from the course and from feedback?

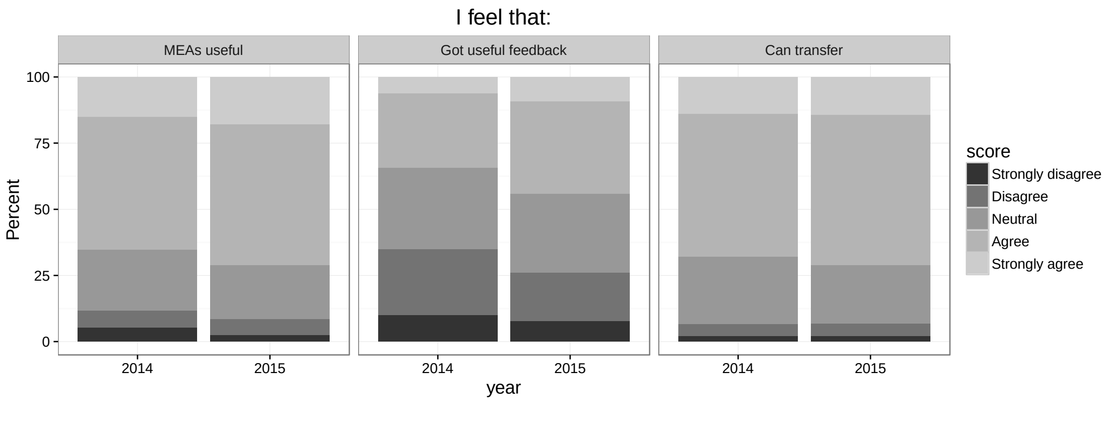
There are two sources of student data about the value of feedback and scaffolding in developing complex problem solving skills.

a) In the 2013-2014 academic year students were asked to complete a survey that asked about the value of the feedback provided in Google Drive activities (Frank and Hamgini 2014). Approximately 55% of the class responded, and of those slightly less than half of the students agreed or strongly agreed that feedback from the instructor in class on assignments completed in Google Drive helped them learn. Approximately 31% of the students said they had technological problems with using Google Docs in class, suggesting that some of the dissatisfaction with in-class assignments was the result of technological issues, such as difficulties sharing their documents with their teammates, and network connectivity problems. Another factor that year was that these activities took place in a standard tiered lecture hall, which is not an ideal space for group work. The instructor could not move in between the rows, making it very challenging to answer questions or provide feedback to groups.

b) In the 2014-2015 and 2015-2016 academic years students were asked to complete an end-of-term survey that asked them to self-report learning gains and rate the value of course activities. Figure 2 shows the results from these two years in response to questions like: “This module helped improve my ability to…”.

[Figure 2 goes here]





As can be seen, overall students feel the module helped them develop fundamental skills, with a slight improvement between 2014 and 2015. The students were also asked to provide feedback on specific approaches in the course, including the use of MEAs, whether they received useful feedback, and whether they felt they could transfer skills to other areas. The weakest area is on receiving useful feedback, where a significant fraction of the students are neutral or disagree that they received useful feedback. This reflects the challenging nature of providing feedback on the complex problem solving activities where there is no single solution to which responses can be compared.

Question 2: Is there a significant change in student course performance after new formative feedback and scaffolding are added?

Since the course was first developed, the MEAs have been scored using analytic rubrics on multiple outcomes, which vary somewhat year to year but do contain consistent expectations. Student characteristics have been quite stable over the three years in the study, as shown in Table 5. Table 6 shows the outcomes assessed over the last three years on the first and second MEA; the outcomes have been relatively consistent, with minor changes as the tasks change over time.

Table 5: Demographics of 2013-2015 cohorts.

		2013		2014		2015	
		n-mean	%	n-mean	%	n-mean	%
Gender	Male	503	72.5%	514	71.1%	458	68.2%
	Female	191	27.5%	209	28.9%	214	31.8%
Status	Domestic	671	96.7%	693	95.9%	639	95.1%
	International	23	3.3%	30	4.1%	33	4.9%
Entry Age	Male	18.1		18.0		18.1	
	Female	18.0		18.1		18.0	

Table 6: Outcomes assessed on MEAs each year, either as a team or individually.

OUTCOME ⊙ Team ◆ Individual	2013		2014		2015	
	MEA1	MEA2	MEA1	MEA2	MEA1	MEA2
Effective summary	⊙	⊙	⊙	⊙	⊙	⊙
Problem definition	⊙	⊙	⊙	⊙	⊙	⊙
Information evaluation			⊙	⊙		
Propose new process/product	⊙	⊙				
Create model	⊙	⊙	⊙	⊙	◆	⊙
Idea generation and selection		⊙		⊙		
Draw conclusions	◆	◆	◆	◆	◆⊙	⊙
Perform self-assessment	◆	◆	◆	◆	⊙	⊙
Effective argumentation	◆	◆	◆	◆		
Communication	⊙	◆⊙	⊙	◆⊙	⊙	⊙
Ethics		⊙		⊙		◆
Consider social, environmental factors						⊙
Economic analysis		⊙		⊙		⊙

MEA1 is generally submitted around week 7 of the 12-week semester, and MEA2 around week 11 or 12. Each outcome is scored out of 8, with specific descriptors on each outcome for scores of 0-3, 4, 5, 6, and 7-8. An example of the scoring rubric is shown in Table 7. These descriptors are changed slightly each year to address the requirements of that year's MEAs.

Table 7: Portion of a rubric used to score performance on course outcomes in the MEAs.

Outcome	**7-8 Outstanding**	**6 Expectation**	**5 Developing**	**4 Marginal**	**0-3 Not Demonstrated**
Defining a problem	**Meets expectations and:** Includes quality information from sources outside the course to inform process, model, and conclusions.	Clearly defines scope of problem, stakeholders, and required goals. Summarizes and assesses credibility of information used.	Problem definition is clear but missing some elements.	Some important information or biases not identified, or trivial/incorrect information included.	Problem not defined, little useful information, or information directly copied.
Creating a model	**Meets expectations and:** Sophisticated model used incorporating several effects; uncertainty in model's input variables shown by range of output values.	Creates and applies quantitative model using supported analysis, approximations and assumptions.	Model has minor errors or unsupported approximations or assumptions.	Model used has significant errors or uses inappropriate assumptions.	No analysis, or model/analysis selected is inappropriate, or can't draw conclusions.

Student scores on individual outcomes can be compared each year, and total scores on all outcomes. Figure 3 and Table 8 show distributions of the mean scores of all outcomes for all students (scored out of 8) for each of the last three years on MEA1 and MEA2. As can be seen, in each of the last three years the mean score has increased, and in the last two years' scores increased between MEA1 and MEA2. The course instructor, expectations, and the calibration and training for scorers has remained relatively constant.

[Figure 3 here]

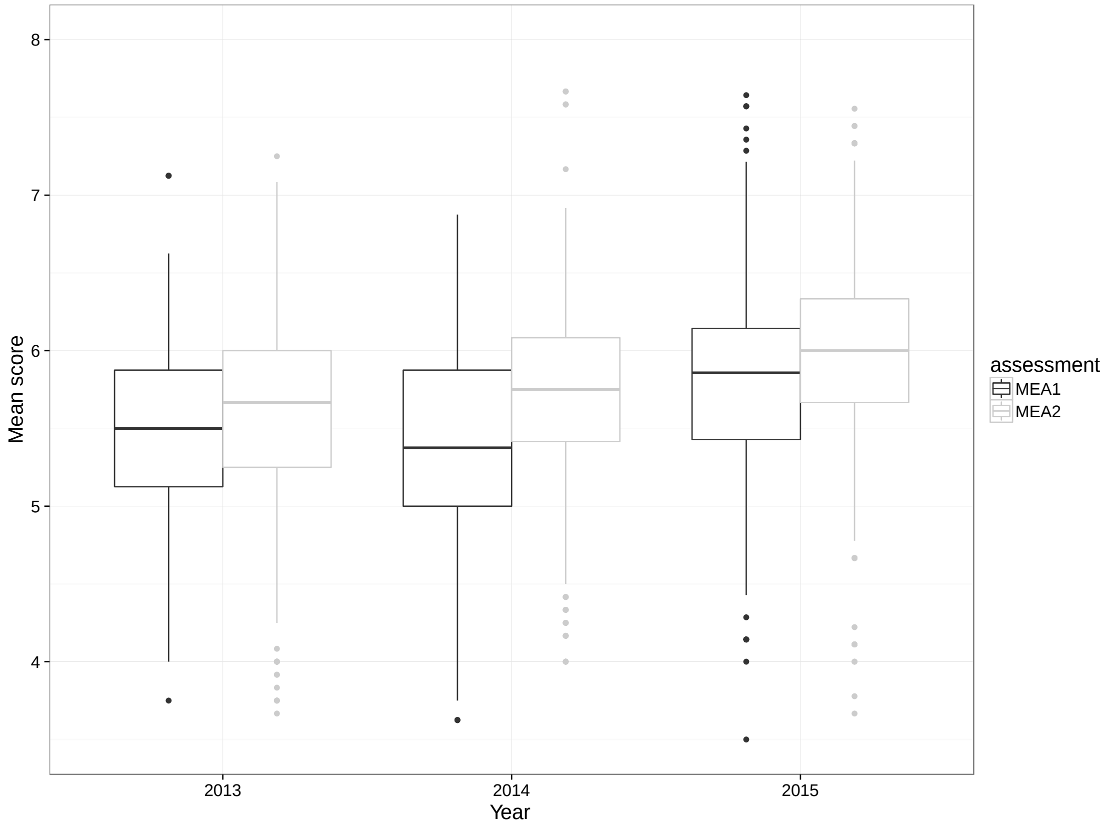


Table 8: Descriptive statistics of performance (count, mean, and standard deviation) on each MEA from 2013 to 2015.

	MEA1			**MEA2**		
Semester	**n**	**mean**	**sd**	**n**	**mean**	**sd**
Fall 2013	732	5.52	0.53	725	5.58	0.58
Fall 2014	773	5.36	0.71	769	5.74	0.54
Fall 2015	794	5.74	0.68	792	6.0	0.58

In 2014, when the course was moved into an active learning space and more feedback was provided in class, there was some improvement seen in MEA2 scores, but not

MEA1. The biggest increase in score was in the 2015 year when the heavily scaffolded formative assignment was added before MEA1. This data is consistent with the graders' impressions; the graders who had been involved in the 2014 and 2015 years commented that they felt performance on MEA1 was better in 2015 due to the addition of the scaffolded formative assignment.

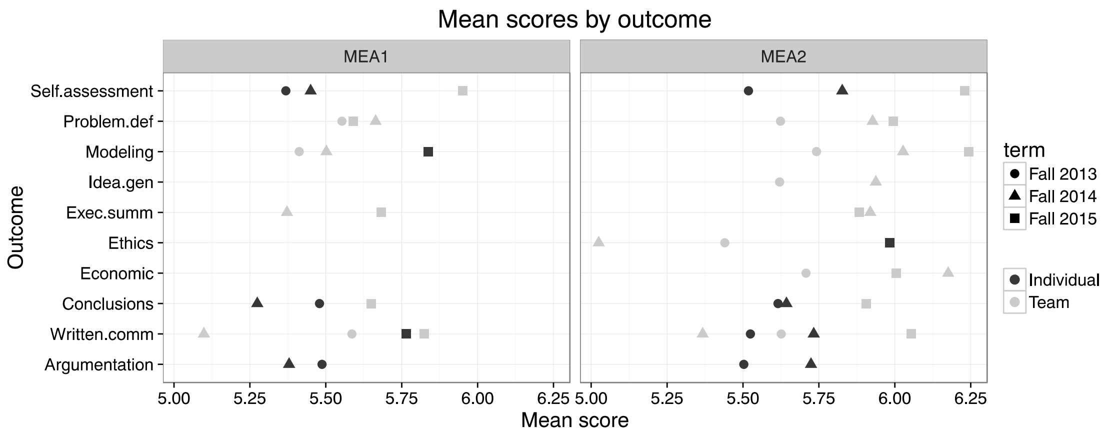
The key point is that the score changes between years (e.g. 2013 to 2015), and between MEAs within a year, are almost all significant and meaningful. The score distributions exhibited significant non-homogeneity of variance using Levene's test, so Welsh's F and non-parametric tests were used to analyze the data. The changes between terms are significant; using a one-way independent ANOVA with Welsh's F for MEA1 scores, $F(2,1518.8)=59.6$, $p<0.001$, and for MEA2 $F(2, 1513.6) = 102.3$, $p< 0.001$. Effect sizes between 2013 and 2015 using pooled SD are $d=0.33$ standard deviations, and $d=0.71$ for MEA1 and MEA2, respectively. In 2015, student performance was significantly higher on MEA2 than on MEA1, $p<0.001$ with an effect size of $d=0.4$. A summary of effect sizes is shown in Table 9.

Table 9: Summary of effect size (d) gains significant at p<0.001.

Effect size between years for:	**2013 to 2014**	**2014 to 2015**	**2013 to 2015**
MEA1		0.53	0.33
MEA2	0.29	0.46	0.71
Effect size within year:	**2013**	**2014**	**2015**
MEA1 compared to MEA2	0.13	0.58	0.4

Mean performance on individual outcomes like problem definition, modeling, and self-assessment can also be compared, as shown below in Figure 4. Student performance on all outcomes by Fall 2015 was uniformly improved over 2014 and 2013; particularly large changes occurred in self-assessment and modelling. However, in Fall 2015 self-assessment was done as a team whereas it was assessed individually in previous years, and since team activities tend to lead to higher performance (Cohen, 1994) it is not appropriate to compare performance on this outcome. In contrast to this, modelling was done as a team in 2013 and 2014 and changed to individual in 2015, so the net effect of changing from a team assessment to individual would be expected to be negative and the measured effect size should under-estimate the actual effect. There is a significant difference between score on modelling on MEA1 in Fall 2013 (median=5) vs. Fall 2015 (median=6), $W=213040$, $p<0.001$, with an effect size of $d=0.46$, and also significant between MEA1 and MEA2 in the 2015 academic year, $p<0.001$, $d=0.42$. This is encouraging as one of the primary goals of adding the scaffolded formative assignment in fall 2015 was to improve student ability to model systems in MEA1.

[Figure 4 here]



Question 3: Does a student cohort demonstrate improved performance over time?

The previous section showed that there is a significant increase in student scores over the duration of the module. However, that comparison relies on different tasks with slightly different rubrics, so conclusions drawn from that data alone are weak. However, stronger conclusions may be drawn from a four-year longitudinal study involving many students in the 2013 cohort that is examining development of transferable outcomes including problem solving, critical thinking, and written communications. One of the measures was the Critical Thinking Assessment Test. It is a 60-minute paper-based test developed by researchers at Tennessee Technological University, with funding from the National Science Foundation. The test comprises a series of questions, with problem-based real-world scenarios, and was designed to assess critical thinking, creative thinking, as well as non-routine problem solving and effective communication (Stein, Haynes, and Redding 2016).

There were 118 consenting students who completed the CAT at the beginning of first year, and again toward the end of second year. The results in Figure 5 and Table 10 demonstrate a significant improvement ($t(118)=3.12$, $p=.002$) between first and second year, suggesting that the students improved in the above listed skills as assessed on the CAT. A confounding issue with using standardized low-stakes tests to measure development is a lack of motivation that can develop when writing the same test repeatedly, particularly as students move from the first to upper years of the program.

Another approach used in this longitudinal study involves scoring signature student academic activities in each year of the students' program. Graders external to the course used the Valid Assessment of Learning in Undergraduate Education (VALUE) Rubrics developed by the American Association of Colleges and Universities. These discipline-neutral rubrics have four levels of performance criteria, from the benchmark level of a student entering university to the capstone level of a student who has just completed their undergraduate experience. One of these rubrics focuses on problem solving; the dimensions are listed in Table 10. There has been considerable effort in establishing the validity and reliability of conclusions drawn using the VALUE rubrics (Finley 2012; Rhodes and Finley 2013).

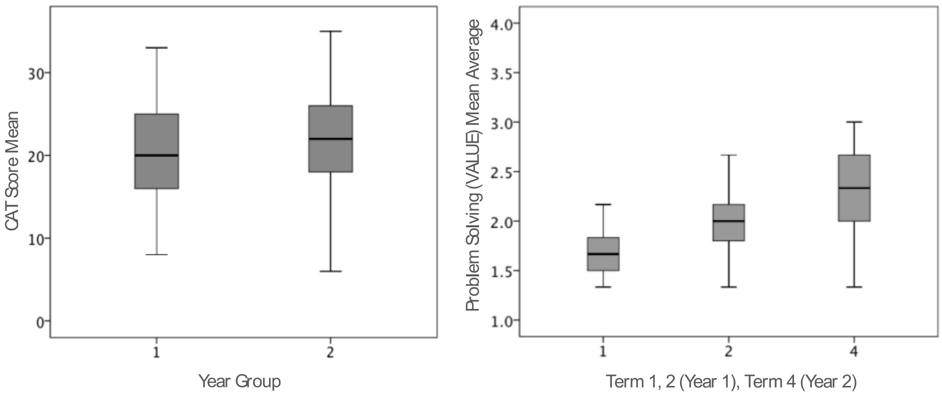
Trained research assistants used the VALUE rubrics to score the student work samples. The process involved independent ratings for each work sample, followed by calibration to come to consensus. Fifty-two Semester 1 group reports (MEA 2's) were scored (contributed from 150 consenting students). The weakest dimensions were "implement solutions", and "evaluate outcomes", with mean scores of 1.1, and 1.3 respectively (see Table 10). The rubric descriptions for level one on these dimensions describe implementation in a manner that does not directly address the problem, and superficial review of results with no consideration of need for further work. Twenty-four Semester 2 group reports were scored (contributed from 147 consenting students), and with much larger group sizes in the second year course, there were fifteen Semester 4 group reports scored (contributed from 200 consenting students). A one-way multivariate analysis of variance was conducted, and found significant gains in all of the dimensions from semester one through to semester 4 (Wilks' Lambda = .37, $F(12, 461) = 49.0$, $p < .001$). The multivariate effect size was estimated at .39, which implies that 39% of the variance in was accounted for by educational semester. The Cohen's *d* effect sizes in Table 10 display the standardized differences between the Semester 1 and 4 means. Gains in excess of 1.6 standard deviations were found for the dimensions of "solution/ hypotheses", "implement solutions", and "evaluate outcomes". Score distributions are shown in Figure 5.

Table 10: First and second year score means and standard deviations (SD) on dimensions of the Critical Thinking Assessment Test and VALUE rubric scores.

Instrument	Dimension	First Year						Second year			Effect size *d* *(Term. 1-4)*
		Term 1			Term 2			Term 4			
		n	Mean	SD	n	Mean	SD	n	Mean	SD	
CAT	Score total	162	20.3	5.42				349	21.6	5.44	.24
Problem Solving VALUE rubric	Define problem	150	1.8	.49	112	2.0	.39	200	2.2	.67	.68
	Identify strategies		2.0	.14		2.1	.41		2.3	.51	.80
	Solution/ hypotheses		2.0	.16		2.2	.56		2.7	.51	1.85
	Evaluate solution		1.6	.60		1.9	.51		2.2	.77	.87
	Implement solution		1.1	.55		2.0	.81		2.2	.79	1.62
	Evaluate outcomes		1.3	.50		1.7	.59		2.5	.64	2.09

Note: Cohen's d calculated using Term 1 Mean- Term 4 Mean / pooled SD

[Figure 5 here]



4. Conclusions

Through iteration and observation the delivery a first year engineering course seeking to develop complex problem solving skills has effectively expanded its use of formative feedback and scaffolding. In 2014, the course was moved from a large lecture hall into an active learning space that allowed the teaching personnel to circulate among the students during active lectures and provide feedback on scaffolded tasks. This had a small effect on the outcomes measured by the final summative team report (MEA2), with no significant change in the first task (MEA1), and a medium effect size between MEAs. In 2015, a short scaffolded formative individual assignment was added prior to the first major summative team report (MEA1). This provided the opportunity for students to receive detailed feedback directed at the task and process level, and more detailed written feedback provided on MEA1 directed at task, process, and self-regulation levels. There were significant gains between mean scores on MEA1 and MEA2, and significant improvement in both MEA1 and MEA2 compared to previous years. By the final large assignment (MEA2) in 2015, the median score corresponded to meeting the course expectations for most outcomes on the analytic rubric. The scaffolded formative assignment has proven particularly effective at teaching students how to create a model. Feedback provided on a scaffolded formative assessment appears to be particularly useful to student learning as measured on outcomes measured from MEA1 submitted several weeks later, particularly the elements of communicating an effective model.

There are some limitations to the degree that the scores can be attributed to course changes. It is also possible that the expectations of graders changed from year to year, and there were minor changes in rubric descriptors. Taking all of this into account, it is still noteworthy that in all three years the outcome scores increase from MEA1 to MEA2, with the largest effect size in 2015 of d=0.4, using paired student scores when graders and calibration are consistent. It is also possible that variations in latent traits between years (e.g. general intelligence, self-regulation, work ethic) impacted performance. However, changes in the overall class grade point average (GPA) between years on the order of a few percent, much smaller than the effect size observed in score improvements. Additionally, larger gains (d=0.42) are seen on the outcome of creating a model after adding a formative activity that directly targets this outcome, and students self-report learning gains. The most significant confound arises from the changes between MEAs themselves; the tasks must be changed every couple of years so that students are assigned a problem for which many solutions are not easily available from past students.

The measured effect sizes here fit in the range of learning gains found in the literature. Arum's longitudinal study of critical thinking and written communication measured an effect size of d=0.18 over three semesters (Arum, Roska, and Cho 2011), and a meta-analysis by Huber predicts an 80% confidence interval for critical thinking gains of -0.44 to 0.44 over a single semester (Huber and Kuncel 2015). Pascarella & Terrenzini's synthesis describe gains in post-formal reasoning from d=0.65 to 1.37 over two years (Pascarella and Terenzini 2005). In this study there was no measurement of ability at the beginning of the course; the reported effect size in 2015 of d=0.4 represents measurement points that span only half of the 12-week semester, so actual gains measured from the beginning to end of the course could be reasonably expected to be higher.

Supporting evidence arises from the ongoing longitudinal study of the 2013 cohort that is finding that there are significant increases in performance of problem solving from first to second year when academic work rescored externally using program-wide rubrics. Large gains were seen in student's presentation of solutions and hypotheses; the demonstrated VALUE rubric criteria suggested that by the end of second year, on average, students were able to demonstrate comprehension of the problem that was sensitive to contextual factors such as ethical, logical, or cultural dimensions. The largest gains however were seen in the student's demonstration of *evaluation of outcomes*, with more than two standard deviation differences between Term one and four.

Hattie's massive synthesis concluded that the effect size of feedback (described at the beginning of this paper) has a mean of d=0.73 (Hattie 2008). Kluger's meta-analysis of feedback interventions concluded that:

> a [feedback intervention] provided for a familiar task, containing cues that support learning, attracting attention to feedback-standard discrepancies at the task level, and is void of cues … that direct attention to the self is likely to yield impressive gains in performance, possibly exceeding 1 SD. (Kluger and DeNisi 1996)

The effect size between measured performance from 2013 to 2015 as measured by course outcome scores is on the low end compared to these studies, but when consistent raters used the VALUE rubrics to evaluate problem solving, the effect sizes were well in excess of the (Hattie 2008) results. Student comments about feedback suggests there is still room for improvement, and Shute (2008) has some promising ideas, such as providing feedback on small tasks without grades presented by computer or in written form, with content directly linked to learning goals.

An approach being piloted this year is assigning a scaffolded task (on information evaluation, argumentation, modeling, etc.) on the learning management system that requires students to provide a short response, then read a prepared high quality response, and finally write a self-reflection on how they could have improved their response.

Acknowledgements

This work was supported by the DuPont Canada Chair in Engineering Education Research and Development, and the Higher Education Quality Council of Ontario through the Learning Outcomes Assessment Consortium. The longitudinal study reported here was approved by the Queen's University General Research Ethics Board.

References

Ambrose, Susan A., Michael W. Bridges, Michele DiPietro, Marsha C. Lovett, and Marie K. Norman. 2010. *How Learning Works: Seven Research-Based Principles for Smart Teaching*. John Wiley & Sons.

Arum, Richard, Josipa Roska, and Esther Cho. 2011. "Improving Undergraduate Learning: Findings and Policy Recommendations from the SSRC-CLA Longitudinal Project." Social Science Research Council. http://www.ssrc.org/publications/view/D06178BE-3823-E011-ADEF-001CC477EC84/.

Autor, D. H., and Brendan Price. 2013. "The Changing Task Composition of the US Labor Market: An Update of Autor, Levy, and Murnane (2003)." *Unpublished Manuscript*.

Beed, Penny L., E. Marie Hawkins, and Cathy M. Roller. 1991. "Moving Learners toward Independence: The Power of Scaffolded Instruction." *The Reading Teacher* 44 (9): 648–55.

Casner-Lotto, Jill, and Linda Barrington. 2006. *Are They Really Ready to Work? Employers' Perspectives on the Basic Knowledge and Applied Skills of New Entrants to the 21st Century U.S. Workforce*. Partnership for 21st Century Skills. http://eric.ed.gov/?id=ED519465.

Cho, Kyoo-Lak, and David H. Jonassen. 2002. "The Effects of Argumentation Scaffolds on Argumentation and Problem Solving." *Educational Technology Research and Development* 50 (3): 5–22. doi:10.1007/BF02505022.

Crouch, C. H., and E. Mazur. 2001. "Peer Instruction: Ten Years of Experience and Results." *American Journal of Physics* 69: 970.

Davidson, Janet E., and Robert J. Sternberg. 2003. *The Psychology of Problem Solving*. Cambridge University Press.

Diefes-Dux, Heidi A., Judith S. Zawojewski, Margret A. Hjalmarson, and Monica E. Cardella. 2012. "A Framework for Analyzing Feedback in a Formative Assessment System for Mathematical Modeling Problems." *Journal of Engineering Education* 101 (2): 375–406. doi:10.1002/j.2168-9830.2012.tb00054.x.

Dunlosky, John, Katherine A. Rawson, Elizabeth J. Marsh, Mitchell J. Nathan, and Daniel T. Willingham. 2013. "Improving Students' Learning With Effective Learning Techniques Promising Directions From Cognitive and Educational Psychology." *Psychological Science in the Public Interest* 14 (1): 4–58. doi:10.1177/1529100612453266.

Dym, C., A. Agogino, O. Eris, D. Frey, and L. Leifer. 2005. "Engineering Design Thinking, Teaching, and Learning." http://digitalcommons.olin.edu/mech_eng_pub/22/.

"Employability Skills Framework - Source Matrix." 2016. Accessed January 8. http://cte.ed.gov/employabilityskills/index.php/framework/source_matrix.

Engineers Canada, and Canadian Council of Technicians and Technologists. 2009. "Engineering and Technology Labour Market Study Final Report." http://www.engineerscanada.ca/files/w_Engineering_and_Technology_Labour_Mark_Study.pdf.

Finley, Ashley. 2012. "Reliable Are the VALUE Rubrics?" *Peer Review* 13/14 (4/1.).

Frank, Brian, and Behnam Hamgini. 2014. “Collaborative Cloud-Based Documents for Real-Time Bi-Directional Feedback in Large Lecture Activities.” In *ASEE 2014 Annual Conference*. Indianapolis, IN.

Frank, Brian, David Strong, Rick Sellens, and Lynann Clapham. 2013. “Progress with the Professional Spine: A Four-Year Engineering Design and Practice Sequence.” *Australasian Journal of Engineering Education* 19 (1): 63.

“Graduate Attributes and Professional Competencies.” 2013. International Engineering Alliance. http://www.washingtonaccord.org/IEA-Grad-Attr-Prof-Competencies.pdf.

Greiff, Samuel, Andreas Fischer, Matthias Stadler, and Sascha Wüstenberg. 2015. “Assessing Complex Problem-Solving Skills with Multiple Complex Systems.” *Thinking & Reasoning* 21 (3): 356–82. doi:10.1080/13546783.2014.989263.

Greiff, Samuel, Daniel Holt, and Joachim Funke. 2013. “Perspectives on Problem Solving in Educational Assessment: Analytical, Interactive, and Collaborative Problem Solving.” *The Journal of Problem Solving* 5 (2). doi:10.7771/1932-6246.1153.

Griffin, Patrick, and Esther Care, eds. 2015. *Assessment and Teaching of 21st Century Skills*. Dordrecht: Springer Netherlands. http://link.springer.com/10.1007/978-94-017-9395-7.

Hake, R. R. 1998. “Interactive-Engagement versus Traditional Methods: A Six-Thousand-Student Survey of Mechanics Test Data for Introductory Physics Courses.” *American Journal of Physics* 66: 64.

Hammer, Sara Jeanne, and Wendy Green. 2011. “Critical Thinking in a First Year Management Unit: The Relationship between Disciplinary Learning, Academic Literacy and Learning Progression.” *Higher Education Research & Development* 30 (3): 303–15. doi:10.1080/07294360.2010.501075.

Hattie, J. 2008. *Visible Learning: A Synthesis of over 800 Meta-Analyses Relating to Achievement*. Routledge.

Huber, Christopher R., and Nathan R. Kuncel. 2015. “Does College Teach Critical Thinking? A Meta-Analysis.” *Review of Educational Research*, September, 0034654315605917. doi:10.3102/0034654315605917.

International Engineering Alliance. 2013. “Graduate Attributes and Professional Competencies.” http://www.washingtonaccord.org/IEA-Grad-Attr-Prof-Competencies.pdf.

Isaksen, Scott G., K. Brian Dorval, and Donald J. Treffinger. 2010. *Creative Approaches to Problem Solving: A Framework for Innovation and Change*. SAGE Publications.

Jiao, H. 2015. “Enhancing Students’ Engagement in Learning through a Formative E-Assessment Tool That Motivates Students to Take Action on Feedback.” *Australasian Journal of Engineering Education* 20 (1): 9–18. doi:10.7158/D13-002.2015.20.1.

Jonassen, David H. 2010. *Learning to Solve Problems: A Handbook for Designing Problem-Solving Learning Environments*. New York: Routledge.

———. 2014. *Engineers as Problem Solvers (ch.6)*. Edited by Aditya Johri and Barbara M Olds. Cambridge Handbook of Engineering Education Research. Cambridge University Press.

Kaupp, Jake, and Brian Frank. 2013. “Impact of Model Eliciting Activities on Development of Critical Thinking.” In *ASEE Annual General Conference*. Atlanta, GA.

Kleij, Fabienne M. Van der, Remco C. W. Feskens, and Theo J. H. M. Eggen. 2015. "Effects of Feedback in a Computer-Based Learning Environment on Students' Learning Outcomes A Meta-Analysis." *Review of Educational Research* 85 (4): 475–511. doi:10.3102/0034654314564881.

Kluger, A. N., and A. DeNisi. 1996. "The Effects of Feedback Interventions on Performance: A Historical Review, a Meta-Analysis, and a Preliminary Feedback Intervention Theory." *Psychological Bulletin* 119 (2): 254.

O'donnell, Angela M, Donald F Dansereau, and Richard H Hall. 2002. "Knowledge Maps as Scaffolds for Cognitive Processing." *Educational Psychology Review* 14 (1): 71–86.

Pascarella, Ernest T, and Patrick T Terenzini. 2005. *How College Affects Students: A Third Decade of Research*. San Francisco: Jossey-Bass.

Rhodes, Terrel, and Ashley Finley. 2013. "Using the VALUE Rubrics for Improvement of Learning and Authentic Assessment." Association of American Colleges and Universities.

Rosenshine, Barak, and Carla Meister. 1992. "The Use of Scaffolds for Teaching Higher-Level Cognitive Strategies." *Educational Leadership* 49 (April): 26–33.

Sawyer, R. Keith. 2011. *Explaining Creativity: The Science of Human Innovation*. 2 edition. Oxford University Press.

Sheppard, S., K. Macatangay, A. Colby, and W. M. Sullivan. 2009. *Educating Engineers: Designing for the Future of the Field*. Jossey-Bass San Francisco, CA. http://www.lavoisier.fr/livre/notice.asp?id=OLSWL6ALLK2OWA.

Shute, Valerie J. 2008. "Focus on Formative Feedback." *Review of Educational Research* 78 (1): 153–89. doi:10.3102/0034654307313795.

Stein, Barry, and Ada Haynes. 2011. "Engaging Faculty in the Assessment and Improvement of Students' Critical Thinking Using the Critical Thinking Assessment Test." *Change: The Magazine of Higher Learning* 43 (2): 44–49. doi:10.1080/00091383.2011.550254.

Stein, Barry, Ada Haynes, and Michael Redding. 2016. "National Dissemination of the CAT Instrument: Lessons Learned and Implications." Accessed July 28. http://www.enfusestem.org/wp-content/uploads/2016/04/AAAS-Paper-2016-Stein_Haynes_Redding-revision.pdf.

Sweller, John, Paul Ayres, and Slava Kalyuga. 2011. "Cognitive Load Theory in Perspective." In *Cognitive Load Theory*, 237–42. Explorations in the Learning Sciences, Instructional Systems and Performance Technologies 1. Springer New York. http://link.springer.com/chapter/10.1007/978-1-4419-8126-4_18.

Trevelyan, James. 2014. "The Making of an Expert Engineer." *CRC Press*. September 22. https://www.crcpress.com/The-Making-of-an-Expert-Engineer/Trevelyan/9781138026926.

Williams, Rosalind. 2002. "Retooling: A Historian Confronts Technological Change." *MIT Press*. https://mitpress.mit.edu/books/retooling.

Zahner, D. 2013. "Reliability & Validity - CLA+." Council for Aid to Education. http://cae.org/images/uploads/pdf/Reliability_and_Validity_of_CLA_Plus.pdf.

Zawojewski, Judith S, Heidi Diefes-Dux, and Keith Bowman. 2008. *Models and Modeling in Engineering Education*. Designing Experiences for All Students. Sense Pub.

Zheng, Wei, and Yanhua Cao. 2015. "Implementation and Outcomes of Scaffolding Collaborative Learning in Multiple STEM Courses." In *ASEE Annual Conference. American Society of Engineering Education*. https://www.asee.org/public/conferences/56/papers/13967/view.

Zheng, Wei, Yanhua Cao, Himangshu Shekhar Das, and Jianjun Yin. 2014. "Scaffolding Cyber-Enabled Collaborative Learning in Engineering Courses and Its Impacts of on Students' Learning." In *ASEE Annual Conference. American Society of Engineering Education*. http://www.asee.org.proxy.queensu.ca/file_server/papers/attachment/file/0004/4750/ASEE-2014-_Scaffolding_Cyber-final_paper-4-reference_chang_.pdf.