8-2016

Promoting Teaming Metacognition

Lee Kemp Rynearson
Purdue University

Graduate School Form
30 Updated 12/26/2015

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Lee Kemp Rynearson

Entitled
Promoting Teaming Metacognition

For the degree of Doctor of Philosophy

Is approved by the final examining committee:

Heidi Diefes-Dux
Chair

Matthew Ohland

Morgan Hynes

Edward Berger

To the best of my knowledge and as understood by the student in the Thesis/Dissertation Agreement, Publication Delay, and Certification Disclaimer (Graduate School Form 32), this thesis/dissertation adheres to the provisions of Purdue University's "Policy of Integrity in Research" and the use of copyright material.

Approved by Major Professor(s): Heidi Diefes-Dux

Approved by: David Radcliffe 7/6/2016
Head of the Departmental Graduate Program Date

PROMOTING TEAMING METACOGNITION

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Lee K Rynearson

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

August 2016

Purdue University

West Lafayette, Indiana

ACKNOWLEDGEMENT

I would like to acknowledge and thank all members of my doctoral committee for their effort and support. This dissertation was enormously improved through their insight and advice. I am excited to continue the work begun here, and believe that the guidance received from members of the committee will bear its greatest fruit in studies yet to come.

TABLE OF CONTENTS

Page

Page

Page

Page

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Rynearson, Lee K. Ph.D., Purdue University, August 2016. Promoting Teaming Metacognition. Major Professor: Heidi Diefes-Dux.

Improving students' capacity to effectively perform teaming skills is a crucial outcome for engineering education, and has been the subject of considerable prior and ongoing research. Based upon review of research on teaming, it was hypothesized that greater awareness of appropriate opportunities to use teaming skills in authentic contexts would lead to greater teaming skills employment over time. Further, it was hypothesized that greater psychological safety in student teams would lead to more students choosing to employ appropriate teaming skills over time. An intervention to achieve such increases could therefore be expected to promote student teaming skills performance improvement. Seeking to evaluate potential new methods for teaming skills instruction and development in engineering contexts, a suite of interventions was designed to support growth in student metacognition and promote psychological safety in student teams.

This dissertation took the form of a quantitative study implemented in nine sections of the Purdue First-Year Engineering classes ENGR131 (five sections) and ENGR141 (four sections). Multiple psychometric instruments were administered across a semester. Results were investigated to assess the efficacy of the experimental interventions in supporting student teaming metacognition, raising psychological safety, altering relationships between measured variables, and ultimately in raising teaming skills performance. The experimental interventions used in this study incorporate tools and techniques that do not appear to have been previously employed in engineering education. Results suggest that the instrumentation for metacognition was not satisfactory, but that the intervention may have had effects on psychological safety in student teams. These findings are discussed along with directions for further inquiry in the design, implementation, and evaluation of teaming skills instruction.

CHAPTER 1. INTRODUCTION

1.1 Background

Researchers in engineering education have pursued evidence for the value of various engineering instructional techniques for decades, with the most notable and widespread success being the conclusive demonstration of the efficacy of active learning techniques in improving numerous undergraduate learning outcomes in comparison with traditional lecture-based methods (Johnson, 1991; Johnson, Johnson, & Smith, 1998; Prince, 2004, p. 625; Slavin, 1996). While active learning methods are understood to have broadly positive results, a need remains to gather evidence on methods in active learning that are efficacious in improving particular student outcomes, in keeping with Jamieson and Lohmann's (2012) call for evidence-based educational innovation. Understanding which active learning methods are most effective in addressing a given learning outcome, under what circumstances, and why, would permit the design of undergraduate engineering learning experiences that reliably surpass the efficacy of generic 'active learning' efforts.

Researchers have begun efforts to develop deeper understandings of active learning, proposing models of learning incorporating results from many research fields (Vanasupa, Stolk, & Herter, 2009) and working to understand and differentiate more and less effective active learning paradigms (Chi, 2009). Incorporating research finding from fields complimentary to engineering education is likely imperative to achieving the goal of understanding active learning design, as many known facets of learning, from student motivation to effective methods for memorization, have been primarily investigated by researchers outside of engineering education. However, interventions intended for engineering would be most strongly supported by evidence of their efficacy in classroom settings, not just lab environments, as are frequently found in some fields of research. This study is conducted in the authentic context of two first-year engineering (FYE) courses.

1.2 Research Purpose and Research Questions

The purpose of this study was to develop, deploy, and evaluate a classroom intervention promoting student performance of teaming skills in a first-year engineering (FYE) context, continuing and expanding the work of a pilot study in this area (Rynearson & Hynes, 2015). This work contributes to the effort to develop more effective active learning techniques by targeting specific learning outcomes and drawing from theory and educational research from inside and outside of the field of engineering education.

As will be discussed in the literature review, teaming skills are critical for engineers. Team projects are a common method employed to facilitate the growth of teaming skills and are often employed in FYE or senior design courses. Existing teaming instruction in the Purdue University FYE sequence provides information about the importance of teaming, general information about teaming processes and practices, opportunities to put teaming into practice, and feedback to each student about their teaming performance, and has been shown to be successful in cultivating teaming skills growth (Jimenez-Useche, Ohland, & Hoffmann, 2015). This dissertation's experimental intervention overlays existing instruction with a focus on cultivating metacognitive knowledge and awareness and student psychological safety in working teams with the understanding that growth in these areas is likely to support higher overall teaming performance. The experimental intervention was administered to a subset of participating engineering students, such that a control group was formed to allow comparison of results. Therefore, this study addresses the following research questions:

1) To what extent did the experimental intervention promote the target aspects of metacognition in FYE students in the experimental group in comparison to the control group?
2) To what extent did the experimental intervention promote psychological safety of student teams in the experimental group in comparison to the control group?
3) In the event that the intervention succeeds in supporting more metacognitive skills growth and/or psychological safety in the experimental than the control groups, to what extent did the teaming performance of the experimental group then improve beyond that of the control group?
4) To what extent did the improvement in the targeted metacognitive capabilities or psychological safety correlate with improved individual student teaming performance?

1.3 Research Context and Current Practices

This section provides an introduction to the context in which the study was undertaken, including practices in teaming skills improvement currently employed. A limited effort will be made to identify and reference best practices currently employed to support this contextualization. This section does not provide a comprehensive review of best practices in teaming skills development.

As previously mentioned, the study took place in authentic engineering classroom settings. These settings were provided by two Purdue University First-Year Engineering courses, Transforming Ideas to Innovation 1 (ENGR131) and Honors Creativity and Innovation in Engineering Design 1 (ENGR141). ENGR131 is a four-contact-hour, two-credit course meeting twice a week for one semester. ENGR141 is a six-contact-hour, three-and-a-half credit course meeting three times a week for one semester. Both are the first course in a two-course sequence, with ENGR131 being followed by Transforming Ideas to Innovation 2 (ENGR132) and ENGR141 being followed by Honors Creativity and Innovation in Engineering Design 2 (ENGR142). Aside from mentioning that these follow-up courses also offer numerous opportunities for students to benefit from improved teaming skills, the follow-up courses do not require discussion.

1.3.1 ENGR131

ENGR131 is the largest FYE course offered to students of Purdue University and serves the majority of undergraduate engineering students in their first semester. The total number of students in the course was approximately 1800 in the fall 2015 semester, divided into sections of roughly 120 students each. Students in ENGR131 have relatively strong academic preparation (manifested through SAT/ACT scores, high school class GPA and class rank, etc.) as would be expected of students in a highly selective engineering program. The Purdue Engineering Office of Future Engineers reports the incoming class in the fall of 2015 had an SAT middle 50% range of 1800-2080 and a high school GPA middle 50% range of 3.8-4.0 (Engineers, 2016). The Purdue University College of Engineering attracts a high proportion of international students, approximately 25% of enrollment (Jimenez-Useche et al., 2015).

ENGR131 covers topics including diversity and teaming skills, engineering major selection and career preparation, mathematical modeling, Microsoft Excel®, the engineering decision making, the design process, and information literacy.

ENGR131 features teaming skills instruction early in the course and revisits the topic in-class around the midpoint of the course. Team activities and projects are incorporated and times for peer evaluation occur periodically throughout the course. Teams are instructor-formed using the CATME Team-Maker tool (Layton, Loughry, Ohland, & Ricco, 2010), based on criteria including each student's daily schedule, gender, Under-Represented-Minority (URM) status, and international/domestic status. In general, students from historically vulnerable populations are grouped with at least one like team member. Additionally, it is ensured that no two teammates share a first or last name for administrative convenience (though one student's first name can be the same as another's surname). Teams are predominantly composed of four students with some teams of three, and are permanent once formed except under extraordinary circumstances. These practices are specifically recommended for student teams in engineering by Oakley, Felder, Brent, and Elhajj (2004) and Wankat and Oreovicz

(2014). In fall 2015, all experimental and control groups of ENGR131 participated in versions of the following activities:

1. *In-class discussions of diversity and basic information about the motivation for teaming skills along with team working roles and processes*. This coverage constitutes “guidance from the instructor on effective teamwork” as recommended by Oakley, Hanna, Kuzmyn, and Felder (2007, p. 271) and also the discussion of “the importance of teamwork, problems experienced with student teams, ways to resolve these problems...” specified by Stephens (2001, p. 337).
2. *The development of a 'Code of Cooperation' by each team laying out their internal rules and expectations for working together and a 'team poster' with team pictures and associated information*. The 'Code of Cooperation' exercise extends Stephens' recommendation to have teams “formulate a contract which specifies criteria for non-performance” (Stephens, 2001, p. 337) and falls in line with Felder's suggestion that faculty have students “prepare and sign a list of ground rules they all agree to observe” (Felder & Brent, 2001, p. 3).
3. *Journal entries with question prompts relating to teaming issues outside of class*. These journal entries, integrated into the ENGR131 curriculum and part of normal coursework, appear to be targeted at providing students with an initial engagement with the topic prior to further classroom discussion. This is a fairly standard flipped-classroom active learning practice aimed at stimulating engagement as discussed in Millard (2012).
4. *Many active learning activities and exercises on various topics completed in teams*. This heading covers the large volume of classwork for all sections already

comprised of team-based active learning materials independent of this study. The efficacy of in-class active learning practices has already been established, and these activities predominantly fall into Chi's 'interactive' category of active learning activities, which is most likely to lead to learning gains (Chi, 2009).

5. *A roughly three-week team mathematical modeling project, substantially undertaken in-class*. This project maps to interactive active learning in Chi's taxonomy.
6. *A roughly four-week design project, substantially undertaken in-class*. This project is also an interactive active learning experience according to Chi's taxonomy.
7. *A team 'practical' on engineering design, which is essentially a team examination of applied design process management and design capabilities in a timed format*. Another class day is allocated for practice of this activity. These tasks are also interactive active learning in Chi's taxonomy.
8. Four CATME peer evaluations. CATME (Ohland et al., 2012) is recommended as a resource to help students improve their teaming skills (Hrivnak, 2013; Weimer, 2013) and has been found to help students improve their teaming behavior (Pung & Farris, 2011). It provides "mid-term peer evaluations with feedback" as recommended by Stephens (2001, p. 337). It should also be noted that students can, but are not required to, elicit faculty advice or feedback outside of this system, and that faculty can give direct feedback on their own. While such feedback can be very helpful to student teams, it is not an assessment target of this

intervention and no special efforts with regards to it differentiate the experimental and control sections.

As can be seen, ENGR131 has a schedule that already incorporates a large number of team-based activities instantiating existing best practices in teaming skills development. The existing teaming skills development opportunities offer a strong control to compare the results of this study's experimental intervention against. As will be made clear in the methods section, the experimental intervention designed and implemented in this study included activities and practices not currently found in engineering teaming instruction.

1.3.2 ENGR141

ENGR141 is a smaller FYE course offered to students of the Purdue University Honors College to meet certain requirements of that academic body. The total number of students in the course was approximately 270 in the fall 2015 semester, divided into four sections of over 60 students each. Students in ENGR141 have higher average SAT/ACT scores and high school GPA's than students in ENGR131 and a higher probability of having been highly ranked in high school. Prospective students are selected for invitation to join the Honors College for these and other characteristics. Students who accept the invitation of the Honors College and pursue an engineering degree typically take ENGR141, but it is not mandatory. The international contingent of ENGR141 is much smaller than that of ENGR131, with fewer than 5% of fall 2015 semester students being of international origin. Honors College invitations to international students interested in engineering have been increased roughly five-fold in recent years to enroll more international students in ENGR141. However, in the fall of 2015, students of ENGR141 were not representative of most undergraduate populations. Results from ENGR141 do present an interesting opportunity to assess the effects of the intervention on domestic students entering engineering with higher than average preparation.

ENGR141 is designed to be extremely challenging and is widely so regarded by students. ENGR141 introduces multiple topics simultaneously, gives several large assignments in a typical week, and calibrates overarching team project difficulty such that approximately 10% of teams are successful at a majority of project goals. Some students consider the workload and challenge of ENGR141 to be more extensive than the extra 1.5 academic credits if offers over ENGR131 merit. ENGR141 covers topics including teaming skills, the design process, project management, algorithm development and documentation, programming in Python and MATLAB, mathematical modeling including data cleaning, curve fitting, correlation, descriptive statistics and the development of models for novel situations. Additional course activities support additional limited learning goals outside of these topics. Pedagogy in ENGR141 features much more extensive lectures (punctuated by discussion and active learning exercises) than ENGR131.

As with ENGR131, teaming skills instruction includes some relatively brief units early in the course followed by extended team activities and projects, dotted with times for peer evaluation and reflection throughout the course. As with ENGR131, teams are instructor-formed, predominantly composed of four students with some teams of three, and are permanent once formed. The CATME Team-Maker tool was not used in the fall of 2015 in ENGR141 in favor of an ENGR141-specific team formation tool that did not account for out-of-class availability but did ensure women and URM students were placed with some like students. The ENGR141 team creation tool attempts to distribute students with previous programming experience, strong high school science experience, and strong high school math experience (based on student surveys) evenly across teams, such that all or nearly all teams had a variety of class-relevant strengths and experiences. It may be possible to gather and sort by this information using CATME Team-Maker, but ENGR141 administrators did not explore this option. In fall 2015, both the experimental and control groups of ENRG141 participated in versions of the following activities:

1. *In-class lectures and discussions of the importance of teaming and basic information about team working processes, similar to those discussed for ENGR131*.
2. *A reading or video assignment to promote student teaming motivation and understanding prior to classroom discussions of the topic*.
3. *The development of a 'Code of Cooperation' by each team laying out their internal rules and expectations for working together and an 'E-card' with team pictures and a team name*. References supporting these activities appear in the description of similar items in ENGR131
4. *Three CATME peer evaluations*. A more detailed description of these activities and references supporting such peer evaluations are given in the discussion of ENGR131's efforts.
5. *In-class reflections and discussions on team processes*. These in-class reflections and discussions (which are forms of active learning) offer opportunities to capitalize on feedback to benefit their teaming skills as recommended by

(Stephens, 2001). Additionally, these activities model appropriate reflective practice for students, a valuable skill employed by professional engineers (R. S. Adams, Turns, & Atman, 2003; Svarovsky & Shaffer, 2006).

6. *A two-week out-of-class project to develop a straw tower as a team*. Resource scarcity of the building materials means that teams not working together may be subject to critical design or construction errors by individuals or sub-groups. This project constitutes interactive active learning in Chi's taxonomy.
7. *Eleven class sessions primarily dedicated to team learning of programming concepts and languages or additional practice in specific course skills*. Teams are typically limited to working on a single computer with students rotating control of the computer. These working periods offer interactive active learning as laid out in Chi's taxonomy.
8. *Two class days entirely composed of timed and graded team engineering challenges supporting specific learning objectives*. These challenges are not competitive between teams – the limited time and the need for individuals on teams to work together efficiently to accomplish their tasks in the given time have historically been sufficient motivation. These tasks constitute interactive active learning in Chi's taxonomy.
9. *An eight-day in-class project on gathering, cleaning, analyzing, and building, and reporting on mathematical models from experimental data*. This project constitutes interactive active learning in Chi's taxonomy.
10. *A fourteen-week primarily out-of-class team term project in robotics*. This project constitutes interactive active learning in Chi's taxonomy.

11. *Many additional assorted smaller active learning activities and exercises in the context of class lectures on various topics, the majority of which will fall into the constructive and interactive categories of Chi's taxonomy, the two most likely to lead to learning gains*.

As can be seen, ENGR141 also has a full schedule that incorporates a large number of team-based activities instantiating best practices in teaming skills development. As with ENGR131, existing teaming skills development opportunities are robust and offer a strong control to compare the results of the intervention against. As will be discussed in the next section, the goal of the intervention is not to replace or divert class time and attention away from these worthwhile activities wholesale, but to intervene in a limited fashion with the goal of making teaming skills acquisition and improvement during some activities (principally team working time) more efficient.

1.4 Overview of Intervention Concept

This section serves to orient readers to the basic ideas and structure of the study prior to in-depth discussions of specific aspects of the study in later chapters. At its most basic level the intervention works to induce students to get more practice in the authentic use of teaming skills, in keeping with the general principles of active learning. Through this practice, students' teaming skills are expected to improve. As the classroom context for the intervention already includes substantial teaming skills learning and performance opportunities in accordance with known best practices, the study deployed and assessed new methods derived from research that may be effective beyond the results achieved through current methods. Control cohorts in both ENGR131 and ENGR141 provided a method for distinguishing the effects of the intervention from other teaming skills development activities.

To induce students to get more practice in the authentic use of teaming skills, it was useful to examine a model of decision-making, the idea being that students could be put on the path to making decisions to employ teaming skills with greater frequency. The AIDA (Awareness, Interest, Desire, Action) Hierarchy of Effects model and its numerous descendants (Wijaya, 2015) in the field of marketing provided a starting point for a model. In AIDA and other traditional marketing models of decision making, the subject is seen as transitioning through a number of stages of decision making from being completely unaware of a product to having bought it. In some newer models discussed by Wijaya (2015), stages past purchase are added reflecting items such as customer satisfaction and customer use of social media in relation to the product, along with acknowledgements that customer decision-making processes are not always linear. Progression through the stages is sometimes referred to as the ‘sales funnel’, and it is expected that some proportion of potential customers will fail to advance at each stage.

Refocusing this decision-making hierarchy onto the employment of teaming skills, the main thrust of this study was to employ research in the cultivation of metacognition to promote the awareness or recognition of opportunities to employ teaming skills via active learning practices, and thereby to increase the base of the ‘sales funnel’ for student decisions to employ teaming skills. An additional target of this study was to employ research into psychological safety to better support student motivation to employ teaming skills by reducing the perceived risk of doing so. Other aspects of the intervention support other stages of the decision-making process, and further support was provided by standard instructional practices in both ENGR131 and ENGR141. Existing practices in both courses were retained for both the experimental and control groups as it would have been unethical to deliberately remove best practices likely to support learning from a classroom intervention. Overall, the study integrates new techniques based on research literature into practice and compares the results against strong existing modern instruction.

The AIDA model, being aimed at sales to the general public, does not align precisely with processes that may be anticipated for an audience of students being asked to perform teaming skills. This study employed an adapted version of the model with the following five stages:

1. Awareness
2. Motivation
3. Selection
4. Implementation
5. Reflection

The rationale for each stage in the adapted model is given in the following sections. This five-step process primarily serves to contextualize aspects of the study rather than to determine granular aspects of it; the stages of the process are so broad as to require further review of the literature and selection of methods. However, it is helpful in understanding what different aspects of the intervention are intended to accomplish. Each stage will now be discussed in more detail, along with aspects of the experimental intervention and best-practices present in both experimental and control sections. Note that the stage names are henceforth italicized when referring specifically to a stage to differentiate them from other uses of the words (e.g. *awareness* is the stage, awareness is generic).

1.4.1 Awareness

Students must have awareness of the opportunity to employ targeted teaming skills. In response to problems or opportunities in the team environment, it is possible for students to act with greater or lesser degrees of thought and intentionality. Reacting without thought to a situation may result in appropriate employment of teaming skills. However, a thinking reaction is required to result in the target behavior: the use of teaming skills in situations where teaming skills would otherwise not have been used. Therefore, student *awareness* of specific opportunities to employ teaming skills is a first step towards additional practice of teaming skills. As stated, this study employs research in the cultivation of metacognition to promote the recognition of opportunities to employ teaming skills via active learning practices. The primary method for this is the development and deployment of worksheets (Strategy Evaluation Matrices) that could work to increase student *awareness* of teaming opportunities. The previously mentioned pilot study showed some promise in promoting awareness of opportunities to employ teaming skills (Rynearson & Hynes, 2015) using metacognition.

1.4.2 Motivation

Subsequent to recognizing the opportunity to deploy teaming skills, students must have *motivation* with regards to whether to take an action in response to the teaming problem or opportunity. '*Motivation*' replaces 'Interest' from the AIDA model in this instance, being more appropriate and descriptive for students in mandatory coursework. Experience with FYE student teams, including feedback on team dynamics and performance via the CATME system, suggest that not all students regularly apply effort towards addressing issues and opportunities in their teams. For the purposes of this intervention it was generally seen as desirable to employ some variety of deliberate actions as a means to practice teaming skills. This could include actions such as actively listening and similarly quiet or less-visible skills.

There are many potential avenues to motivate students to more frequently decide to actively employ teaming skills. All sections of ENGR131 and ENGR141 employ some basic measures to motivate students, such as presenting information on the importance of teaming in industry and the potential for strong teaming to produce better products, which could positively affect student grades. Beyond these common measures, the experimental intervention focused on improving intra-team psychological safety. Psychological safety will be explored in more depth in the literature review, but is essentially "a sense of confidence that the team will not embarrass, reject, or punish someone for speaking up" (Edmondson, 1999, p. 354). It was expected that improved student psychological safety would lead to reduced student concerns about potential negative team reactions to attempts to perform teaming skills. Thus, factors reducing *motivation* to perform teaming skills would be reduced, effectively increasing *motivation* and therefore the amount of teaming skills practice undertaken by students. Increased psychological safety also potentially contributes to teaming skills development through feedback and example. For example, all members of a team with a relatively dominant member might benefit if one less-dominant member began to work to correct the conduct of the dominant member in view of the team.

Additionally, psychological safety may increase *motivation* to interact as a team through mitigating conflict. One reason that students may avoid employing teaming skills is in response to real or perceived conflict in their teams. Conflict can lead to team members interacting less as “team members try to disengage from those with whom they experience conflict, and further limit their interactions” (Langfred & Moye, 2014, p. 33). As psychological safety makes it more likely that team issues will be broached before they become crises, it has the potential to reduce team conflict and increase *motivation* to employ teaming skills in this way as well.

Psychological safety is supported by the experimental intervention through the development and deployment of student team Codes of Cooperation in the experimental sections, along with brief in-class presentations on psychological safety. Codes of Cooperation are already used by both ENGR131 and ENGR141, but the experimental sections’ altered Code of Cooperation assignments explicitly required teams to develop and commit to a plan to ensure strong psychological safety in their team environment.

1.4.3 Selection

Either concurrent with or subsequent to the second step, students must select an approach to the teaming problem or opportunity. This step is necessary in the model of student teaming decision making in addition to the steps in the AIDA model as there is more than one possible final action to take (all possible teaming actions, versus simply buying the advertised item). The student must therefore select the action to be implemented. This step depends heavily on student knowledge of potential actions and their suitability for employment in response to the problem or opportunity the student is aware of. All sections of ENGR131 and ENGR141 are broadly introduced to appropriate teaming actions through course presentations and the use of the CATME BARS peer evaluation system (Loughry, Ohland, & Woehr, 2014). The worksheets employed in the experimental intervention (Strategy Evaluation Matrices) worked to scaffold the *selection* of appropriate teaming activities in response to teaming problems or opportunities.

1.4.4 Implementation

Fourth, subsequent to selecting a teaming skill to perform, students must *implement* the use of the teaming skill in their team environment. This step depends on things like student knowledge of and skills in communication, ability to regulate their attitude when interacting with teammates, and to empathize with the perspectives of others. These items were not assessed in the study, and an uneven distribution of these factors across the sample is a potential confounding factor. *Implementation* is lightly addressed by some current ENGR131 and ENGR141 class practices, such as in-class discussions of how teaming interactions might best be approached, and the Code of Cooperation assignments may, at a student team's discretion, contain guidelines for when and how students should implement some teaming interactions. The experimental intervention was expected to result in improvements in student ability in this area through additional in-class practice during the course of normal class operations rather than explicit practice in the form of role plays, exercises, or other activities.

1.4.5 Reflection

The final step is for students to *reflect* on their teaming skills performance. This step is not included in the AIDA model and may not be helpful for selling products. It is, however, helpful for learning. Reflective practice allows students to derive more benefit from past teaming skills performances through considering personal beliefs, stronger and weaker skills, and deciding on any changes in outlook or practice to be attempted in future iterations. This step is addressed by some current class practices in both ENGR131 and ENGR141, which already featured reflective practice to some extent. Various in-class and out-of-class prompts for *reflection* relevant to the intervention were implemented in support of the intervention's goals in ENGR131 and ENGR141.

1.5 Time and Logistical Limitations to the Intervention

It is important to account for the amount of classroom time consumed in inducing and supporting additional teaming skills practice. While classroom interventions to improve teaming skills abilities have been deployed in engineering, including many already deployed in Purdue University FYE courses, interventions designed to push beyond current practice can be time-intensive, such as the five-course minor in Engineering Communication and Performance reported on by Seat, Parsons, and Poppen (2001). As it is well known that engineering curricula are quite full and noting legislative action in Indiana to reduce the quantity of credits required to graduate (Wheldon, 2013), this intervention worked to act in the context of existing courses, consuming two or fewer in-class hours. This quantity of class time was determined to approach the maximum acceptable to faculty teaching the experimental sections of this study. This time limitation did affect the selection and scheduling of intervention activities. Spending this amount of time on new material in an FYE course may be seen as plausible to FYE instructors more broadly, as the displacement of material and/or credits is much more limited than, for instance, a five-course minor. The concepts and methods employed in this experimental intervention could be scaled up such that they consumed more class time, potentially to greater effect.

1.6 Intervention Overview Summary

In summary, the intervention was designed to improve student teaming performance by increasing the amount of teaming skills practice that students get in the course of normal teaming activities in ENGR131 and ENGR141. This additional practice comes from promoting *awareness* of opportunities for teaming practice through metacognition and other activities that support students in progressing through the sequence of *awareness*, *motivation*, *selection*, *implementation*, and *reflection* in teaming. *Awareness* and *selection* were primarily supported by Strategy Evaluation Matrices developed for engineering teaming, *motivation* was primarily supported by Code of Cooperation assignments emphasizing psychological safety, and reflection, while already present in the target courses, was refocused on the target behaviors. Chapter 3 provides detailed descriptions of the interventions and assessments. Administration timelines can be seen in Table 3.2 and Table 3.3, and study activities by research question can be seen in Table 3.4. Existing best practices in student motivation, teaming skills knowledge and training, and reflection were generally retained, but were not areas of focus of this study.

The intervention was designed to consume fewer than two class hours for logistical reasons.

CHAPTER 2. LITERATURE REVIEW

In this chapter, reviews of research literature relating to the need for engineering teaming skills development, aspects of metacognition useful for developing teaming skills via the previously discussed 5-step sequence, methods for the development of metacognitive skills and knowledge, the concept of psychological safety and its effects on team performance, and fostering team psychological safety are conducted. Given the breadth of possible material on the topics reviewed, only key features of the research literature with respect to this study are provided, not comprehensive general literature reviews for each research area. As this study enacts research-to-practice, the focus of many sections of this review are on instructional implications that can be derived from the literature and methods that could be employed to achieve the educational goals of the intervention. Literature relating to assessment appears in Chapter 3, alongside the specific methods employed in this study. Some review of best practices in teaming skills development appeared in Chapter 1 in the description of the study's context.

2.1 The Need for Engineering Teaming Skills Development

The intervention employed in this study sought to improve student teaming skills. This section motivates the need for and benefits of teaming skills for engineering students. It has been broadly established that engineers require interpersonal, communication, and teaming skills in industry, even when employed in predominantly technical roles. Prominent voices from industry articulated problems in these areas two decades ago, stating that "most major American universities overemphasize engineering science at the expense of engineering practice" including "cooperative learning/teamwork" (McMasters & Matsch, 1996, p. 1). A more recent review by Martin, Maytham, Case, and Fraser (2005) stated that "Surveys...of industry perceptions of engineering graduates have consistently identified communication and teamwork as important attributes where "competency gaps" are frequently found." (p. 168). Similar difficulties have recently been reported in Canadian engineering programs (May & Strong, 2011).

In response to the identification of these industry needs, movement began towards a greater emphasis on teamwork, among other aspects of engineering practice, in the education of engineers. The introduction of the notable CDIO engineering syllabus used in whole or in part by dozens of engineering schools around the world states that "there is a growing recognition that young engineers must possess a wide array of personal, interpersonal, and system building knowledge and skills that will allow them to function in real engineering teams and to produce real products and systems." (Crawley, 2001, p. 1) Mandating some emphasis on teamwork at virtually all American engineering schools, ABET has required engineering programs to present documented evidence that graduating students possess the abilities to "communicate effectively" and "function on multidisciplinary teams" (ABET, 2013, p. 3) to retain accreditation since the turn of the century. A recent review of First-Year Engineering curricula found that "Teaming skills and communication were by far mentioned in most if not all syllabi and group discussion" and that these items are still the "most cited characteristics from employers" (Reid, Hertenstein, Fennel, Spingola, & Reeping, 2013, p. 7).

In addition to the industrial and accreditation demand for teaming skills, engineering students with strong teaming skills may accrue benefits in their technical education. Cooperative and other forms of collaborative learning have become widespread in higher education in general (Johnson & Johnson, 2009) and engineering is no exception. Research on the efficacy of active (including cooperative) learning in engineering is compelling, with authors such as Smith, Sheppard, Johnson, and Johnson (2005) stating that "cooperative learning and problem-based learning can advance academic success, quality of relationships, psychological adjustment, and attitudes toward the college experience" (p. 96). Prince (2004) states that "the best available evidence suggests that faculty should structure their courses to promote collaborative and cooperative environments" (p. 7). As engineering faculty increasingly respond to this evidence by adopting active learning methods that frequently require students to work in teams, students with strong teaming skills may find themselves more prepared for team-based classroom environments. It is very plausible that teaming-prepared students will learn more from such team-based learning experiences.

While educational psychology is still actively exploring methods and outcomes in collaborative learning, it has been shown that some between-team differences in generating correct problem solutions in mathematics can be dependent on the nature of team interaction (Barron, 2003), with more teaming-capable teams outperforming comparable groups who failed to coordinate well. Additional educational psychology literature has shown that students in collaborative learning teams acquire knowledge from teammates (Jeong & Chi, 2007), learn more when explaining work to peers (Coleman, 1998), and are more successful in creating abstract representations of situations of interest when discussing them with teammates (Schwartz, 1995). These works taken together suggest the idea that students who are more skilled at teaming behaviors may derive more learning from team-based learning activities - at least when the materials and tasks are well-suited for teamwork (Sears & Reagin, 2013). According to Sears and Reagin (2013), tasks well-suited for learning in teams tend to be:

- More challenging or complex than an individual student is likely be able to accomplish on their own (increasing the need for students to work together).
- Demonstrable or explicable, in the sense that it is possible for one student to demonstrate a good answer or a portion of it and convince other students.
- Possessed of complimentary roles, giving each student reasons to participate and areas where their participation is expected. These complimentary roles may revolve around knowledge or resources assigned to specific students without which the overall task cannot be completed, or around specific actions to be taken by individual students (such as recorder or timekeeper).

It is noted that these conditions are found in much of the work undertaken by teams in ENGR131 and ENGR141, most notably the large team design projects.

Preparing students to work in teams successfully will also likely reduce the frequency and severity of individual students holding back learning activities by disrupting their group or team. The negative outcomes of such disruptive behavior are illustrated by Felps, Mitchell, and Byington (2006) in their work on the effects of 'bad apples' on the atmosphere and performance of teams, along with Hsiung's work showing that dysfunctional cooperative learning teams could be identified through examination of individual exam scores (Hsiung, 2010).

It will be noted here that while dedicated courses in engineering teaming have been developed (Seat & Lord, 1999; Seat et al., 2001), the majority of engineering courses with learning objectives in teaming skills development integrate teaming content and practice into existing courses also covering other engineering topics, as in E. A. Adams (2014) and Ostafichuk, Hodgson, Sophie Bartek, and Naylor (2010). Team projects are commonly but not ubiquitously found in such integration efforts.

In summary, there exists both a compelling and documented need for engineering graduates with strong teaming skills in industry and reason to believe that teaming skills do and will increasingly benefit engineering students in learning technical engineering content. Many faculty integrate teaming skills development into engineering courses not solely dedicated to teaming. This background informs this study's efforts to develop, implement, and assess a new teaming skills development intervention.

2.2 Review of Metacognition Research

The intervention employed in this study targeted metacognitive development in students with a special focus on awareness and action relating to teaming skills use, as previously discussed. This section broadly reviews the development and status of research literature in metacognition, laying the groundwork for deeper discussion of specific aspects of metacognition of high relevance to this study in later sections.

Metacognition at the broadest level is often defined as 'thinking about thinking' and though various authors employ different vocabulary, the conceptions of metacognition held by researchers "...all emphasize the role of executive processes in the overseeing and regulation of cognitive processes." (Livingston, 2003, p. 3)

While the concept of metacognition has existed in human thought and writing at least as far back as Plato (Spearman, 1923, p. 52), metacognition as a research topic was pioneered in a series of papers by John Flavell in the 1970's (Flavell, 1970, 1976, 1979) and has been an active area of research since that time. Flavell's widely cited 1979 paper defined metacognition to be "knowledge and cognition about cognitive phenomena" and cognitive monitoring to be "monitoring of their (one's) own memory, comprehension, and other cognitive enterprises" (p. 906) before going on to identify subcomponents of these ideas and provide extensive examples. In the years since then, "Flavell's definition was followed by numerous others, often portraying different emphases on (or different under-standing of) mechanisms and processes associated with metacognition" (sic) according to Georghiades (2004, p. 365). Tarricone (2011b) provided extensive analysis of various "key models" (p. 127) of metacognition, including concept maps for each (2011b, pp. 132-154). In addition to Flavell, Tarricone identified the work of Brown (Brown, 1978, 1981) , Borkowski (Borkowski, 1985; Borkowski, Carr, & Pressley, 1987), and Kuhn (Kuhn, 1999, 2000a, 2000b) as providing "important conceptual contributions to metacognition" (p. 127). Brown's work, appearing early in the development of metacognition as a research field, helped define the construct of metacognition alongside Flavell. Borkowski contributed concepts of metacognitive strategy knowledge, or knowing about methods for managing thinking, to the overall definition of metacognition. Kuhn's work situated metacognition as part of the broader area of meta-knowing.

Despite the introduction of different models, nomenclature, and emphases, the major elements of Flavell's 1979 characterizations survived with few fundamental changes and the field of metacognitive research approached a theoretical consensus more detailed than but essentially very similar to what Flavell originally proposed. The model used in this study clearly descends from Flavell's work, as is discussed later. Nelson wrote in 1998 that there was "an ongoing shift from theory to practice" (p. ix) in metacognition, noting in Hacker, Dunlosky, and Graesser (1998):

> While it would be incorrect to think that the theories of metacognition are currently so highly developed that the applications to education are straightforward, it would also be incorrect to assume that our current ideas about metacognition are so fragmented and poorly developed that any application to education would be premature (p. ix).

Embodying this idea, many works in metacognition subsequent to Nelson were either directly applied to education, as in Donald (2002), Gilbert (2005), Mithaug, Mithaug, Agrain, Martin, and Wehmeyer (2007), and Nessel and Graham (2006), or investigated metacognitive processes from the perspective of cognitive psychology as in Alter, Oppenheimer, Epley, and Eyre (2007) or Thompson, Prowse Turner, and Pennycook (2011). Works applying metacognitive pedagogies in engineering education also began to appear (Boiarsky, 2004; Case, Gunstone, & Lewis, 2001; Cunningham, Matusovich, Hunter, & McCord, 2015; Newell, Dahm, Harvey, & Newell, 2004) alongside uses of metacognition in related subject areas like programming (Breed, Mentz, & Van der Westhuizen, 2014) and science education (Zohar & Barzilai, 2013).

Summarizing the works in engineering education, Boiarsky's paper discusses the use of reflection to support transfer of student knowledge between similar situations in engineering writing. Case, Gunstone, and Lewis used reflective journaling to help students improve their learning methods and increase conceptual understanding in chemical engineering coursework. Cunningham, Matusovich, Hunter, and McCord announced a project to develop scalable teaching and assessment methods for metacognition in engineering but provided limited details. Newell, Dahm, Harvey, and Newell used an assignment analogous to ENGR131 and ENGR141's Code of Cooperation along with learning styles inventories and team discussions on potential sources of conflict to improve team environment and function in long-term chemical engineering projects for junior and senior-level students. However, Newell et al. (2004) did not target metacognitive awareness of teaming opportunities or metacognitive development.

Returning to the discussion of metacognition as a whole, the relatively mature status of metacognition is reflected by Tarricone's synthesis of previous works, which yielded a "final taxonomy of metacognition" (Tarricone, 2011b, p. 193). It is a synthesis of previous metacognition literature that presents a unified and detailed framework for understanding different elements of metacognition and the relationships between them. Therefore, "a comprehensive understanding of the construct" of metacognition (Tarricone, 2011b, p. 220) is available for use by educators, including engineering educators interested in developing teaming skills performance. The highest level of the taxonomy includes “knowledge of cognition” and “regulation of cognition” (Tarricone, 2011b, p. 192). It is noted that the two top level categories are essentially the same as those laid out by Flavell in 1979. As might be expected of a construct detailing human thought, the taxonomy of metacognition is complex, with more than 80 identified elements. A figure showing the complete taxonomy is available online from the book's publisher (Tarricone, 2011a) and is too large to reproduce here in full. A partial representation can be seen in Figure 1 and Figure 2. Note that the representations of the taxonomy given in these figures omit the final and most detailed level of the hierarchy, eliminate cross-references, and simplify some of the label text to permit more compact diagrams. These alterations may make it more difficult for the reader to determine the differences between elements of the taxonomy with similar names. This literature review briefly covers the structure of the taxonomy, before focusing on the elements of the taxonomy of central relevance to this study and how they might be employed to promote engineering teaming.

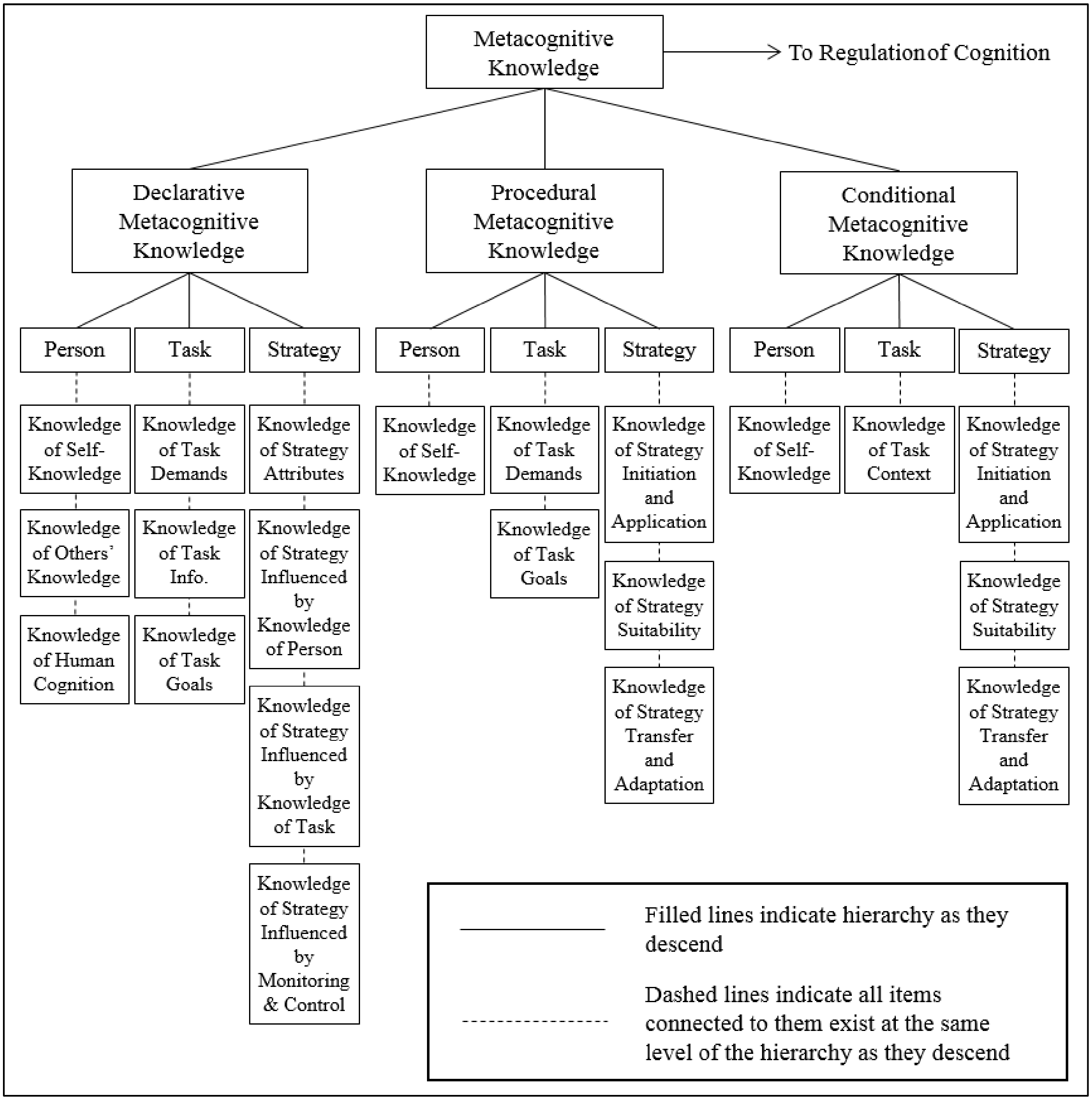


Figure 1: Structure of Metacognitive Knowledge

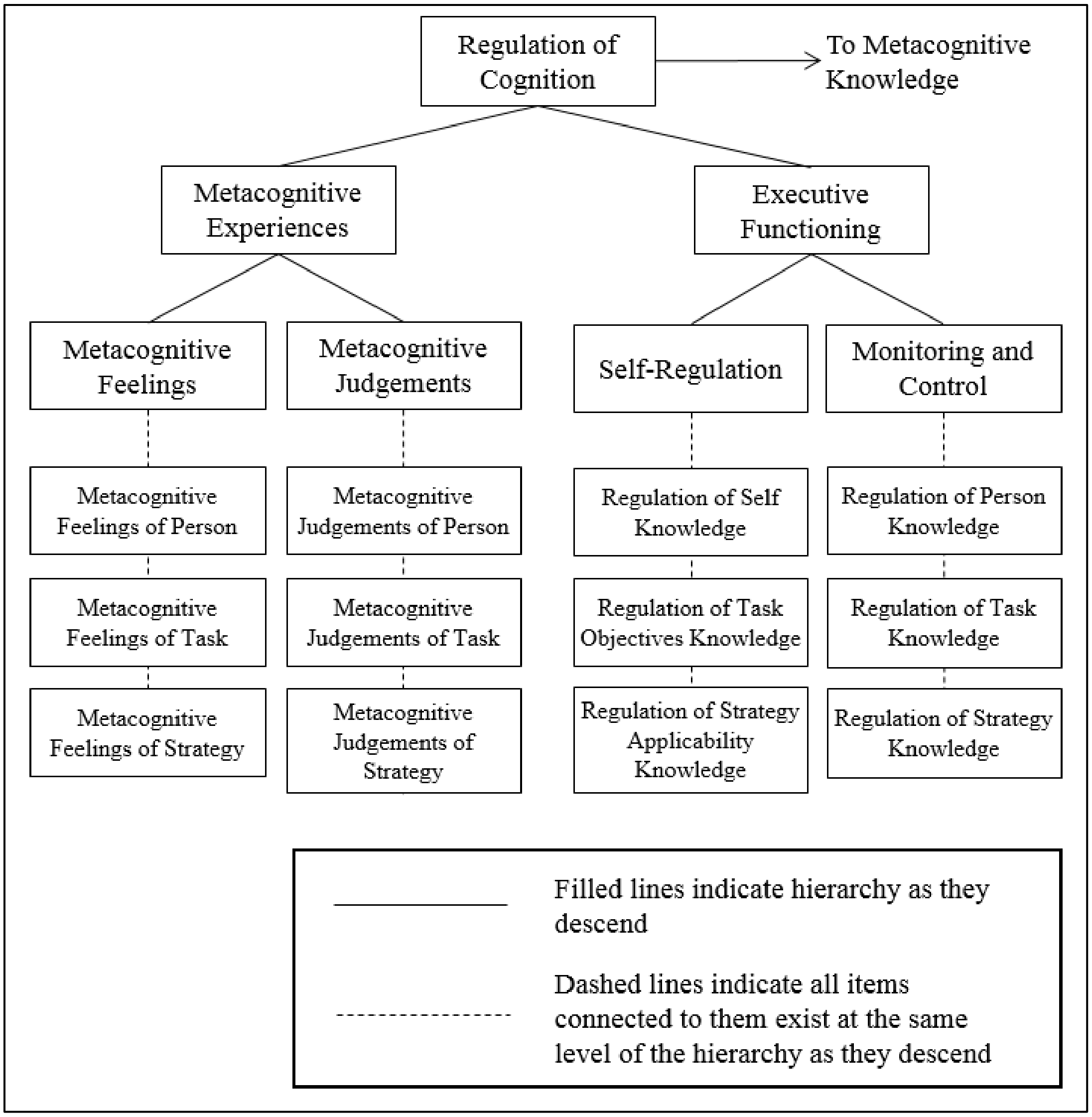


Figure 2: Regulation of Cognition

2.3 The Taxonomy of Metacognition

This section briefly summarizes the contents of the Taxonomy of Metacognition produced by Tarricone (2011b), used as the source for terminology and understanding of the structure of metacognition in this study. Some examples are provided to connect terms to real-world situations. Reference to the full, published taxonomy is recommended for readers seeking a deeper understanding of the structure of metacognition. Throughout sections 2.3, 2.4, and 2.5 the names of taxonomic categories of metacognition will appear in italics to assist in referencing Figure 1 and Figure 2.

The highest level of the taxonomy includes the categories of knowledge of cognition (or *metacognitive knowledge*) and *regulation of cognition* (Tarricone, 2011b, p. 192). Metacognition in the broadest sense can feature either of these categories or the interplay of both. Given the large number of different types of *metacognitive knowledge* and forms of *regulation of cognition* in the taxonomy, the potential breadth of this interplay is staggering. The category of *metacognitive knowledge* is introduced first.

2.3.1 Metacognitive Knowledge

It is important to differentiate knowledge of cognition, more compactly named *metacognitive knowledge*, from purely cognitive knowledge. Cognitive knowledge is essentially any knowledge that is not about thinking or learning. *Metacognitive knowledge* must be about thinking or learning in some meaningful way. For instance, ‘Student A’s report shows that she is capable of synthesizing information from many different sources’ states a fact about Student A’s ability to think, and is therefore metacognitive in nature. The statement ‘Student A’s report was submitted late’ does not express any facts about Student A’s thinking, and reflects the speaker’s cognitive knowledge, not their *metacognitive knowledge*. *Metacognitive knowledge* can also be subjective or false (Tarricone, 2011b, p. 157). Thus both the idea that ‘human cognition is well-understood’ and the idea that ‘human cognition is not well-understood’ could both qualify as *metacognitive knowledge*.

Metacognitive knowledge can be declarative "knowing about knowing", procedural "knowing how to know", or conditional "knowing when, where, and why" about persons, tasks, or strategies (Tarricone, 2011b, pp. 194-195). This review does not cover all categories of the taxonomy, but the examples provided below may be helpful in illustrating some of the high-level differences. For these examples, a hypothetical student has been given instruction in solving a particular type of equation, and thereafter has the following thoughts:

1. *"The homework assignment that requires solving these equations is due tomorrow."*

    This fact is not about thinking. It is cognitive knowledge.

2. *"I do not understand the techniques taught in class. I am confused."*

    These facts relate to the status of the student's mind and thoughts – and have no conditional or procedural characteristics – so they are *declarative metacognitive knowledge*. They are about a person rather than a task or strategy, so they are *declarative metacognitive knowledge of person*.

3. *"Figuring out these problems is really tough."*

    This is also a fact about thinking, without procedural or conditional implications, and it is therefore *declarative metacognitive knowledge*. Recall that *metacognitive knowledge* can be subjective or false. As this fact relates to a cognitive task (figuring out these problems) rather than to a person or strategy, it is *declarative metacognitive knowledge of task*.

4. *"If I just sit here and look at the examples like I'm doing right now, I am not going to be successful in figuring this out."*

Under the current cognitive conditions (confusion, lack of progress), the student observes the fact that the current thought strategy (reviewing examples) is not appropriate. The fact is metacognitive, conditional and relates to a specific strategy - it is *conditional metacognitive knowledge of strategy*.

5. *"I need to know how to break the overall solution down into smaller steps."*
    a. The student is thinking about what procedures are required to complete a cognitive task – thinking about how to know. This is *procedural metacognitive knowledge*. As the fact relates to the nature of a task, is it *procedural metacognitive knowledge of task*.

2.3.2 Regulation of Cognition

For an act of regulation to be metacognitive it must regulate, or at least influence, thought or the mind. Therefore, human acts of regulation such as maintaining heartbeat, pouring water into a glass to an appropriate level of fullness, or completing a multi-step process by rote are not typically metacognitive in a given moment. One example of *regulation of cognition* is planning to study for an exam in a quiet environment to eliminate distractions. Note that the idea 'studying in a quiet environment may limit distractions and improve cognitive function' is *metacognitive knowledge* – it is the mental action taken to determine and enact the plan that constitutes the *regulation of cognition*. Another example of *regulation of cognition* could be becoming aware of personal anger, connecting that *awareness* with the *metacognitive knowledge* that anger that can skew decision-making, and abstaining from making decisions until the anger has subsided.

Further examples accompany discussion of the subcategories of *regulation of cognition*. *Recognition of cognition* is composed of the child categories of *executive functioning* and *metacognitive experiences*. *Executive functioning* is discussed first.

2.3.2.1 Executive Functioning

Executive functioning is composed of *monitoring and control* and *self-regulation*. *Monitoring and control* is more active and task-centric, focusing on “evaluation and control” and “predicting, planning, cognitive monitoring, diagnosing, regulating, checking, and evaluating learning processes, difficulties, and outcomes” (Tarricone, 2011b, p. 166). *Self-regulation* is a metacognitive subset of the larger overall process of self-regulation and is more reflective, touching on “self-awareness, self-judgment, self-concept and self-efficacy” along with “motivational elements” (Tarricone, 2011b, p. 169). However, the distinction between *monitoring and control* and *self-regulation* is in some places the narrowest in the taxonomy, especially when focused on the *person* dimension as opposed to that of *task* or *strategy*. Some elements of these two categories appear identical or interchangeable - compare “(monitoring and control) involves metacognitive processes that facilitate and support the evaluation and control of the learning process” (Tarricone, 2011b, p. 166) with “self-regulation involves processes such as control, monitoring, and regulation of learning processes” (Tarricone, 2011b, p. 168). The difficulty in distinguishing between the two subcategories of *executive functioning* seems primarily due to the close interaction between the two subcategories. It should be noted that self-regulation extends considerably beyond the borders of metacognition, including self-efficacy and self-esteem among other areas (Tarricone, 2011b, p. 168). The taxonomy only deals with the metacognitive aspects of *self-regulation*.

Examples of *monitoring and control* include assessing whether a studying technique is effective, noting an error in thought and correcting it, and evaluating whether something has been learned sufficiently well or not. Examples of metacognitive *self-regulation* include noting that fatigue is inhibiting clear thought, updating one's self-efficacy in problem-solving after completing a difficult task, or using internal dialogue to 'tell' oneself how to proceed.

2.3.2.2 Metacognitive Experiences

Metacognitive experiences interact closely with *monitoring and control* and can be further subdivided into *metacognitive feelings* and *metacognitive judgements*. *Metacognitive feelings* are mental experiences distinct from emotions that are metacognitive in nature. For example, a student feeling that she is (or is not) confused, that a task is familiar (or unfamiliar), or a task is difficult (or easy) are all *metacognitive feelings*. It should be noted that *metacognitive feelings* can be implicit (a person may be confused without consciously noting that this has occurred) or explicit, which is often linked to mental or physical vocalization of the *metacognitive experience*. *Metacognitive feelings* of particular relevance to teaming skills employment include feelings of confusion, dissatisfaction, and unfamiliarity when engaged in teamwork. *Awareness* of these *metacognitive feelings* could prompt teaming skills employment leading to practice and improvement.

Metacognitive judgements are judgements made about cognitive enterprises, sometimes in response to *metacognitive feelings*. Example *metacognitive judgments* include the "judgment of memory correctness", "judgment or estimate of learning", "judgment or estimate of solution correctness", context-specific judgments of difficulty, and "judgments or estimates of effort expenditure" (Tarricone, 2011b, p. 212). Examples of a teaming-skills specific to *metacognitive judgement* would be the judgement by a student that they have correctly remembered target teaming skills or that they have understood the communications of a teammate.

2.4 Targeted Elements of Metacognition

It is likely that educational implications exist relating to most if not all of the many metacognitive processes in the taxonomy. However, this review emphasizes aspects of metacognition most relevant to the experimental intervention through the five-step process derived from the AIDA model as discussed in Section 1.4. To support the review of methods that can promote metacognitive growth and development in Section 2.5, this section briefly identifies and discusses areas of metacognitive occurrences and development relevant to the intervention. Cognitive knowledge or other aspects of the intervention that are not metacognitive in nature are largely omitted.

2.4.1 Awareness

Increasing the *awareness* of students with regard to opportunities to practice teaming skills was a key part of the experimental intervention. ‘Awareness’, however, is a broad term with many potential meanings in the context of metacognition. In this case, awareness is ideally conscious recognition of specific opportunities to employ and improve teaming skills in the team environment. This might be a mental statement along the lines of ‘Now would be a good time to check on whether I understood what my teammate is trying to say’. A conscious declaration of this sort constitutes student-constructed *metacognitive knowledge*.

However, before arriving at conscious awareness, at least some students pass through *metacognitive experiences* or *metacognitive judgements*, triggering or cuing higher orders of thought through *monitoring and control*. In the example above, the student might feel a lack of confidence in their own understanding. This is a *metacognitive experience*. A similar *metacognitive judgement* could be an estimate by the student that there is only a moderate probability that their understanding is correct.

Monitoring and control can direct cognition to subject *metacognitive feelings* or *judgements* to further thought, employ existing metacognitive and cognitive knowledge, and potentially create the actionable *metacognitive knowledge* sought by the intervention. For instance, the student could combine a *metacognitive feeling* of uncertainty with the *metacognitive knowledge* that a feeling of uncertainty is an appropriate prompt to regulate thought and action, creating the new *metacognitive knowledge* that ‘now would be a good time to check whether I understood that’. Therefore, the promotion of the target *awareness* involves *metacognitive feelings* and *judgements*, *monitoring and control*, and *metacognitive knowledge*.

2.4.2 Motivation

As previously discussed, the primary method employed by the experimental intervention to support *motivation* is enhanced psychological safety, and the primary goal of this step is for students to employ teaming skills in situations when they previously would not have taken any action. Psychological safety is reviewed in sections 2.6 and 2.7. However, some *metacognitive knowledge* potentially supports *motivation* and is discussed here as well.

Declarative metacognitive knowledge relevant to *motivation* includes the student's knowledge of their own motivation to acquire teaming skills. *Procedural metacognitive knowledge* includes the fact that the procedure to learn about teaming and develop teaming skills was to engage fully with the teaming activities presented by the class. Similarly, motivating *conditional metacognitive knowledge* was that students should attempt to use and improve their teaming skills whenever they interact with their teammates during team-based activities. Therefore, the promotion of *motivation* in the target context includes *declarative*, *procedural*, and *conditional metacognitive knowledge*.

2.4.3 Selection

During the *selection* step, students use what they know or believe about their own cognition, the cognition of others, the situation in the team, and cognitive knowledge of teaming skills to select an appropriate teaming skill to employ. This step may be very quick, as a student chooses in an instant between communicating feedback gently or harshly. This step may also be extended in time, as a student or students grapple with ongoing team problems and create a plan to address those problems. The experimental intervention primarily supports *metacognitive knowledge* for this step of the process. Specifically, *declarative metacognitive knowledge* of sources for teaming skills implementation strategies. This step may draw on *declarative metacognitive knowledge* of a particular student's strengths and weaknesses in teaming and *conditional metacognitive knowledge* related to the team's operations.

2.4.4 Implementation

Metacognition during the actual *implementation* of teaming skills was not a target of the experimental intervention, which focused on getting students to notice opportunities to employ teaming skills and making the decision to do so. It is likely that some students evaluated their own thinking, or that of teammates, during their *implementation* of teaming skills, which would involve metacognition. These metacognitive behaviors could potentially improve teaming skills performance. However, the goals of the intervention were satisfied if the *implementation* of teaming skills did not include metacognition, as long as metacognition supported the previous steps of *awareness*, *motivation,* and *selection*. The intervention set no metacognitive targets for this step of the process.

2.4.5 Reflection

The *reflection* step asks students to perform metacognitive self-evaluation (part of *monitoring and control*) and *self-regulation* in order to update existing *metacognitive knowledge* and construct new *metacognitive knowledge*. For instance, students may construct *declarative metacognitive knowledge* of their own strengths and weaknesses in teaming and *conditional metacognitive knowledge* related to the team's operations. By updating and constructing *metacognitive knowledge* (and cognitive knowledge) relevant to teaming skills employment, students are better prepared to act upon future teaming skills employment opportunities.

2.5 Methods for Metacognitive Development

Despite the availability of a taxonomy of metacognition, literature speaking to developmental methods for specific aspects of metacognition identified in the taxonomy are typically not available at this time. Studies and reviews advocating methods for improving metacognition often predate the taxonomy and feature varyingly precise definitions of metacognition. Educational techniques are often presented as supporting metacognition without being tied to specific metacognitive processes. Therefore, to target development of specific metacognitive processes, some interpretation of prior work is required in terms of likely targets and effects. It is also noted that Tarricone identified more than 50 avenues for future research in understanding and developing metacognition (Tarricone, 2011b, pp. 215-220), reflecting the breadth of metacognition research work not yet done, much of it in methods to support the development of specific aspects of metacognition. This study to some extent answers her call to use the framework "to form the basis of new and exciting empirical studies" (Tarricone, 2011b, p. 220).

2.5.1 Development of Metacognitive Knowledge

Metacognitive knowledge is a subset of cognitive knowledge and remains amenable to development through standard undergraduate pedagogic techniques. *Metacognitive knowledge* is factual in nature (though it can be subjective or false), whether the facts have to do with the student personally or tasks and strategies they are exposed to. For instance, students could gain *metacognitive knowledge of strategy* by being informed of resources they could use to identify basic teaming skills and strategies, such as the CATME website and the course instructor. This is *metacognitive knowledge* because it relates to changing and updating knowledge or thinking. In an example from the literature, Schraw (2001, p. 119) recommends that "teachers to take the time to discuss the importance of metacognitive knowledge and regulation" and to accompany that time with "group discussion and reflection". No special techniques to support learning of this content knowledge beyond those commonly employed in undergraduate education are required. The acquisition of *metacognitive knowledge* could employ techniques supported by authoritative sources such as Bransford, Brown, and Cocking (2000) and (Wankat & Oreovicz, 2014) when designed in an intentional and appropriate way per Hansen (2012). Cognitive knowledge may be required to enable *metacognitive knowledge* to be useful. For instance, *metacognitive knowledge* about how to gain more information about teaming is not necessarily useful unless follow-through to actually learn the additional teaming information occurs.

2.5.2 Development of Regulation of Cognition

Metacognitive feelings, *metacognitive judgements*, *monitoring and control*, and *self-regulation* are all aspects of *regulation of cognition* of relevance to the experimental intervention. However, the experimental intervention, while making use of *metacognitive feelings* and *metacognitive judgements*, did not seek to promote them. These feelings and judgements are naturally part of the experience of human consciousness (falling into the taxonomic category *metacognitive experiences*), and it is not clear that promoting them is possible, or would be useful. Therefore, the development of *metacognitive feelings* and *metacognitive judgements* is not discussed. This section reviews methods supporting the development of metacognitive *monitoring and control* and *metacognitive self-regulation*, beginning with *monitoring and control*.

Schraw presents several techniques for promoting student *monitoring and control* in the classroom environment (Schraw, 2001). One method of potential interest to this study was the classroom use of "strategy evaluation matrices", typically abbreviated as SEM's. As the acronym 'SEM' is most often understood in engineering education to mean structural equation modeling, in this dissertation strategy evaluation matrices will be abbreviated as 'SM'. SM's list potential strategies to be employed in a task along with brief snippets summarizing when, where, and why to use them. Schraw recommends rotating the class focus of attention to different strategies at different points in the course (potentially one strategy per class or week) but using the same overall SM for extended periods of time so that students become very familiar with the information contained. While the sample SM given by Schraw presents strategies for reading comprehension, for this experimental intervention a teaming skills SM was developed for use in engineering teaming contexts. In this case, the ‘strategies’ were manifestations of specific teaming skills of interest. This technique is said to pair well with regular group reflection and self-evaluation and to “promote explicit metacognitive awareness” of the contents of the SM (Schraw, 2001, p. 120).

Schraw also recommends the use of a “regulatory checklist” that scaffolds monitoring and control of cognition. The example checklist includes questions like “Do I have a clear understanding of what I am doing?” and “Am I reaching my goals?” (Schraw, 2001, p. 121). Such a checklist is provided as a reference to students who are then prompted to review it periodically or as needed. Schraw’s example RC is presented in the context of a solitary task or problem, which is perhaps not ideally suited to teaming skills development on engineering teams, as students in such teams need to allocate effort to multiple problems at once, including the problem of practicing teaming skills. However, one or more checklists could be created to prompt consideration and review of teaming-relevant metacognitive feelings or situations. This would likely increase metacognitive awareness of the listed conditions if used over time. The use of an RC can be complimentary to the use of an SM.

In a potential contrast to Schraw, Gourgey (2001, p. 84) noted that “Many studies have found that metacognitive activities that are externally imposed (i.e., the teacher generates questions or dictates strategies to use for clarification) are less effective than those generated by the students themselves” and stated that “it is recommended that instruction encourage students to generate and use their own strategies and self-questions; this approach has been found more effective for promoting independent learning and transfer.” This would seem to discourage the use of the scaffolded methods suggested by

Schraw. However, a middle ground between these positions is likely tenable upon closer inspection. Gourgey’s comments relate to metacognition in reading comprehension and cite the effects of question generation training in the same context (Davey & McBride, 1986). The students in the Davey and McBride study were more successful in learning about the contents of the readings when they generated their own questions about the readings, but their process for reviewing the readings by generating questions had been substantially scaffolded by the training provided. This process would appear to be scaffolding some methods for *monitoring and control* in support of learning in another area, which is comparable to the processes suggested by Schraw. However, in keeping with the principles of student-centered, active learning, it is plausible that avoiding an excess of pre-made scaffolding and tasking students with creating or extending their own scaffolding for the processes of *monitoring and control* would promote greater engagement and learning.

Approaching *monitoring and control* from a different angle, Alter et al. (2007) found that students were more often induced into deliberate, analytic thinking (characteristic of high levels of *monitoring and control*) when faced with tasks possessing two characteristics. The first characteristic was being more difficult. The second characteristic was possessing some aspect that triggered in students’ minds the impression that an instinctive or heuristic response might be in error. One example given was problems made deliberately difficult to read. While not presented using the language of the taxonomy of metacognition, this appears to be a case of *metacognitive feelings* and *metacognitive judgements* triggering *monitoring and control*. While it is not necessarily desirable to make all course materials related to teaming difficult to read or emplace similar artificial barriers, thought can be given in the design of learning experiences towards methods that might be employed to disrupt heuristic thought and direct students towards controlled, analytic thought. The SM and RC prompts suggested by Schraw may serve this purpose, as could more generic prompts by a teammate, TA, or instructor to pause and reflect on or record recent events in the team. It is likely that creating assignments or situations that are too difficult or too complicated may have a negative effect on student learning through a surplus of cognitive load (Sweller, 1988), but tasks that are too simple or too easy can also impair efficient learning in teams (Sears & Reagin, 2013).

Moving on to metacognitive *self-regulation, reflection* is recommended by virtually all sources reviewed for promoting metacognition, including authors working in engineering contexts (Boiarsky, 2004; Case et al., 2001). *Reflection* promotes *self-regulation* via processing and updating information about the self. Reflections in the undergraduate context often take the form of group or class discussions or individual writing assignments or prompts and may be optional or mandatory. While post-hoc reflection does not directly stimulate awareness of teaming skills employment opportunities, it is widely noted for the potential to improve student *motivation*, update *metacognitive knowledge,* and promote *awareness* or *monitoring and control* in the future.

2.5.3 Summary of Methods for Metacognitive Development

In summary, a review of the literature on the development of metacognition identified information and methods applicable to the goals of this study. First, the development of *metacognitive knowledge* can be supported by methods suitable for cognitive knowledge, especially reflection. Second, strategy evaluation matrices and regulation checklists may scaffold metacognitive *monitoring and control*. Metacognition literature and active learning principles suggest that encouraging students to construct their own or elaborate upon provided scaffolding may increase the effectiveness of *monitoring and control* scaffolding. Third, the capacity for materials that are difficult, confusing, or otherwise work to break the flow of heuristic thinking may be applicable to the promotion of metacognitive thought. While it is not seen as desirable to integrate deliberately obnoxious learning materials into the curriculum to serve this purpose, periodic reflective prompts, prompts to engage with an SM or RC, or other minor disruptions to in-class student working periods could be contemplated to serve a similar purpose. Finally, there is widespread support for regular post-hoc reflections on learning to promote metacognitive *self-regulation*, in addition to the benefits to *metacognitive knowledge*.

2.6 Psychological Safety

Psychological safety has been identified as an important factor in team performance, and is principally of interest to this study as a mediating factor in student motivation to employ teaming skills. Psychological safety is defined by Edmondson as “a shared belief that the team is safe for interpersonal risk taking” and adds that “the term is meant to suggest neither a careless sense of permissiveness, nor an unrelentingly positive affect but, rather, a sense of confidence that the team will not embarrass, reject, or punish someone for speaking up” (1999, p. 354). Bradley, Postlethwaite, Klotz, Hamdani, and Brown (2012) state that “psychological safety may amplify the involvement of each team member and the intensity of interaction among teammates without endangering the harmony of the team, thereby increasing team performance” (p. 151).

The importance of balancing challenge and personal security for high performance was identified in organizational research in the mid-20th century (Pelz & Andrews, 1966) but has recently gained prominence in the public eye after a study by Google articulating the importance of psychological safety was reported in the New York Times (Duhigg, 2016). A detailed review of the development and current status of psychological safety research literature can be found in Edmondson and Lei (2014), but there are several key takeaways. These include the fact that “psychological safety is associated with learning” at individual, group, and organizational levels, that there are “clear and significant relationships between psychological safety and performance” (more psychological safety tends to lead to higher group and organizational performance), and that “psychological safety in organizational life can best be considered a phenomenon that lives at the group level.” (Edmondson & Lei, 2014, p. 37) The first two outcomes of higher psychological safety are clearly desirable in classroom environments. For the third takeaway, it is not entirely clear whether a class corresponds most closely to an organization or group. It seems likely that student teams form the most important group for psychological safety during team activities but that the larger class might be ‘the group’ when classwide discussions or Q&A are ongoing.

In the context of the five-stage process for student teaming skills practice (*awareness*, *motivation*, *selection*, *implementation*, and *reflection*), psychological safety is most likely to support *motivation*. With a higher level of psychological safety, team members are more likely to 'stick their necks out' to personally work to address problems or opportunities the team may be facing. As previously mentioned, this is in accordance with conflict avoidance literature, which shows that conflict can lead to team members interacting less as “team members try to disengage from those with whom they experience conflict, and further limit their interactions” (Langfred & Moye, 2014, p. 33). Team members are also likely to be more comfortable attempting activities that they are aware they may not be good at in a team with high psychological safety. Many such issues will either deal directly with teaming, or else either require or benefit from the application of teaming skills. For instance, a less-dominant member of the team may take action to regulate the behavior of a more-dominant team member knowing that the team environment supports the giving of appropriate feedback. Thus, an increase of psychological safety means that more opportunities to practice teaming skills will be acted upon by students due to their greater *motivation*, resulting in greater practice overall and a consequent improvement in teaming skills performance. Recalling that the intervention developed in this study sought to increase the number of opportunities perceived by students to employ teaming skills, the synergy with psychological safety in promoting teaming skills performance is clear.

2.7 Developing Psychological Safety in Teams

This section reviews some recommended methods for creating working environments with high psychological safety. Principles for creating groups with high psychological safety have largely been developed and tested in professional, rather than scholastic, environments. However, the majority of techniques appear to be translatable to undergraduate team environments as most behaviors of individuals and teams relevant to psychological safety in organizational research have parallels in engineering undergraduate student teams. These parallels and their implications will become clear through the following discussion on developing psychological safety in student teams.

Edmonson identifies several practices for promoting psychological safety (Edmondson, 2002). Some behaviors are explicitly associated with team leaders, including taking communal feedback in decision making, accessibility of team leaders, and acknowledgement of failure or fallibility by team leaders. While student teams in ENGR131 and ENGR141 do not have explicit leaders, all students share some responsibility for the management of the team. Teams are already encouraged to make evidence-based communal decisions but further emphasis could be placed on this during teaming instruction and self-evaluation/reflection. Accessibility is one goal in the use of CATME's team formation capabilities that work to ensure that student teams have time out-of-class to meet, but teams could also be encouraged to integrate thoughts on mutual accessibility into their Codes of Cooperation and reflect on the extent to which they have promoted a healthy level of accessibility. Finally, the fallibility of all student team members could be addressed in a supportive way through class discussion, group reflection, or other activities. Edmonson relates that some organizations schedule times to admit mistakes and learn from them (Edmondson, 2002, p. 21). While asking students to admit mistakes in their teaming to their teammates in too formal a way or on too forced a schedule may incite resistance, scaffolding discussion of how teams wish to discuss or deal with mistakes and apologies into the Code of Cooperation assignments and making errors and apologies part of reflection prompts may encourage teams to develop methods to address team member fallibility in positive ways that promote psychological safety.

Edmonson also recommends learning practices widely employed in education, such as explicit goals and periods of reflection where progress is evaluated against the explicit learning goals, for the cultivation of psychological safety and learning. In classroom settings, this likely includes discussion of course norms and expectations with regards to behavior that could support or detract from psychological safety. Both ENGR131 and ENRG141 have some discussion of course norms and expectations, but psychological safety could likely be more explicitly and firmly discussed and motivated as a goal.

The role of the course instructor as a leader in the classroom should also be considered. An instructor who models behavior that contributes to high psychological safety, such as consistently treating questions as worthwhile and students with respect, being accessible for questions, and relating a certain amount of their fallible humanity to the class could go far in establishing course norms. Explicitly stating such goals and class norms, abiding by them, and referencing them when appropriate would also likely promote an overall atmosphere of psychological safety. A method of reporting problems in teams or in the class that detract from psychological safety, potentially anonymously (unlike CATME), might allow the instructor to take action to enforce or reinforce course norms.

In summary, to promote psychological safety in the classroom and student working teams, clear goals and norms relating to psychological safety should be set for the class and by each team and adhered to, methods for supporting communal decision making on teams should be selected, student teams should be encouraged to develop positive methods for admitting and addressing mistakes and fallibility, and accessibility of students on teams to each other and the instructor to students should be considered.

CHAPTER 3. METHODS

This chapter discusses the study's sample, the experimental methods and analyses, and some relevant standard practices of each class that could to influence the results. After discussing the sample characteristics, the remainder of the chapter is organized by research question, in the order that the research questions appear in Section 1.2. Methods relevant to a set of case studies appear after the discussion of all four research questions. These case studies do not directly address the research questions, but provide examples for discussion and context. It is noted that some data was collected during this study not relevant to the research questions; methods relating to this data collection will not be discussed outside of the case studies.

3.1 Sample

This study targeted a sample size above six hundred first-year engineering students at Purdue University across five sections of ENGR141 (roughly 250 students) and four sections of ENGR131 (roughly 600 students). A previous investigation of metacognitive performance in this FYE context (Rynearson & Hynes, 2015) found that variability in initial student teaming performance between sections of ENGR131 could be large, making it difficult to render meaningful conclusions with pre-post testing between only two sections of the course. Sampling from nine sections across two FYE courses was intended to secure the following benefits: an increased probability that the control and experimental groups will be representative of the Purdue FYE populations, a decreased probability that students in the control and experimental groups will have strongly differing characteristics, the mitigation of confounding factors including instructor performance and the time-of-day each course section is offered in relation to the others, and the opportunity to compare the effect of the intervention in the Honors (ENGR141) and non-Honors (ENGR131) student contexts.

The sections of ENGR141 and ENGR131 included in the study were not randomly selected. For ENGR141, all three course instructors supported the study in their sections. Thus, for ENGR141, the entire course population participated as either part of the experimental or control groups. The three instructors covered four sections of the course, with two instructors covering one experimental section each and a third instructor teaching the two control sections. This sampling configuration is susceptible to confounding factors such as time-of-day and instructor performance, but represents the best sample that can be achieved within real-world constraints in ENGR141.

In ENGR131, the larger number of total class sections means that the three experimental and two control sections do not comprise the entire student population. While an even larger sample would accrue additional benefits in terms of claims to a representative sample, five sections approached the maximum sample size that could be achieved without incorporating sections with additional clear confounding factors. Sections of ENGR131 serving themed learning communities along with sections taught by first-time instructors were excluded from the sample. As this study required personal and significant in-class actions at several points during the term by course instructors, it was necessary to recruit willing facilitators from among the body of course instructors. Five of the remaining seven instructors self-selected into the experimental and control categories. This self-selection represents a potentially meaningful confounding factor.

Each section of ENGR131 nominally has a population of 120 students and sections of ENGR141 typically have a population between 60 and 70 students. As noted elsewhere, the study was administered in the course of normal class operations and all students were expected to participate. Therefore, approximately 850 students were enrolled in the study as either part of the experimental or control groups. However, student attrition and noncompliance with some or all of the assessment procedures resulted in a considerably smaller number of responses. Some characteristics of the sampled sections are given in Table 3.1. The abbreviation 'EXP' is used for experimental sections and 'CTRL' for control sections.

Table 3.1 Sample Characteristics

Section	Total Students	International Students	Time of Day	Attendance	Final Grade Difference*	Final Grade Std. Dev.*
ENGR131						
EXP 1	120	24	7:30-9:30AM	96%	2.4	4.1
EXP 2	119	18	3:30-5:30PM	98%	-2.5	4.2
EXP 3	119	31	11:30-1:30PM	98%	-0.3	4.5
CTRL 1	119	14	7:30-9:30AM	95%	0.1	6.5
CTRL 2	119	15	11:30-1:30PM	98%	-0.3	5.5
ENGR141						
EXP 1	69	2	11:30-1:30PM	>95%	0.8	7.6
EXP 2	65	5	3:30-5:30PM	>95%	-0.7	8.9
CTRL 1	61	2	9:30-11:30AM	>95%	-1.1	8.6
CTRL 2	68	1	1:30-3:30PM	>95%	1.0	6.8

*ENGR131 and ENGR141 grading scales were mapped to a 100-point grade scale for these calculations

For ENGR131, attendance rates were calculated by dividing the number of marked absences by the total number of potential attendances. For ENGR141, exact attendance data was not captured by this study. However, it is known that no attendance penalties were assessed in any section, meaning that no student had more than two unexcused absences. Full attendance is typical in ENGR141. The ‘Final Grade Difference’ column shows the difference between the average of final grades across the sampled sections and each individual section’s final grades average, out of 100 points. For example, ENGR131 EXP 1 has a ‘Final Grade Difference’ of 2.4, indicating that the average final grade for this section, out of 100, was 2.4 points higher than the average of the entire ENGR131 sample. It can be seen that the sampled sections do not exhibit major differences across most categories. The experimental group in ENGR131 tended to have more international students than the control section, which could be a confounding factor.

3.2 Intervention and Assessment Outline and Schedule

This section presents tabular summaries of interventions and assessments employed in the study, along with scheduling information for each. Table 3.2 and Table 3.3 show all intervention and assessment activities, scheduling information, whether these activities applied to the experimental group, the control group, or both, and some additional information to characterize the activity. Table 3.4 groups activities and interventions chronologically by the study research question they are most closely associated with for ease of reference and discussion. This chapter will discuss intervention and assessment activities in the order given in Table 3.4. Table 3.5 shows the full text of the study research questions.

Table 3.2: ENGR131 Intervention and Assessment Schedule

Week	Start	End	Interventions and Assessments	Population	Method	Purpose
1	24-Aug	25-Aug	Intro to Teaming & Diversity	EXP, CTRL	Lecture, Discussion	Control
2	2-Sep	3-Sep	Psychological Safety & Norm Setting	EXP	Lecture	Intervention
2	2-Sep	9-Sep	Standard Code of Cooperation	CTRL	Assignment	Control
2	2-Sep	9-Sep	Revised Code of Cooperation	EXP	Assignment	Intervention
2	2-Sep	10-Sep	CATME Skills for Metacognition	EXP	Lecture	Intervention
4	16-Sep	17-Sep	Standard ENGR131 Teaming Recap	EXP, CTRL	Lecture, Discussion	Control
5	21-Sep	1-Oct	Psychological Safety Survey 1	EXP, CTRL	Out-of-class Survey	Assessment
5	23-Nov	24-Nov	Introduction of Teaming SM	EXP	Lecture	Intervention
5	23-Nov	24-Nov	SM Administration 1	EXP	In-class Worksheet	Intervention
6	30-Sep	1-Oct	Metacognitive Frequency Survey 1*	EXP, CTRL	In-class Survey	Assessment
6	2-Oct	11-Oct	CATME BARS 1	EXP, CTRL	Out-of-class Rating	Assessment
7	5-Oct	6-Oct	Standard Teaming Reflections 1	CTRL	Assignment	Control
7	5-Oct	6-Oct	Teaming & Psych Safety Reflections 1	EXP	Assignment	Intervention
10	26-Oct	27-Oct	SM Administration 2	EXP	In-class Worksheet	Intervention
10	28-Oct	29-Oct	Standard Teaming Reflections 2	CTRL	Assignment	Control
10	28-Oct	29-Oct	Teaming & Psych Safety Reflections 2	EXP	Assignment	Intervention
11	1-Nov	8-Nov	CATME BARS 2	EXP, CTRL	Out-of-class Rating	Assessment
11	2-Nov	3-Nov	Metacognitive Frequency Survey 2**	EXP, CTRL	In-class Survey	Assessment
12	9-Nov	10-Nov	SM Administration 3	EXP	In-class Worksheet	Intervention
13	16-Nov	17-Nov	Standard Teaming Reflections 3	CTRL	Assignment	Control
13	16-Nov	17-Nov	Teaming & Psych Safety Reflection 3	EXP	Assignment	Intervention
13	18-Nov	19-Nov	Metacognitive Frequency Survey 3	EXP, CTRL	In-class Survey	Assessment
14	23-Nov	3-Dec	Psychological Safety Survey 2	EXP, CTRL	Out-of-class Survey	Assessment
16	6-Dec	13-Dec	CATME BARS 3	EXP, CTRL	Out-of-class Rating	Assessment

* Section EXP 1 re-did this data point on 10/22, ** Section CTRL 2 re-did this data point on 11/10

Table 3.3: ENGR141 Intervention and Assessment Schedule

Week	Start	End	Interventions and Assessments	Population	Method	Purpose
1	28-Aug	28-Aug	Standard ENGR141 Intro to Teaming	EXP, CTRL	Lecture, Discussion	Control
1	28-Aug	28-Aug	Psychological Safety & Norm Setting	EXP	Lecture	Intervention
1	28-Aug	4-Sep	Standard Code of Cooperation	CTRL	Assignment	Control
1	28-Aug	4-Sep	Revised Code of Cooperation	EXP	Assignment	Intervention
2	4-Sep	4-Sep	CATME skills for Metacognition	EXP	Lecture	Intervention
4	18-Sep	26-Sep	Psychological Safety Survey 1	EXP, CTRL	Out-of-class Survey	Assessment
4	18-Sep	25-Sep	CATME BARS 1	EXP, CTRL	Out-of-class Rating	Assessment
5	23-Sep	23-Nov	Introduction of Teaming SM	EXP	Lecture	Intervention
5	23-Sep	23-Nov	SM Administration 1	EXP	In-class Worksheet	Intervention
6	30-Sep	30-Sep	Metacognitive Frequency Survey 1	EXP, CTRL	In-class Survey	Assessment
7	2-Oct	2-Oct	Teaming & Psych Safety Reflections 1	EXP	Assignment	Intervention
10	30-Oct	30-Oct	SM Administration 2	EXP	In-class Worksheet	Intervention
10	30-Oct	30-Oct	Teaming & Psych Safety Reflections 2	EXP	Assignment	Intervention
11	2-Nov	2-Nov	Metacognitive Frequency Survey 2	EXP, CTRL	In-class Survey	Assessment
12	11-Nov	18-Nov	CATME BARS 2	EXP, CTRL	Out-of-class Rating	Assessment
13	18-Nov	18-Nov	SM Administration 3	EXP	In-class Worksheet	Intervention
14	25-Nov	25-Nov	Metacognitive Frequency Survey 3	EXP, CTRL	In-class Survey	Assessment
15	30-Nov	30-Nov	Teaming & Psych Safety Reflection 3	EXP	Assignment	Intervention
15	30-Nov	9-Dec	Psychological Safety Survey 2	EXP, CTRL	Out-of-class Survey	Assessment
16	7-Dec	13-Dec	CATME BARS 3	EXP, CTRL	Out-of-class Rating	Assessment

Table 3.4: Interventions and Assessments by Research Question

Interventions and Assessments	Course	Population	Method	Supports Process Step(s)	Purpose
Research Question 1					
Presentation of CATME skills for metacognition	Both	EXP	Lecture	Awareness, Motivation	Intervention
Introduction of Teaming SM	Both	EXP	Lecture	Awareness, Selection	Intervention
SM Administrations 1, 2, 3	Both	EXP	In-class Worksheet	Awareness, Selection	Intervention
Metacognitive Frequency Survey 1, 2, 3	Both	EXP, CTRL	In-class Survey	Awareness	Assessment
Research Question 2					
Psychological Safety Introduction & Norm Setting	Both	EXP	Lecture	Motivation	Intervention
Standard Code of Cooperation Assignment	Both*	CTRL	Assignment		Control
Revised Code of Cooperation Assignment	Both*	EXP	Assignment	Motivation	Intervention
Psychological Safety Survey 1, 2	Both	EXP, CTRL	Out-of-class Survey	Motivation	Assessment
Research Questions 1 and 2					
Standard Teaming Reflections 1, 2, 3	ENGR131	CTRL	Assignment		Control
Teaming and Psych Safety Reflections 1, 2, 3	Both	EXP	Assignment	Reflection	Intervention
Research Question 3					
Standard Intro to Teaming & Diversity	ENGR131	EXP, CTRL	Lecture, Discussion		Control
Standard Intro to Teaming	ENGR141	EXP, CTRL	Lecture, Discussion		Control
Standard Teaming Refresher	ENGR131	EXP, CTRL	Lecture, Discussion		Control
CATME BARS 1, 2, 3	Both	EXP, CTRL	Out-of-class Rating		Assessment

*Different assignments were used in ENGR131 and ENGR141 for this purpose

Table 3.5: Research Question Numbers and Text

Research Question	Research Question Text
Research Question 1	To what extent did the experimental intervention promote the target aspects of metacognition in FYE students in the experimental group in comparison to the control group?
Research Question 2	To what extent did the experimental intervention promote psychological safety of student teams in the experimental group in comparison to the control group?
Research Question 3	In the event that the intervention succeeds in supporting more metacognitive skills growth and/or psychological safety in the experimental than the control groups, to what extent did the teaming performance of the experimental group then improve beyond that of the control group?
Research Question 4	To what extent did the improvement in the targeted metacognitive capabilities or psychological safety correlate with improved individual student teaming performance?

3.3 Interventions, Assessments, and Analyses

This section describes the interventions, assessments, and analyses employed in this study. As appropriate, the threshold for results or findings to be meaningful is also discussed. The interventions and assessments are discussed in the order given in Table 3.4, which groups them by the research question they are most closely associated with. Some interventions are associated with multiple research questions but will only be fully described the first time they are discussed. Information relating to all research questions, including the assembly of the overall data set, is reviewed before specific interventions and assessments. Discussion of research question 4, which does not appear in Table 3.4, follows discussion of research questions 1 through 3.

3.3.1 General Data Set Assembly

The overall data sets prepared for analysis for this study included results from a number of different instruments and sources. In assembling disparate data sources into an overall data file for ENGR131 and ENGR141, it should be noted that a bias towards the retention of records was in place, and few records were discarded or destroyed. In some cases, seemingly conflicting records required remediation. For instance, in ENGR131 some students appeared in different sections according to different records. In most cases, it was apparent that students had transferred between sections subsequent to the formation of the initial rosters. In such cases, CATME BARS and final grading records were used to determine which section a student attended for the majority of the term (typically, ENGR131 students are not permitted to switch sections after week 2). Also in ENGR131 records, approximately 10 students not in any initial course rosters appeared in later data (typically, new students cannot join ENGR131 after week 3). These students were added to the overall data set. Students appearing in the data set for which absolutely no data was collected during the semester (four in ENGR131 and none in ENGR141) were assumed not to have shown up for the class and were removed from the overall data set. These students appeared in some records as being fifth members of teams, which is not a typical team configuration in ENGR131. Instances of five-member teams where four members submitted data was taken as conclusive that the fifth student did not participate in the class.

3.4 Interventions, Assessment, and Analysis Primarily Associated with RQ1

Research Question 1: To what extent did the experimental intervention promote the target aspects of metacognition in FYE students in the experimental group in comparison to the control group?

Several interventions targeting metacognition in students were implemented across the semester in the experimental sections, including a discussion of CATME teaming skills, the introduction and use of a strategy evaluation matrix (SM), and reflections on teaming. Each of the interventions embodies recommendations for prompting development of the target metacognitive capabilities identified in Chapter 2. Each activity is discussed in more detail. One main assessment was administered in both the experimental and control sections, the Metacognitive Frequency Survey.

3.4.1 Presentation of CATME Skills for Metacognition

An in-class presentation on CATME teaming skills was performed in the experimental sections early in the term so that students were explicitly made aware of the skills to be improved through practice. This intervention was expected to support the *awareness* and *motivation* steps of the teaming skills practice decision model, by listing and illustrating items to be aware of an emphasizing the importance of these skills to the class. This introduction accompanied a discussion of teams and teaming in each class (see 3.6.1, 3.6.2, and 3.6.3).

This review primarily constituted cognitive knowledge and metacognitive knowledge of task. Section 2.5.1 established that standard pedagogic techniques are appropriate for metacognitive knowledge. Concepts in metacognitive development were necessarily embedded in this discussion, but not discussed explicitly. The talking points and slides were coordinated to lower variation across experimental sections, but instructors were permitted to adjust the materials to fit their format or needs. The exact content, timing, and duration of this presentation depended to some extent on the instructors, but the experimental content was expected to consume approximately five minutes of class time based on the content and number of sample slides. More intensive instruction or exercises in this material may have increased the impact of this review, but the limitations for class time allocated for experimental materials mandated a brief presentation. The sample slides appear in Appendix Figure C.7.

3.4.2 Introduction of Teaming SM

A strategy evaluation matrix (SM) (Schraw, 2001) was developed featuring teaming as the process to be regulated and can be seen in Appendix Figure B.4. The SM drew from specific CATME competencies and included information connecting strategies to potentially incite metacognitive experiences. It was expected that student use of the SM would be salutary to development of metacognitive *monitoring and control* in the target context. The cues and potential teaming actions in response to the cues listed on the SM may also reinforce *metacognitive knowledge*. Use of an RC (Schraw, 2001) might also support development of metacognitive *monitoring and control* alongside an SM or in addition to it. An RC was not developed or employed in this study to reduce the number of activities and the time required for the experimental intervention.

While personalization of SM and SM-like tools by students is noted in the literature review as a method to potentially increase efficacy (see Section 2.5.2), personalization was not a focus of this study for two reasons. First, personalization of an SM by students would require a meaningful quantity of additional student working time not available to the study. Second, differences in SM personalization between sections might contribute a confounding factor to the study. However, an SM without personalization is a legitimate tool for the promotion of metacognition. This SM was intended to support *awareness* by listing cues for teaming actions, and *selection* by suggesting potential courses of action based on those cues. *Motivation* may also be supported by listing potential benefits of implementing appropriate teaming actions. Spaces to mark occurrences of *awareness* of opportunities to employ specific teaming skills, and additional spaces to mark completed teaming actions were available. These sets of spaces were intended to suggest to students that for each instance of *awareness*, they should continue through to action, logging both. This tool was introduced by experimental section faculty at the beginning of a team-based in-class activity, in a presentation expected to take about five minutes. Sample slides for this introduction can be seen in Appendix Figure C.8.

3.4.3 SM Administrations 1, 2, 3

At three points in the semester, students in the experimental sections were asked to engage with the SM for specific periods of time (10 minutes) during in-class team-based work. As discussed in the previous section, use of the SM was intended to support *awareness, selection,* and potentially *motivation*. This exercise was introduced as brief but mandatory practice in teaming skills monitoring and control. The SM was printed on a worksheet so that students could log by hand the number of times that opportunities for or actualizations of specific teaming skills were noted as having occurred in the team. The order of SM contents was not randomized due to the logistical constraints of producing and distributing paper worksheets, and therefore the order of contents was the same for all students in all sections in this study. The limited working timespan had benefits in enhanced ability to ensure student interest and compliance. These worksheets were collected at the end of the ten minute working periods. The collection of the SM worksheets, even without a grade impact, may have increased a sense of accountability for students with respect to engaging with the SM. See Table 3.2 and

Table 3.3 for administration dates in ENGR131 and ENGR141, respectively.

The SM worksheets were also intended to support future efforts to assess metacognition with data gathered from the real-time flow of cognition and metacognition, also referred to in metacognition research as ‘online’ data (Veenman, Van Hout-Wolters, & Afflerbach, 2006, p. 9). As the collected data cannot be used to compare the experimental and control sections, it is not directly relevant to the research questions of this study, but could support future inquiries into the suitability of SM worksheets for assessing ‘online’ metacognition.

Using the SM worksheets on more occasions, potentially for longer, giving more extensive training in the use of SM’s, and rotating the order of the SM prompts could all potentially increase the impact of the intervention, and were all ruled out due to logistical constraints.

3.4.4 Teaming Reflections 1, 2, 3

Structured teaming reflections occurred three times across the term for metacognitive development purposes. As discussed in Chapter 2, reflection is widely recommended and employed in support of metacognitive development. The reflection activities in this study were intended to support the *reflection* step of the teaming skills employment model. Students performed metacognitive self-evaluation (part of *monitoring and control*) and *self-regulation* in order to update existing *metacognitive knowledge* and construct new *metacognitive knowledge*.

Teams were prompted by course instructors to consider the state of their teaming skills, their ability to monitor and control their teaming skills, and consider any potential issues or areas of improvement and what actions if any need to be taken to address them. The reflection prompts can be seen in Appendix Figure C.10, Appendix Figure C.11, Appendix Figure C.12, and Appendix Figure C.13. This exercise was allocated approximately three minutes in class, for a total of approximately ten in-class minutes across the semester. Some teams may have updated their Code of Cooperation or held further discussions outside of class based on these reflections, but such activities were not tracked.

It should be noted that in ENGR131, reflection was already incorporated into course assignments prior to the intervention. The topics of the reflections varied, tracking course topics and needs. The intervention reflections in ENGR131 refocused the existing exercises around teaming and psychological safety. More reflections, more types of reflections, and more training in reflective practice could all potentially increase the impact of the intervention, but were not feasible with the logistical constraints of the course.

3.4.5 Metacognitive Frequency Survey 1, 2, 3

Assessment of metacognition and metacognitive development remains an area under active research and development. Veenman reports that "Research on metacognitive instruction often merely reports product measures (i.e., the effects on learning outcomes)" (Veenman et al., 2006, p. 9). In engineering education, the product measurement approach can be seen in the work of Boiarsky (2004) and Newell et al. (2004). Various methods have been used to assess metacognition, and it may be that different methods are appropriate for different aspects of metacognition. Methods including surveys, think-aloud protocols, observations, stimulated recall, and computer use tracking have been employed in assessments of metacognition (Veenman et al., 2006, p. 8). In engineering education, Case et al. (2001) performed qualitative analyses of student journals and follow-up interviews to identify areas of metacognitive development in their course. Cunningham et al. (2015) report ongoing efforts to develop instruments to assess metacognition in engineering, but results are not yet available.

Given the sample size of this study, a scalable assessment method for metacognition was required to permit data collection and analysis. No instruments for metacognition specifically relating to teaming were found in a review of the literature. Therefore, a survey was developed in Rynearson and Hynes (2015) and revised for use in this study. The survey was found in a sample of more than 100 ENGR131 students to have a Cronbach's alpha of 0.78, which is adequate for experimental use. The version of the survey employed here omitted three of the original prompts found to decrease reliability of the instrument, leaving seven of the original ten prompts in use. The modified instrument had a Cronbach's alpha of 0.79 for the first administration in ENGR131, demonstrating that the reduction in the number of items did not adversely affect reliability.

This survey assessed the frequency of metacognitive *awareness* of opportunities to practice appropriate teaming skills with prompts specifically related to the CATME performance areas of Keeping the Team on Track (K) and Interacting with Teammates (I) during the activity that preceded the survey. Only two of the five CATME skill areas were employed in prompts for parsimony, acknowledging that this activity required class time in both experimental and control sections. This assessment may promote *reflection* in the control sections, which could have skewed measurements in the control sections upwards. This effect is not thought to be large given that instructors did not specifically prompt students to reflect or allocate class time for doing so. A representation of the survey in tabular format can be seen in Table 3.6. A representation of the survey as it appeared to students can be seen in Appendix Figure A.1.

Table 3.6: Metacognitive Frequency Survey

Item	Question Text
Question Prompt	During the team activity preceding this survey, how frequently did the following situations occur?*
Question 1	I monitored the progress my team made overall and tried to change things if the progress was not adequate
Question 2	I monitored the progress an individual team member was making on a given task
Question 3	I noticed that I gave a teammate timely, specific, and constructive feedback
Question 4	I noticed that I asked for and showed an interest in a teammate's ideas and contributions
Question 5	I noticed that I made sure that teammates stayed informed and understood each other
Question 6	I noticed that I provided encouragement or enthusiasm to the team
Question 7	I noticed that I asked for or respectfully received feedback from a teammate

* All questions are 5-point-scale multiple choice and possible responses are: never, rarely, sometimes, often, and all the time.

The survey was administered in class directly following in-class team working periods three times across the semester. This instrument was designed to be administered immediately after working period as ‘offline’ metacognitive data collection (Veenman et al., 2006, p. 8). Students were emailed Qualtrics survey links that were only usable during the intended class period, and timestamps for collected data were inspected to identify results not collected at the appropriate time. Instructors determined when during their class periods the data collection would be performed, given that it must occur directly after a ten minute team working period. Administration dates for this survey were selected to avoid overlapping administration of the SM worksheets, and can be seen in Table 3.2 and Table 3.3. Dates selected targeted team activities to directly precede the survey each time. In ENGR141, programming activity days were used. Programming activity days require student teams to face programming challenges collaboratively. In ENGR131, dates were selected for appropriate activities in conjunction with ENGR131 administrators. Administration dates featured times when teams could be expected to be working collaboratively on projects or other challenging, graded work. Note that when the survey was administered, the items were not separated into two screens; separating the items was necessary for static display of the survey in this document.

3.4.6 Data Cleaning for Research Question 1

As the only direct measure of metacognition administered to both the experimental and control groups, the metacognitive frequency survey instrument provides the most direct evidence for or against the hypothesis that the experimental interventions act to promote teaming metacognition. This section describes the data cleaning and analysis methods for the survey data, along with a discussion on the meaningfulness of results.

The survey was intended for use directly following team working periods, to place measurement directly after the activity of interest. Surveys completed at incorrect times are likely not comparable, and are inappropriate for inclusion in the study. A number of surveys were identified in both ENGR131 and ENGR141 as being completed either at the very beginning of the class in which the survey was to be taken or substantially after the completion of the majority of surveys in the class. The majority of surveys for each section were completed within an approximately 2-minute span, providing a clear benchmark for the nominal completion time for each section. Examining start and end times for these surveys clearly showed the moment when section faculty directed students to begin the survey. Surveys finished more than three minutes prior to that (finished less than seven minutes into the ten minute working period) were marked ‘early’ but the data was retained in the overall data set to permit analyses including or excluding this data.

Similarly, surveys completed more than ten minutes after the section faculty directed students to begin filling in the survey were marked ‘late’ and retained, again to permit analyses including or excluding this data. The margin for accepting late data was larger than for early data to create more space for students carefully pondering. Submissions more than ten minutes late were distant enough from the prompt that reasonable compliance with study procedures could not be assumed. The number of early or late submissions for each administration of the survey is shown in Table 3.7. Data submitted that does not fall into the ‘early’ or ‘late’ categories was labeled ‘timely’ and is described using that term in this study. Finally, opened but entirely blank surveys were discarded.

Table 3.7 Early and Late Submissions of Metacognitive Frequency Survey

Course	Survey 1 Early/Late	Survey 2 Early/Late	Survey 3 Early/Late
ENGR131	26	19	63
ENGR141	6	5	6

3.4.7 Analysis for Research Question 1

Response rates were calculated by section for both ‘Timely’ and all submitted surveys. Timely response counts and rates are shown in Table 4.1 and Table 4.2. For each student, the score for a given administration was calculated as the average response across the seven items. As each item is itself on a five-point scale, this resulted in a five-point scale for the overall instrument. However, scale scores do not fall exclusively on integer values on this scale due to the averaging of item scores. As this survey was intended to represent a single scale of metacognitive *awareness* in the teaming context of interest, scale scores (rather than item scores) are the analytic unit of interest.

Averages, medians, and standard deviations were calculated for each section, experimental condition, and administration of the survey to characterize the results for these conditions. Determining whether observed differences are meaningful in an educational sense is relatively difficult with an immature instrument such as the metacognitive frequency survey employed in this study. The sensitivity and performance of this instrument are not well understood. Some bounds can be established. For instance, a difference in the average scale score of a full point on the five-point scale across entire sections, corresponding to a broad difference in reported metacognitive frequency, would almost certainly be meaningful. A difference of only 0.01 on the same five-point scale, roughly corresponding to a one-point increase in one student's reported metacognitive frequency, cannot be interpreted as meaningful without further strong validation and assessment of the instrument itself. For the purposes of this analysis, an effect size of 0.5 on the 5-point scale was used as a benchmark for being educationally meaningful, which is both 10% of the scale and a typical standard deviation for scale scores observed for a section of students completing this instrument.

Analysis employed Kruskal-Wallace tests to determine whether differences between the experimental and control cohorts, differences between individual sections in the study, and differences between administrations were statistically significant. One-way ANOVA was not appropriate as the ordinal data collected is not normally distributed, and Kruskal-Wallace tests are the nonparametric analogue to ANOVA. In testing differences between only two sections or cohorts, Mann-Whitney U tests could have been used, but Kruskal-Wallace provides comparable results and was used for all comparisons. The Kruskal-Wallace test assesses the probability that the mean rank of two distributions is the same, rather than the mean, as would be the case with ANOVA (McDonald, 2014, p. 158). This has the effect of assessing whether compared samples come from the same distribution. It should be noted that “The null hypothesis of the Kruskal–Wallis test is often said to be that the medians of the groups are equal, but this is only true if you assume that the shape of the distribution in each group is the same. If the distributions are different, the Kruskal–Wallis test can reject the null hypothesis even though the medians are the same.” Therefore, though median values are reported in several tables in this study, they are not the determining factor in assessing differences between distributions. In this case, the distributions are made up of individual student scale scores. While the scale for each of the seven items is five points, student scale scores have far more than five values because they represent the average of the item scores for each student.

For the Kruskal-Wallace difference testing performed for this study, a significance level of $\alpha=0.05$ was selected. This was seen as an appropriate balance of experimental rigor and statistical sensitivity for testing an unproven set of educational interventions. Differences between distributions detected through Kruskal-Wallace testing at this level of significance were interpreted as statistically significant or distinct, but not necessarily meaningful.

3.5 Interventions, Assessments, and Analyses Primarily Associated with RQ2

Research Question 2: To what extent did the experimental intervention promote psychological safety of student teams in the experimental group in comparison to the control group?

Several interventions supporting psychological safety in student teams were implemented in the experimental sections across the semester. These include an explicit introduction to the concept, introduction of a classroom norm promoting psychological safety, alteration of the Code of Cooperation assignments to incorporate psychological safety, and in-class reflection prompts on psychological safety. Psychological safety was also assessed using a Psychological Safety Survey. Each of the interventions embodies recommendations for creating environments with high psychological safety identified in Chapter 2. All intervention activities were intended to support the *motivation* step of the teaming decision making process by reducing perceived risks of attempting the performance of teaming skills. Each intervention and the assessment are discussed in more detail, followed by analysis procedures and a discussion of what constitutes meaningful results in the context of this study.

3.5.1 Psychological Safety Introduction & Norm Setting

Instructors held a short classroom discussion early in the term focusing on psychological safety as an explicit course norm and goal. Sample talking points and slides were coordinated across experimental sections and can be seen in Appendix Figure C.9. In these introductions, psychological safety for all class members was explicitly discussed as a goal for the class, and the benefits of psychological safety to team performance were briefly reviewed. This introduction of material took approximately ten minutes. This activity was seen as analogous to a team or organizational leader establishing positive norms, as recommended by Edmondson (2002) and discussed in Section 2.7. Follow-up and continuous demonstration and reinforcement of course norms of psychological safety by instructors, GTA's, and peer instructors could increase the impact of this norm setting. However, these methods were seen as generally infeasible to implement without monitoring and authority over the individuals involved.

3.5.2 Revised Code of Cooperation Assignment

Both ENGR131 and ENGR141 typically ask newly formed student teams to create Codes of Cooperation detailing expected team processes and potential penalties for failure to adhere to expected team processes. In ENGR131, the Code of Cooperation assignment was already highly scaffolded, employing a fill-in-the-blanks structure. In ENGR141, this assignment is typically much more freeform. For ENGR141, the assigned task was altered to provide more explicit scaffolding for considering different teaming questions or issues and a requirement for the development of a team plan to promote psychological safety was added as a mandatory area of the Code of Cooperation. The prompt for this plan explicitly referenced recommended methods for promoting psychological safety identified in Section 2.7. For other areas, a list of questions that may be helpful to students in developing a strong Code of Cooperation were provided, but teams were not be required to address every possible question. The revised Code of Cooperation assignments for ENGR131 and ENGR141 can be seen in Appendix Figure B.5 and Appendix Figure B.6, respectively.

The altered Code of Cooperation assignments were expected to require student teams to spend additional time to complete them when compared with the unmodified assignments. The difference in time required was not measured. However, in ENGR131 the unmodified assignment was composed of three sections, requiring student responses in the areas of general contact information, guidelines for individual team members, and guidelines for the team as a whole. The altered assignment added a fourth section, on psychological safety. As this addition was in the same format and required the same number of responses as the previous two sections, it may be speculated that the altered assignment would take one-third to one-half longer than the unmodified Code of Cooperation assignment. It is plausible that the modified ENGR141 Code of Cooperation assignment required a similar additional quantity of time to complete.

3.5.3 Psychological Safety Reflection Prompts

In the three teaming reflections already occurring for metacognitive development purposes, teams were also prompted by course instructors to consider the state of psychological safety on their team and in the class, including potential issues or areas of improvement, and what actions if any needed to be taken to address them. This exercise took approximately 3 minutes each time, for a total of ten in-class minutes. Variation between sections in this time depended on individual instructors. Some teams may have updated their Code of Cooperation or held further discussions outside of class, but these activities were not tracked. The reflection prompts can be seen in Appendix Figure C.10, Appendix Figure C.11, Appendix Figure C.12, and Appendix Figure C.13. Reflective activities where problems and solutions can be considered were recommended to enhance psychological safety by Edmondson (2002) and discussed in Section 2.7.

3.5.4 Psychological Safety Survey 1, 2

As supporting student psychological safety in teams was one of the major goals of this study, it was imperative that it be appropriately assessed. Edmondson (1999) investigated several aspects of teaming performance and developed a 7-item instrument assessing team psychological safety that has been employed by other researchers (Carmeli & Gittell, 2009). This instrument's questions and wording do not require modification to be used in the undergraduate teaming environment, so the original questions were implemented in a survey via Qualtrics. The survey as it appeared to students can be seen in Appendix Figure A.2, though prompt order was randomized for a given user. Note that these figures would appear seamlessly as a single screen to students; separate images are used in this dissertation to permit static display. As with the previous survey, results were identifiable but students were reminded that their grades would not be affected by their responses to this survey. Supporting steps again consisted of a brief in-class introduction to the survey and reminders to complete it along with listings on course calendars and assignment lists. Psychological safety in a specific team was not seen as likely to be meaningful before teams had been formed and had at least some time to interact. For this reason, assessment of psychological safety began later in the courses than assessment of most other items. Taking this survey was not expected to contribute meaningfully to improved teaming performance.

3.5.5 Analysis for Research Question 2

As the direct measure of psychological safety administered to both the experimental and control groups, this instrument provides the most direct evidence for or against the hypothesis that the experimental interventions act to promote psychological safety. As this survey was administered outside of regular class time and no rationale for discarding responses prior to analysis was seen, response rates reflect only total submission rates.

Instrument scores for each student were calculated as the average response across the seven survey items, on a seven point scale, for each experimental condition, section, and administration. Condition, section, and administration averages, medians, and standard deviations were calculated based upon these student instrument scores.

The level at which differences or changes in psychological safety scale scores are educationally meaningful has not been precisely established in previous literature. Edmondson (1999) assessed the internal reliability of the psychological safety instrument and its validity (with interview and other data) and found both satisfactory for use, but has more recently stated that “consistent and accurate measures of the construct of psychological safety” remains a “methodological challenge. This study takes 5% of the scale range to be the smallest educationally meaningful measured difference (0.3 on a scale from 1-7). This corresponds to each student in a section improving one point on two of the seven items, or half the students improving four points across the seven items.

To establish whether the differences in outcomes between experimental and control conditions and between individual sections were statistically significant, Kruskal-Wallace tests were used. As previously discussed one-way ANOVA was not appropriate as the ordinal data collected is not normally distributed, and Kruskal-Wallace tests are the nonparametric analogue to ANOVA. It is again noted that the Kruskal-Wallace test assesses the probability that the mean rank of two or more distributions is the same, rather than the mean, as would be the case with ANOVA. For testing differences between condition, section, and time-point mean rank distributions, all data from those circumstances was used. For testing differences in growth (student scores on the second administration minus first administration scores), only data from students who completed both surveys was employed.

As previously discussed, a significance level of $\alpha=0.05$ was selected for the Kruskal-Wallace difference testing performed for this study. Differences in mean rank between measurements detected through Kruskal-Wallace testing at this level of significance were interpreted as statistically significant or distinct, but not necessarily meaningful.

3.6 Content, Assessments, and Analyses Primarily Associated with RQ3

Research Question 3: In the event that the intervention succeeds in supporting more metacognitive skills growth and/or psychological safety in the experimental than the control groups, to what extent did the teaming performance of the experimental group then improve beyond that of the control group?

No interventions directly targeting teaming skills development were implemented as part of this study. However, as previously addressed, existing course content on teaming was retained. ENGR131 and ENGR141 both present information on teaming early in the course, and ENGR131 also presents on the value of diversity in engineering teams. Teaming skills were assessed with the Comprehensive Assessment of Team Member Effectiveness Behaviorally Anchored Rating Scale (CATME BARS) three times across the semester. The teaming-related non-intervention course content and the CATME assessment are discussed in more detail, followed by analysis procedures and a discussion of what constitutes meaningful results in the context of this study.

3.6.1 Standard Intro to Teaming and Diversity (ENGR131)

In the first week of class, ENGR131 students reviewed online modules on teaming and diversity, then performed a number of brainstorming and ethics challenges and discussions in class in teams. The importance of working in diverse engineering teams to generate ideas and perspectives in the face of complex challenges or issues was emphasized. These challenges and discussions were scheduled to occupy about 80 minutes of class time. It is noted that a substantial percentage of this time was primarily devoted to questions of ethics, but with the integrated materials the specific amount focused on teaming versus ethics is difficult to distinguish and may have varied by instructor. It is plausible that students extracted some *metacognitive knowledge* about teaming from these modules and activities, which also most likely supported the *motivation* step of the teaming decision making process, by explicitly establishing the importance and necessity of teaming for engineers.

3.6.2 Standard Teaming Refresher (ENGR131)

In the middle of the course, at the close of a team-based mathematical modeling project and prior to the beginning of a team-based design project, ENGR131 briefly revisited earlier teaming content. The importance of engineering teaming to industry was reinforced and student teams were invited to brainstorm potential teaming issues and update their Codes of Cooperation. These activities were scheduled to occupy about 35 minutes of class time. It is again plausible that students extracted some *metacognitive knowledge* about teaming from these activities, which also most likely supported the *motivation* step of the teaming decision making process to some extent.

3.6.3 Standard Intro to Teaming (ENGR141)

During the first week of classes, ENGR141 introduced the importance of engineering teams, the stages of team development, roles for students to fill on teams, and the Code of Cooperation assignment. A mixture of assigned reading, lecture, and team/class discussion was employed. These materials occupied approximately 45 minutes of class time. As with the ENGR131 materials, it is plausible that the students extracted *metacognitive knowledge* about teaming from these activities, which also most likely supported the *motivation* step of the teaming decision making process.

3.6.4 CATME BARS

The ultimate goal of the intervention designed and deployed in this study was to have a positive impact on student teaming performance. Therefore, it was important to assess student teaming performance. Fortunately, both ENGR131 and ENGR141 had already integrated an appropriate teaming performance instrument into their regular operations. The Comprehensive Assessment of Team Member Effectiveness, or CATME, is an online tool designed to support behaviorally anchored peer and self-evaluation of teaming performance (Ohland et al., 2012). Peer evaluation results in general have been supported by evidence showing them to be generally reliable and valid (Malone, 2011; Topping, 1998) and the instrument's creators suggest that CATME can be used to assess "a program's effectiveness in developing students' team skills" (Loughry et al., 2014). CATME captures teaming performance in five areas: Contributing to the Team's Work (C), Interacting with Teammates (I), Keeping the Team on Track (K), Expecting Quality (E), and Having Relevant Knowledge, Skills, and Abilities (H). It is noted that while relationships may exist between the targeted development areas in metacognition and psychological safety and direct teaming skills performance on the CATME instrument, CATME was expected to assess constructs distinct from the development areas as it focuses on specific teaming behaviors, not the thought processes leading up to them. An example of potential overlap is that students in psychologically safer environments are more likely to interact in constructive ways with teammates and invest more effort in keeping the team on track to meet goals. The CATME instrument was administered to student three times throughout the semester as shown in Table 3.2 and Table 3.3. This schedule was selected to place teaming feedback and reflection at logical points in the term in relation to the course content and schedule.

Steps were taken to improve the quality of the assessment provided by the CATME instrument for both the experimental and control sections. First, efforts were be made to assure that student teams were formed employing similar practices. Both ENGR131 and ENGR141 use similar team-formation methods across all course sections. Second, all students were required to use CATME's rater-training calibration system prior to rating teaming performance. This system presents students with text describing teaming behavior and asks them to match the behavior described in the text to the appropriate rating on the scale. Feedback is provided to guide students to more consistently match behavior to ratings on the scale.

As completing the CATME peer and self-evaluations familiarizes students with aspects of strong and weak teaming performances and asks them to evaluate their own performance and that of others, use of the CATME instrument was in itself expected to contribute to growth in student teaming performance, at least for students who participated in the process fully. This is facilitated by the customized feedback delivered to students by the system advising them of areas of teaming performance that they may specifically wish to target improvement in, alongside ideas for how improvement might be pursued. It is not currently possible to track student review of CATME feedback directly, but students in both the experimental and control sections were encouraged to view this feedback as part of typical class practices.

3.6.5 Analysis for Research Question 3

As the main measurement of teaming skills performance administered to both the experimental and control groups, this instrument provides direct evidence for or against the hypothesis that the experimental interventions act to promote teaming skills performance. As this instrument was administered outside of regular class times and no rationale for discarding responses prior to analysis was seen, response rates reflect only total submission rates. Note that in this context, as up to four students provided feedback for each student, a response rate less than 100% does not necessarily mean that some students had no records, but rather that some student’s evaluations do not contain data from all of their teammates. It was found to be exceedingly unusual to have no data at all on an individual student. Only eight students had no data in the third administration of CATME in ENGR131. Therefore, virtually all students had data available from the first and final administrations of the instrument.

Averages and standard deviations of ratings for each item were calculated for each experimental condition, each section, and each administration. If more than one student rated a specific individual on a specific item (which is the case for the majority of the data), these rating were averaged to create a per-student rating, which was then part of the distribution of per-student ratings from which averages and standard deviations were calculated for the course or cohort. Differences in per-student ratings were also calculated between the first and final administration of the instrument to assess growth over time, creating a distribution of student growth values from which an average and standard deviation of growth could be calculated for each experimental condition, section, and time point.

A threshold for educational meaningfulness of results was selected as a difference or change of 0.10 on the 5-point scale, across a section or experimental condition. With a range of four points (1 to 5) on each of the five items, this threshold corresponds to half of the students in a given cohort going up one rating on one item while the other half of students does not make measureable progress.

To establish whether the differences in outcomes between experimental and control cohorts or between sections regardless of experimental condition were statistically significant, Kruskal-Wallace tests were used. As previously discussed, one-way ANOVA was not appropriate as the ordinal data collected is not normally distributed, and Kruskal-Wallace tests are the nonparametric analogue to ANOVA.

To test differences between the experimental and control cohorts, Kruskal-Wallace tests were performed on the mean rank for each item score distribution at the first and third scale administration, and the growth distribution for each item. Testing differences between all individual sections, Kruskal-Wallace tests were performed on the growth for each item between the first and third scale administration.

3.7 Analyses Associated with Research Question 4

Research Question 4: To what extent did the improvement in the targeted metacognitive capabilities or psychological safety correlate with improved individual student teaming performance?

This study's interventions were designed to promote improved teaming skills acquisition and performance through increasing teaming metacognition and psychological safety. The first three research questions investigate the efficacy of the intervention in promoting each of these items. For the final research question, the relationships between these variables are of interest. Therefore, there are no interventions or assessments specifically associated with this research question, only analyses. As improved teaming performance is the key outcome of the study, this analysis primarily examines the correlation between results of the metacognitive frequency survey and psychological safety instrument with CATME BARS growth outcomes. CATME data used was restricted to individuals for whom some evaluations were available on both the first and third scale administration, which removed a total of eight students in two

ENGR131 teams from consideration due to a total lack of data from the final administration.

It should be noted that the most common result in the CATME BARS growth observed in this study is zero in both the experimental and control groups, which could be interpreted to mean that most students do not exhibit growth, that the instrument less often detects limited growth, or some combination of the two. No clear method to estimate the relative role of these two components is available. It seems more likely that the standard course activities and experimental interventions foster at least limited growth in most students. Assuming this is part of the explanation for the common zero-growth phenomenon, the ability of the correlation tests to identify the relationship between growth in the variables and growth in CATME is reduced due to students with limited teaming skills growth falling below the detection threshold of the CATME scale. A more sensitive instrument might detect more limited growth, but would not already have been integrated into standard course operations, as CATME was.

As individual students could potentially test highly for any of the variables in either the experimental or control cohorts, the entire data set (including both experimental and control cohorts) was used to assess correlation unless otherwise noted. As the data to be correlated is in all cases ordinal, the Spearman's Rho test was used throughout.

For the metacognitive frequency survey, individual student ratings at the first and final administration were tested for Spearman's Rho correlation with CATME BARS growth for both individual CATME items and an overall average of the CATME items. Correlation coefficients and p-values were found for both ENGR131 and ENGR141. For both administrations of the psychological safety instrument, the correlation coefficients and p-values for Spearman's Rho correlation between the overall psychological safety scale average and CATME BARS per-item and overall average growth were found.

3.8 Case Study Methods

A set of case studies examining occurrences in nine specific teams across ENGR131 and ENGR14 was constructed. These case studies are minimal and informal, and are intended solely to provide specific examples for context and discussion. The case studies are not intended to directly address any research questions, though they center on psychological safety as results in that area were among the more interesting in the study. The case studies used extreme sampling, seeking teams with extreme characteristics for further investigation and review. This review included examination of data not otherwise incorporated into the study, such as motivation to learn teaming skills, team satisfaction, and final grades. The methods used to collect data in these areas will be discussed later in this section.

Teams were sought with extreme characteristics in four general areas. Two teams each from ENGR131 and ENGR141 were identified with high growth in psychological safety across the semester. Additionally, one team each from ENGR131 and ENGR141 were identified with consistently high psychological safety across the semester and with high initial psychological safety that decreased sharply over the term. One team in ENGR141 was identified with consistently low psychological safety. Therefore, a total of nine teams were selected for review in short case studies. Cases were sought where all or most team members had completed both instances of the psychological safety surveys to remove the focus from extreme cases driven by individual students' measurements.

After identifying teams of interest, demographic data, psychological safety, CATME performance and comments, teaming motivation, teaming satisfaction, and final grades were reviewed. When applicable, potential connections between the data and observed trends in the team were discussed. The collection of data relating to psychological safety and CATME are discussed elsewhere. Other data sources will be discussed here.

Demographic data was taken from Purdue University's student information system late in the semester. Teaming motivation was assessed with a nine question multiple choice survey featuring a mixture of original and adapted prompts from the 'Intrinsic Value' portion of the Motivated Strategies for Learning Questionnaire from Pintrich and De Groot (1990). The survey was administered three times, once at the very beginning of the semester and then twice more alongside the psychological safety survey. The question prompts of the survey are shown in Appendix Figure A.3.

A measure of teaming satisfaction is integrated into the CATME system, and asks students to respond to the statements "I am satisfied with my present teammates", "I am pleased with the way my teammates and I work together", and "I am very satisfied working with this team" on a scale from one to five, with five being very satisfied. This data was collected alongside each CATME administration in both ENGR131 and ENGR141. Finally, final grades were collected from course gradebooks more than two months after the end of the course. Therefore, it is not expected that many grades were changed after data was collected. Each case was reviewed individually, and a short final discussion of findings across multiple cases was also developed. These materials appear in Chapter 4.

CHAPTER 4. RESULTS

This chapter presents the results of the data collection and analyses discussed in Chapter 3. Results appear in order by research question, with the case studies appearing after all of the research questions. The text of each research question is given, followed by the results germane to that research question.

4.1 Research Question 1 Results

Research Question 1: To what extent did the experimental intervention promote the target aspects of metacognition in FYE students in the experimental group in comparison to the control group?

The computerized metacognitive frequency survey administered to all experimental and control sections in ENGR131 and ENGR141 was completed with overall high response rates. Response rate information can be seen in Table 4.1 and Table 4.2. Response rates dropped substantially in the third and final administration for the ENGR131 sections while ENGR141 rates experienced a smaller drop. As previously discussed, 'Timely' responses are those completed at least 75% of the way through the ten-minute pre-sampling period or those finalized less than ten minutes after the end of the sampling period. 'Complete sets' are the responses of participants who submitted complete responses for all three administrations of the instrument.

Table 4.1: ENGR131 Metacognition Frequency Survey Response Rates

Administration 1					
Section	**Students**	**Responses**	**Timely Resp.**	**Resp. Rate**	**Timely Resp. Rate**
EXP 1	120	115	107	95.8%	89.2%
EXP 2	119	112	107	94.1%	89.9%
EXP 3	119	115	115	96.6%	96.6%
CTRL 1	119	111	109	93.3%	91.6%
CTRL 2	119	95	84	79.8%	70.6%
Administration 2					
Section	**Students**	**Responses**	**Timely Resp.**	**Resp. Rate**	**Timely Resp. Rate**
EXP 1	120	113	105	94.2%	87.5%
EXP 2	119	116	114	97.5%	95.8%
EXP 3	119	115	113	96.6%	95.0%
CTRL 1	119	102	95	85.7%	79.8%
CTRL 2	119	113	112	95.0%	94.1%
Administration 3					
Section	**Students**	**Responses**	**Timely Resp.**	**Resp. Rate**	**Timely Resp. Rate**
EXP 1	120	81	71	67.5%	59.2%
EXP 2	119	104	95	87.4%	79.8%
EXP 3	119	101	91	84.9%	76.5%
CTRL 1	119	86	67	72.3%	56.3%
CTRL 2	119	92	78	77.3%	65.5%

Overall			
Section	**Students**	**Complete Sets Collected**	**Complete Set Resp. Rate**
EXP 1	120	62	51.67%
EXP 2	119	77	64.71%
EXP 3	119	86	72.27%
CTRL 1	119	71	59.66%
CTRL 2	119	57	47.90%

Table 4.2: ENGR141 Metacognitive Frequency Survey Response Rates

Administration 1					
Section	**Students**	**Responses**	**Timely Resp.**	**Resp. Rate**	**Timely Resp. Rate**
EXP 1	69	69	68	100.0%	98.6%
EXP 2	65	62	60	95.4%	92.3%
CTRL 1	61	60	60	98.4%	98.4%
CTRL 2	68	66	63	97.1%	92.6%
Administration 2					
Section	**Students**	**Responses**	**Timely Resp.**	**Resp. Rate**	**Timely Resp. Rate**
EXP 1	120	67	67	97.1%	97.1%
EXP 2	119	63	63	96.9%	96.9%
CTRL 1	119	60	59	98.4%	96.7%
CTRL 2	119	63	59	92.6%	86.8%
Administration 3					
Section	**Students**	**Responses**	**Timely Resp.**	**Resp. Rate**	**Timely Resp. Rate**
EXP 1	120	66	65	95.7%	94.2%
EXP 2	119	59	58	90.8%	89.2%
CTRL 1	119	56	56	91.8%	91.8%
CTRL 2	119	65	61	95.6%	89.7%
Overall					
Section	**Students**	**Complete Sets Collected**		**Complete Set Resp. Rate**	
EXP 1	69	62		89.86%	
EXP 2	65	52		80.00%	
CTRL 1	61	54		88.52%	
CTRL 2	68	50		73.53%	

Aggregated results by section and experimental condition can be seen in Table 4.3 and Table 4.4. In these tables the ‘Average’ is the average student survey score on the scale of 1-5, with 5 corresponding to more frequent metacognition. Recall that the score for each student is the average of their responses across the seven survey items. The median and standard deviation results also address survey scores. Only ‘Timely’ results were used.

Table 4.3: ENGR131 Metacognitive Frequency Survey Results

Section	Administration	N	Average	Median	Stdev.
EXP 1	1	107	3.58	3.57	0.56
	2	105	3.65	3.57	0.53
	3	71	3.65	3.71	0.52
EXP 2	1	107	3.48	3.43	0.60
	2	114	3.78	3.79	0.50
	3	95	3.74	3.71	0.58
EXP 3	1	115	3.64	3.57	0.55
	2	113	3.80	3.71	0.48
	3	91	3.85	3.71	0.53
CTRL 1	1	109	3.69	3.71	0.46
	2	95	3.88	3.79	0.62
	3	67	3.79	3.86	0.46
CTRL 2	1	84	3.63	3.57	0.57
	2	112	3.71	3.71	0.65
	3	78	3.60	3.57	0.65
ALL EXP	1	329	3.57	3.57	0.58
ALL EXP	3	257	3.75	3.71	0.55
ALL CTRL	1	193	3.66	3.71	0.51
ALL CTRL	3	145	3.65	3.71	0.64
ALL EXP	3 Minus 1 (Growth)	235	0.15	0.14	0.54
ALL CTRL	3 Minus 1 (Growth)	121	0.04	0.00	0.55

Note: Growth results were calculated using data from students for which 'Timely' data was available for both the first and third administration of the instrument. All other results were calculated using all 'Timely' data available for each administration.

Table 4.4: ENGR141 Metacognitive Frequency Survey Results

Section	Administration	N	Average	Median	Stdev.
EXP 1	1	68	3.40	3.42	0.54
	2	67	3.45	3.43	0.45
	3	65	3.48	3.57	0.62
EXP 2	1	60	3.57	3.57	0.44
	2	63	3.56	3.57	0.50
	3	58	3.59	3.57	0.57
CTRL 1	1	60	3.40	3.43	0.62
	2	59	3.55	3.57	0.48
	3	56	3.74	3.71	0.56
CTRL 2	1	63	3.63	3.57	0.61
	2	59	3.53	3.43	0.54
	3	61	3.68	3.71	0.64
ALL EXP	1	128	3.48	3.57	0.50
ALL EXP	3	123	3.53	3.57	0.60
ALL CTRL	1	123	3.52	3.43	0.63
ALL CTRL	3	117	3.71	3.71	0.61
ALL EXP	3 Minus 1 (Growth)	121	0.16	0.00	0.82
ALL CTRL	3 Minus 1 (Growth)	115	0.28	0.143	0.81

Note: Growth results were calculated using data from students for which 'Timely' data was available for both the first and third administration of the instrument. All other results were calculated using all 'Timely' data available for each administration.

4.1.1 Results Overview

Overall, mean instrument ratings across all sections and time points fall in the range of 3 to 4 out of the five options (1- never, rarely, sometimes, often, 5 - all the time) for reported frequency of the listed metacognitive activities. This suggests that the target behavior of noticing teaming-relevant actions is sometimes or often already performed by many students but room often remains for improvement. Overall mean scores rise in eight of the nine sections, and medians rise in seven of the nine sections. Some increases in metacognition were expected. Veenman (2006, p. 8) states that "metacognitive knowledge and skills… become more sophisticated and academically oriented whenever formal educational requires the explicit utilization of a metacognitive repertoire."

However, the observed differences are marginally meaningful in educational sense, at best. The largest observed increase in median for a single section is approximately 0.3 on a five-point scale, below the 0.5 metric established in Section 3.4.7. It is also possible that this observed difference is at least partly attributable to instrument or sampling error, or other differences between sections such as instructor. A higher proportion of students reported negative growth between the first and third survey administration than expected, with approximately 1/3 of students (see Table 4.5) who completed both the first and third administrations of the survey reporting lower metacognition on the third survey than the first. If metacognition is assumed to trend upward over time, reported decreases in metacognition may be due to instrument error or team activities less conducive to metacognition at the third administration of the survey, despite efforts to sample during similar team activities.

Table 4.5: Metacognitive Frequency Survey Growth (First to Third Administration)

Course	N	Decreases	Increases	No Change
ENGR141	236	81	125	30
ENGR131	356	115	194	47

4.1.2 Kruskal-Wallace Testing

Kruskal-Wallace testing was performed to assess whether a statistical difference exists between the distributions of measured values for administrations, experimental conditions, and between individual sections. Kruskal-Wallace testing uses the mean rank of two or more distributions, rather than the mean, to assess statistical significance. At the time of the first administration of the survey, differences in the mean rank of student score between the experimental and control conditions was not significant at $\alpha=0.05$ for both ENGR131 ($p=0.081$) and ENGR141 ($p=0.857$). Similarly, the mean rank of all individual sections were also not statistically significantly different at $\alpha=0.05$ at the time of the first survey administration for both ENGR131 ($p=0.172$) and ENGR141 ($p=0.074$). These results do not provide strong evidence that student performance was different between the experimental and control conditions or between sections, on the first survey instrument.

Testing for difference between the experimental and control condition at the third administration of the survey, statistically significant results were not found at $\alpha=0.05$ for ENGR131 ($p=0.198$). However, differences between individual ENGR131 sections were found to be statistically significant at $\alpha=0.05$, with $p=0.032$. Given that survey results at the first administration of the survey were comparable, as were third-administration results for the experimental and control conditions, the fact that individual sections were not comparable at the third administration may suggest that sampling procedures or one or more outside factors had a larger influence than the experimental intervention on the survey results in ENGR131. Instructor or cohort effects are some potential candidates. Boxplots showing the distribution of survey results at the first and third administration can be seen in Figure 4.1 and Figure 4.2, respectively. While the experimental and the first control section rise in median and fall or hold steady in standard deviation, the second control section remains constant in median and increases in standard deviation.

No notable demographic or student grading differences between the second control section and the other sections are apparent in Table 3.1, which points towards instructor or sampling procedure rather than student cohort effects.

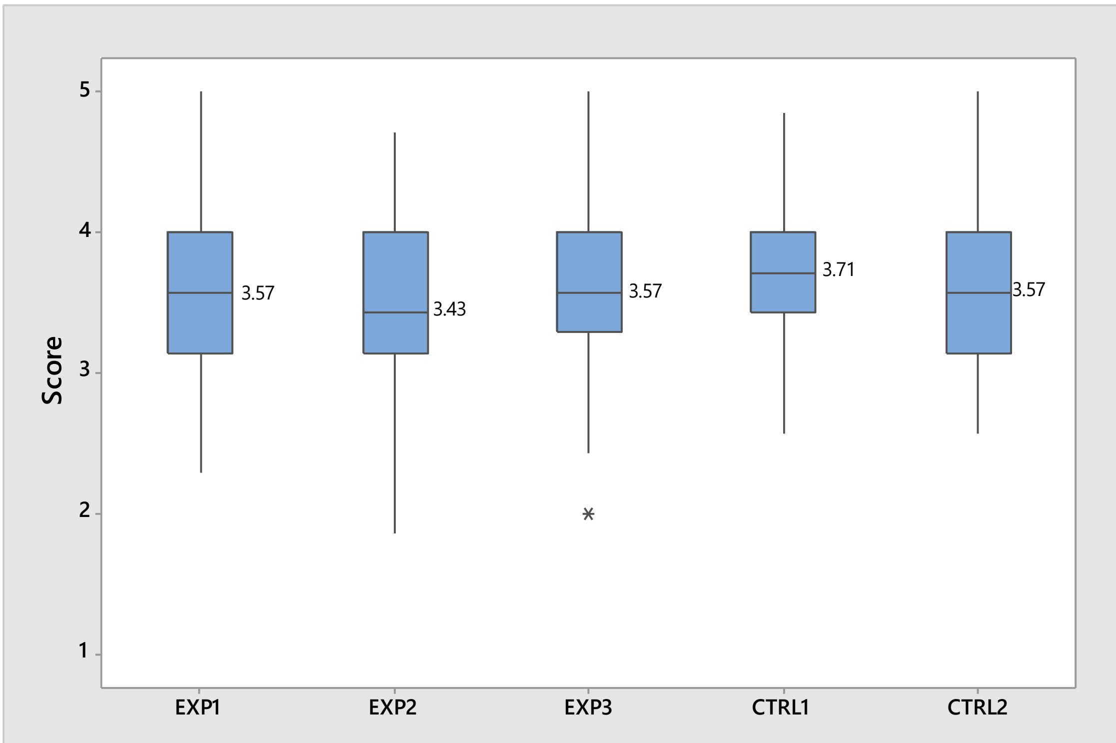


Figure 4.1: ENGR131 Timely Scores, Administration 1, By Section

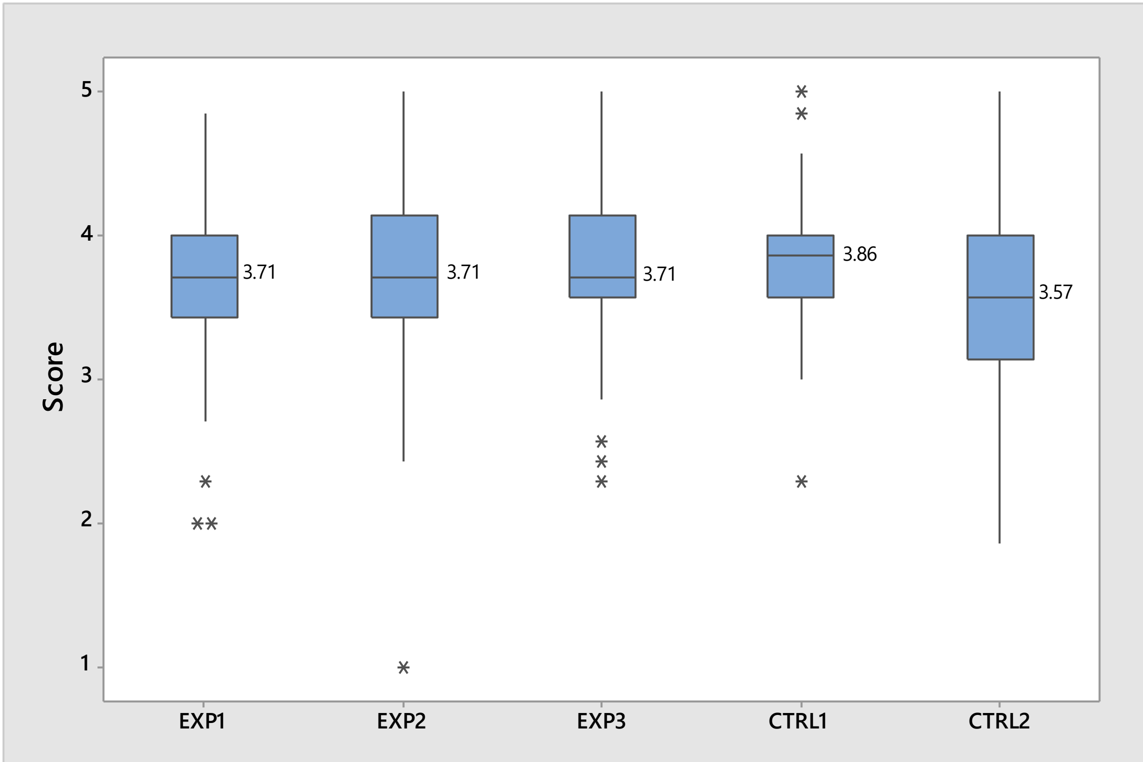


Figure 4.2: ENGR131 Timely Scores, Administration 3, By Section

One potential alternative explanation is that the differences between ENGR131 sections observed in the third administration of the survey could be a product of the sample itself, due to falling response rates over the semester. Kruskal-Wallace testing was performed for ENGR131 per-section differences using only results from students who completed both first and third administration of the survey. This testing also found a statistically significant difference between ENGR131 sections at the third administration of the survey (p=0.024). This finding supports the contention that sampling procedures or a factor such as instructor or student cohort is predominantly responsible for the measured differences between ENGR131 sections in the third survey administration.

Testing for difference between the experimental and control condition at the third administration of the survey, statistically significant results were found at $\alpha=0.05$ in ENGR141 ($p=0.035$). A boxplot of the distributions of scores from the experimental and control conditions can be seen in Table 4.3. In this case, the control section results, which were not statistically distinguishable from the experimental sections at the time of the first survey administration, were found to be higher than those of the experimental sections. This is contrary to the expectations of the study, and it does not seem plausible that the intervention actively retarded acquisition of metacognitive skills, as this result would suggest. This may be further evidence of factors outside of experimental procedures or control influencing the results or weaknesses of the instrument. Kruskal-Wallace testing for difference in mean rank between EXP1 and EXP2 at the third administration of the instrument, which would point to instructor or cohort effects, were not statistically significant ($p=0.294$). Kruskal-Wallace testing for difference in mean rank between CTRL1 and CTRL2, which shared an instructor, were also not statistically significant ($p=0.65$). Further discussion and implications of these results appear in Chapter 5.

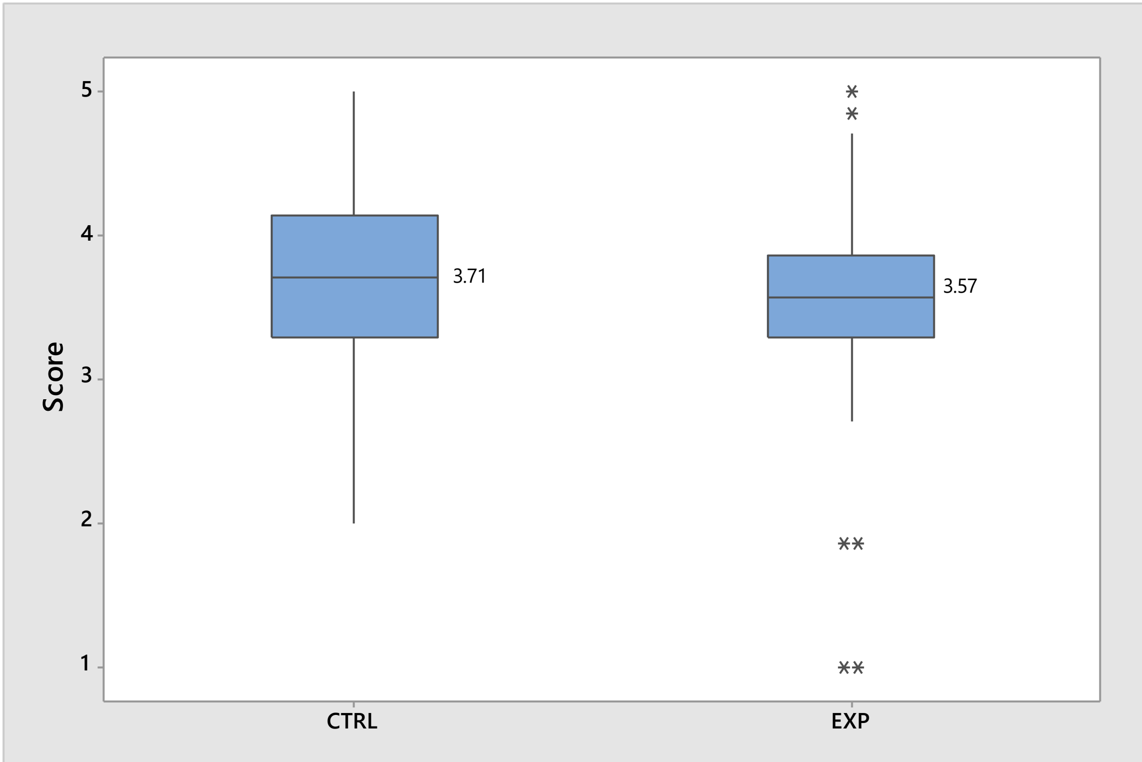


Figure 4.3: ENGR141 Timely Scores, Administration 3, By Experimental Condition

Additional Kruskal-Wallace testing was performed to assess differences in growth on instrument scores between experimental conditions and individual sections. Given comparable survey results on the initial administration, but statistically significant differences in the third administration for both ENGR131 and ENGR141, observing statistically significantly different results for per-section (ENGR131) and per-condition (ENGR141) growth might be expected. However, Kruskal-Wallace tests performed on experimental and control condition growth (ENGR141, $p=0.100$) and per-section growth (ENGR131, $p=0.074$) did not detect statistically significant differences in growth at $\alpha=0.05$. Growth was calculated only for students who submitted both the first and third administrations of the survey, subtracting their score on the first administration from their score on the third administration. This means that the data set used for calculating growth is a subset of the overall available data. It is plausible that differences in initial survey results and in growth below the level of statistical detection combined to statistically differentiate the results of the third survey administration.

4.1.3 Summary

In summary, no results were found to support the efficacy of the intervention. Some evidence was found to suggest that experimental procedures or outside factors, potentially student cohort or instructor, may have a greater influence over the targeted metacognitive development than the intervention. The fact that overall metacognitive development was found to increase for both ENGR131 and ENGR141 is positive and aligns with expectations from literature. The presence of reported negative growth in the results exposes potential weaknesses in the instrument and experimental protocol.

4.2 Research Question 2 Results

Research Question 2: To what extent did the experimental intervention promote psychological safety of student teams in the experimental group in comparison to the control group?

The two computerized surveys of psychological safety administered to all experimental and control sections in ENGR131 and ENGR141 were completed with overall section response rates between 55% and 85%, and with complete set response rates of between 40% and 60%. Response rate information can be seen in Table 4.6 (ENGR131) and 4.7 (ENGR141). ‘Complete sets’ are the responses of participants who submitted complete responses for both survey administrations of the instrument.

Aggregated data by section and comparisons of experimental and control conditions' initial and final measurements can be seen in Table 4.8 and Table 4.9. It is noted that reverse-coded items on the instrument were re-oriented such that higher numbers always reflect higher psychological safety in these tables.

Table 4.6: ENGR131 Psychological Safety Instrument Response Rates

		Administration 1	
Section	**Students**	**Responses**	**Response Rate**
EXP 1	120	95	79.2%
EXP 2	119	84	70.6%
EXP 3	119	99	83.2%
CTRL 1	119	91	76.5%
CTRL 2	119	67	56.3%
		Administration 2	
Section	**Students**	**Responses**	**Response Rate**
EXP 1	120	87	72.5%
EXP 2	119	68	57.1%
EXP 3	119	71	59.7%
CTRL 1	119	67	56.3%
CTRL 2	119	74	62.2%
		Overall	
Section	**Students**	**Complete Sets**	**Complete Set Response Rate**
EXP 1	120	71	59.2%
EXP 2	119	55	46.2%
EXP 3	119	63	52.9%
CTRL 1	119	57	47.9%
CTRL 2	119	50	42.0%

Table 4.7: ENGR141 Psychological Safety Instrument Response Rates

Administration 1			
Section	**Students**	**Responses**	**Resp. Rate**
EXP 1	69	68	98.6%
EXP 2	65	54	83.1%
CTRL 1	61	58	85.3%
CTRL 2	68	52	85.2%
Administration 2			
Section	**Students**	**Responses**	**Resp. Rate**
EXP 1	69	58	84.1%
EXP 2	65	43	66.2%
CTRL 1	61	44	64.7%
CTRL 2	68	41	67.2%
Overall			
Section	**Students**	**Complete Sets Collected**	**Complete Set Resp. Rate**
EXP 1	69	57	82.61%
EXP 2	65	38	58.46%
CTRL 1	61	41	60.29%
CTRL 2	68	39	63.93%

Table 4.8: ENGR131 Psychological Safety Item Averages and Growth

Section	Administration	N	Average	Median	Stdev.
EXP 1	1	95	5.70	5.71	1.00
	2	87	5.86	5.86	0.97
EXP 2	1	84	5.87	6.00	0.96
	2	68	6.02	6.00	0.84
EXP 3	1	99	5.91	5.86	0.91
	2	71	5.72	5.86	0.97
CTRL 1	1	91	5.58	5.71	1.05
	2	67	5.92	6.14	0.96
CTRL 2	1	67	5.36	5.43	1.02
	2	74	5.57	5.71	0.93
ALL EXP	1	278	5.83	5.86	0.96
ALL EXP	2	226	5.87	5.86	0.94
ALL CTRL	1	158	5.48	5.57	1.05
ALL CTRL	2	141	5.73	5.86	0.96
ALL EXP	2 Minus 1 (Growth)	188	-0.01	0.07	0.94
ALL CTRL	2 Minus 1 (Growth)	107	0.25	0.29	0.99

Table 4.9: ENGR141 Psychological Safety Item Averages and Growth

Section	**Administration**	**N**	**Average**	**Median**	**Stdev.**
EXP 1	1	68	5.78	5.78	0.91
	2	58	5.71	5.93	1.12
EXP 2	1	54	5.97	6.00	0.83
	2	43	5.55	5.57	1.04
CTRL 1	1	58	5.77	5.71	0.82
	2	44	5.81	5.86	0.92
CTRL 2	1	52	5.69	5.71	0.92
	2	41	5.72	6.00	1.03
ALL EXP	1	122	5.87	5.86	0.88
ALL EXP	2	101	5.64	5.86	1.09
ALL CTRL	1	110	5.73	5.71	0.88
ALL CTRL	2	85	5.76	5.86	0.97
ALL EXP	2 Minus 1 (Growth)	95	-0.16	-0.14	0.90
ALL CTRL	2 Minus 1 (Growth)	80	-0.06	0.00	0.92

4.2.1 Results Overview

Overall, mean instrument ratings across all sections and time points fall in the range between five and six, out of the seven options (1- very inaccurate to 7- very accurate) for reported psychological safety. This indicates that team psychological safety as described by the scale is consistently fairly strong in the aggregate. While it is excellent news that students predominantly report environments that are not actively psychologically unsafe, there is room for psychological safety to be improved. Standard deviations concentrated around 1 point on the 7-point scale indicate substantial variability in individual experiences of psychological safety in both ENGR131 and ENGR141.

Examination of the results tables for ENGR131 (Table 4.8) and ENGR141 (Table 4.9) shows that experimental sections in ENGR131 and ENGR141 both reported negative growth in average psychological safety, while the control sections reported either less negative growth (ENGR141) or positive growth (ENGR131). The growth across both ENGR131 control sections approaches the mark set for educationally meaningful differences in measurement, at 0.25, against the previously set benchmark of 0.30. Only two section measurements exceed this value, those being the drop in measured psychological safety across the term in the EXP2 section of ENGR141 at 0.42, and the average growth in ENGR131's CTRL1 section, at 0.34. These results appear to show the control sections substantially outperforming the experimental sections in promoting psychological safety in student teams.

However, examination of Table 4.10 and Table 4.11 is helpful in understanding these results. In these tables it can be seen that in both ENGR131 and ENGR141, the experimental cohorts' ratings of psychological safety for the first administration of the scale were higher than the control cohort's ratings. This may be an effect of the intervention's discussion of psychological safety and altered Code of Cooperation assignment for experimental sections raising psychological safety early in the term, and having that safety eroded by actual team working events over time. This would act to reduce apparent growth in the experimental cohorts. It may also represent something of a psychological safety 'ceiling' where conditions in ENGR131 and ENGR141 are conducive to growth in psychological safety up a point, beyond which growth for at least some teams is unlikely. If the large drop in psychological safety in ENGR141's EXP2 (six times larger than the drop in EXP1 and a third larger than any other change in measured average in the study) was removed from consideration, this rationale would be further strengthened. However, it is difficult to provide concrete evidence in support of this rationale with the existing data. Further sampling (potentially trying to observe the 'forming, storming, norming, performing' sequence) and gathering data in additional forms (such as interviews) may be helpful in understanding this dynamic in future work. Students in the experimental cohort ended the semester with a higher average psychological safety than the control in ENGR131, but a lower average rating in ENGR141. The statistical significance of the measured differences will be assessed in the next section.

Table 4.10: ENGR131 Psychological Safety Initial and Final Measurements Comparison

Measurement	**Average**
EXP Administration 1	5.83
CTRL Administration 1	5.48
EXP – CTRL Admin. 1	0.50
EXP Administration 2	5.87
CTRL Administration 2	5.73
EXP – CTRL Admin. 2	0.14

Table 4.11: ENGR141 Psychological Safety Initial and Final Measurements Comparison

Item	**Average**
EXP Administration 1	5.87
CTRL Administration 1	5.73
EXP – CTRL Admin. 1	0.14
EXP Administration 2	5.64
CTRL Administration 2	5.76
EXP – CTRL Admin. 2	-0.12

4.2.2 Kruskal-Wallace Testing

Kruskal-Wallace testing was performed to assess whether a statistical difference exists between the measured values for administrations, experimental conditions, and between individual sections. Kruskal-Wallace testing uses the mean rank of two or more distributions, rather than the mean, to assess statistical significance. First, it is noted that the much lower starting ratings for ENGR131's control sections are significant at $\alpha=0.05$ with a p-value of 0.001. There exists strong statistical evidence that the control sections' average psychological safety was lower than that of the experimental sections on the first administration of the instrument in ENGR131. This could be due to the experimental intervention, which took place before the first data point was collected as previously discussed. However, by the second administration of the instrument later in the semester, there was no longer a statistically significant difference between the experimental and control sections ($p=0.193$). This fits with what was seen the review of average scores in the previous section, with the control cohort largely making up the gap between the low control starting ratings and the high-but-stagnant experimental ratings.

Growth for the control section was also statistically significantly higher than that of the experimental section at $p=0.008$. The mechanism for this growth in psychological safety is likely teams improving their norms and interactions through working together over the term. It is plausible that teams could arrive at higher psychological safety either through an explicitly supported process (as in the experimental sections) and by trial and error across the term (as in the control sections). However, per-section scale averages were found to be statistically significantly different at both the first and second administrations of the instrument for ENGR131 ($p=0.003$ and $p=0.034$ respectively). This indicates that some between-section differences are more pronounced than those between the experimental and control cohorts for the second administration of the instrument. It is likely that the results are again influenced by instructor, GTA, student cohort, or other confounded effects.

In a striking difference from ENGR131, statistical measures of difference for ENGR141 found no statistically significant results. Based on the samples tested, strong statistical evidence is not present to suggest that the experimental intervention had any effects on average ratings for either administration or on growth. Between-section differences also do not appear to be present in average ratings or growth. With a considerably smaller sample and comparable standard deviations versus ENGR131, it may be that differences (including potentially educationally meaningful differences) were not detected by the testing. It is also possible that the ENGR141 coursework eliminates the impact of the additional psychological safety training, or that the ENGR141 student population differs from the ENGR131 population in some way that influences psychological safety. A lower proportion of international students, higher self-efficacy, or lower social anxiety might differentiate the ENGR141 (Engineering Honors) population.

4.3 Research Question 3 Results

Research Question 3: In the event that the intervention succeeds in supporting more metacognitive skills growth and/or psychological safety in the experimental than the control groups, to what extent did the teaming performance of the experimental group then improve beyond that of the control group?

Teaming performance was assessed through the CATME BARS instrument. The CATME BARS was administered three times in both ENGR131 and ENGR141. Each student self-rated and was rated by either two or three teammates, depending on the size of the team. Therefore, if one student failed to complete the scale, some data was absent for up to four students in the overall instrument. In ENGR141, all administrations of the instrument had at least one set of ratings for each student. In ENGR131, all but eight students in the third and final administration had at least one set of ratings. Overall response rates for the CATME BARS exceed 75% in all cases. Table 4.12 documents the response rates for ENGR131 and ENGR141.

Table 4.12: CATME BARS Response Rates

ENGR131			**ENGR141**		
Section	**Administration**	**Resp. Rate**	**Section**	**Item**	**Resp. Rate**
EXP 1	Week 7	84.5%	EXP 1	Week 5	100.0%
	Week 11	88.3%		Week 13	95.0%
	Week 16	94.2%		Week 16	88.0%
EXP 2	Week 7	80.0%	EXP 2	Week 5	87.0%
	Week 11	85.0%		Week 13	84.0%
	Week 16	82.5%		Week 16	77.0%
EXP 3	Week 7	88.3%			
	Week 11	92.5%			
	Week 16	88.3%			
CTRL 1	Week 7	87.0%	CTRL 1	Week 5	79.0%
	Week 11	88.4%		Week 13	90.0%
	Week 16	83.6%		Week 16	81.0%
CTRL 2	Week 7	83.3%	CTRL 2	Week 5	83.0%
	Week 11	91.7%		Week 13	84.0%
	Week 16	90.8%		Week 16	82.0%

CATME rates teaming performance in five areas: Contributing to the Team's Work (C), Interacting with Teammates (I), Keeping the Team on Track (K), Expecting Quality (E), and Having Relevant Knowledge, Skills, and Abilities (H). Results aggregated by experimental condition and section can be seen in Table 4.13 (ENGR131) and Table 4.14 (ENGR141). Multiple raters typically rate each student for each performance area at each time point. The ratings of all raters are averaged for each student, giving a student-level average rating for each performance area. Then, based on the collection of student-level average ratings for each performance area, section-level or experimental-condition-level averages can be calculated. It is the section-level and condition-level averages shown in the table, along with section-level and condition-level standard deviations. For the growth values (shown as administration 3-1 in the tables), growth was calculated for each student (assuming data was available for the first and third administration) to create a distribution of student growth values. The overall averages and standard deviations for growth were then taken from this distribution. Median values rarely varied across the data, and are not shown in the tables. Likewise, the number of participants contributing to each rating is not shown due to the extremely high response rates previously discussed.

Table 4.13: ENGR131 CATME BARS Aggregate Data

Section	**Admin.**	**C**		**I**		**K**		**E**		**H**		**Overall**	
		Avg.	**Stdev.**	**Avg.**	**Stdev.**	**Avg.**	**Stdev.**	**Avg.**	**Stdev.**	**Avg.**	**Stdev.**	**Avg.**	**Stdev.**
EXP 1	1	3.88	0.95	3.98	0.87	3.90	0.93	3.99	0.91	4.14	0.86	3.98	0.91
	2	3.54	0.86	3.64	0.85	3.54	0.88	3.57	0.83	3.80	0.83	3.62	0.85
	3	3.78	0.82	3.87	0.84	3.77	0.85	3.86	0.83	3.89	0.77	3.83	0.82
EXP 2	1	3.83	0.95	3.97	0.84	3.80	0.93	3.88	0.83	4.15	0.84	3.93	0.89
	2	3.81	0.86	3.94	0.86	3.81	0.84	3.89	0.85	4.03	0.82	3.90	0.85
	3	4.03	0.87	4.04	0.84	3.92	0.88	4.03	0.85	4.19	0.80	4.04	0.85
EXP 3	1	3.84	0.96	3.88	0.86	3.88	0.91	3.89	0.80	4.14	0.82	3.93	0.88
	2	3.94	0.74	4.06	0.72	3.94	0.78	4.05	0.69	4.07	0.72	4.01	0.73
	3	4.04	0.81	4.10	0.77	4.02	0.80	3.97	0.80	4.10	0.75	4.05	0.79
CTRL 1	1	3.86	0.91	4.11	0.84	3.84	0.96	4.01	0.83	4.16	0.84	4.00	0.89
	2	3.98	0.82	4.02	0.78	3.99	0.81	4.00	0.72	4.15	0.78	4.03	0.78
	3	4.05	0.84	4.09	0.83	4.01	0.87	4.07	0.79	4.13	0.76	4.07	0.82
CTRL 2	1	3.71	0.99	3.80	0.95	3.67	1.01	3.80	0.90	4.04	0.90	3.80	0.96
	2	3.91	0.89	3.93	0.90	3.88	0.88	3.88	0.88	4.09	0.83	3.94	0.88
	3	3.91	0.93	3.96	0.89	3.89	0.92	3.97	0.87	4.06	0.84	3.96	0.89
All EXP	1	3.85	0.95	3.94	0.86	3.89	0.93	3.92	0.85	4.15	0.84	3.94	0.89
ALL CTRL	1	3.79	0.95	3.96	0.91	3.75	0.99	3.91	0.87	4.10	0.88	3.90	0.93
All EXP	2	3.77	0.83	3.88	0.83	3.76	0.85	3.84	0.82	3.96	0.80	3.84	0.83
ALL CTRL	2	3.95	0.86	3.97	0.84	3.93	0.85	3.94	0.81	4.12	0.80	3.98	0.83
All EXP	3	3.95	0.84	4.00	0.82	3.90	0.85	3.95	0.83	4.06	0.78	3.97	0.83
ALL CTRL	3	3.98	0.89	4.02	0.87	3.95	0.90	4.02	0.83	4.10	0.80	4.01	0.86
All EXP	3-1	0.06	0.60	0.02	0.62	0.01	0.57	0.00	0.62	-0.114	0.58	-0.01	0.45
ALL CTRL	3-1	0.18	0.66	0.09	0.66	0.21	0.73	0.14	0.72	0.00	0.64	0.12	0.55

Table 4.14: ENGR141 CATME BARS Aggregate Data

Section	Admin.	C		I		K		E		H		Overall	
		Avg.	**Stdev.**	**Avg.**	**Stdev.**	**Avg.**	**Stdev.**	**Avg.**	**Stdev.**	**Avg.**	**Stdev.**	**Avg.**	**Stdev.**
EXP 1	1	3.74	0.87	3.72	0.89	3.59	0.87	3.96	0.84	3.84	0.87	3.77	0.88
	2	3.80	0.90	3.87	0.79	3.74	0.87	3.83	0.79	3.94	0.87	3.82	0.85
	3	3.96	0.86	3.87	0.88	3.77	0.83	3.88	0.86	4.08	0.86	3.91	0.86
EXP 2	1	4.01	0.78	3.99	0.85	3.91	0.85	4.03	0.80	4.13	0.80	4.01	0.82
	2	3.83	0.95	3.89	0.98	3.80	0.96	4.03	0.90	4.10	0.86	3.93	0.94
	3	3.88	1.04	3.77	0.92	3.77	0.91	3.90	0.87	4.03	0.91	3.87	0.93
CTRL 1	1	3.85	0.84	4.07	0.78	3.87	0.97	4.08	0.80	3.90	0.75	3.95	0.84
	2	3.90	0.97	3.93	0.92	3.77	0.93	3.87	0.88	3.99	0.88	3.88	0.92
	3	3.97	1.02	3.95	0.87	3.90	0.95	3.97	0.93	3.99	0.84	3.95	0.92
CTRL 2	1	3.87	0.82	3.97	0.97	4.03	0.76	4.13	0.73	4.10	0.80	4.01	0.82
	2	3.92	0.97	3.80	0.96	3.78	0.98	3.97	0.87	4.12	0.88	3.92	0.94
	3	3.94	0.98	3.89	0.99	3.90	0.88	3.97	0.95	4.15	0.92	3.97	0.95
All EXP	1	3.86	0.84	3.84	0.88	3.73	0.88	3.99	0.82	3.97	0.85	3.88	0.86
ALL CTRL	1	3.86	0.83	4.00	0.88	3.95	0.87	4.11	0.76	4.00	0.78	3.98	0.83
All EXP	2	3.82	0.92	3.88	0.88	3.77	0.91	3.91	0.85	4.01	0.87	3.88	0.89
ALL CTRL	2	3.91	0.96	3.86	0.94	3.77	0.95	3.92	0.88	4.05	0.88	3.90	0.93
All EXP	3	3.92	0.94	3.82	0.90	3.77	0.86	3.89	0.86	4.05	0.88	3.89	0.89
ALL CTRL	3	3.95	1.00	3.92	0.93	3.89	0.919	3.96	0.94	4.07	0.88	3.96	0.94
All EXP	3-1	0.04	0.69	-0.03	0.61	0.03	0.65	-0.07	0.65	0.06	0.57	0.01	0.43
ALL CTRL	3-1	0.03	0.73	-0.10	0.74	-0.10	0.68	-0.17	0.71	0.05	0.70	-0.06	0.58

4.3.1 Results Overview

It can be seen that average item ratings typically fall between 3.50 and 4.25, with standard deviations typically between 0.80 and 1.00. It is noted that the assessment items are on a scale from one to five, where five corresponds to behaviors established as high teaming performance. The average values indicate that overall teaming performance was generally considered strong by students, as a rating of three corresponds to expected typical teaming performance. However, the ratings are not so high as to leave no room for improvement in teaming skills for most students, especially given the substantial variability around the averages shown by the standard deviations. As noted in Chapter 3, the most common growth outcome for students rated against a single item is no change, for both ENGR131 and ENGR141.

In Section 3.6.5, a threshold for educational relevance of change or growth in CATME ratings was set as 0.10 on the 5-point scale. Educationally relevant magnitude is not a guarantee of statistical significance (discussed in the next section), but does serve as a metric to identify interesting results for discussion. It can be seen that changes or growth of magnitude 0.10 appear regularly in the tables, across sections and conditions, including several instances of negative growth. In ENGR131, the control sections show an overall growth of 0.12, having begun with lower average ratings than the experimental sections and surpassing the ratings of the experimental sections by the third administration. However, at no point do the overall scale ratings differ by the 0.10 relevance threshold between the experimental and control sections in ENGR131, though some individual items do at some time points. Overall, ENGR131 results show limited growth for both the experimental and control sections.

Examining individual sections in ENGR131, EXP1 is exceptional in that educationally relevant negative growth is observed on all scale items. Contrasting this with EXP2 and EXP3, which both have educationally relevant measured growth on three items and no educationally relevant growth on the others, EXP1 appears to be atypical. Reviewing the sample characteristics in Table 3.1, no special demographic differences would seem to explain EXP1's low performance.

In ENGR141, there is no educationally relevant difference in growth on the overall scale between the experimental and control sections. The only educationally relevant measured difference in item growth is for Expecting Quality, where the control sections have negative growth of -0.17 and the experimental sections have -0.07.

Examining individual ENGR141 sections, it is noted that EXP1 shows educationally relevant growth on four of the five items and no educationally meaningful change in the fifth. There is only one other instance of educationally relevant growth at the section level in ENGR141, in CTRL1, though there are several drops greater than -0.10. Therefore, the performance of EXP1 in promoting CATME growth appears to be exceptional compared with other ENGR141 sections. In an inverse feat, EXP2 shows educationally relevant negative growth on all five items. Given the high performance of EXP1, this suggests that the experimental condition is not the dominant factor in CATME growth promotion in ENGR141. Figure 4.4 illustrates this result.

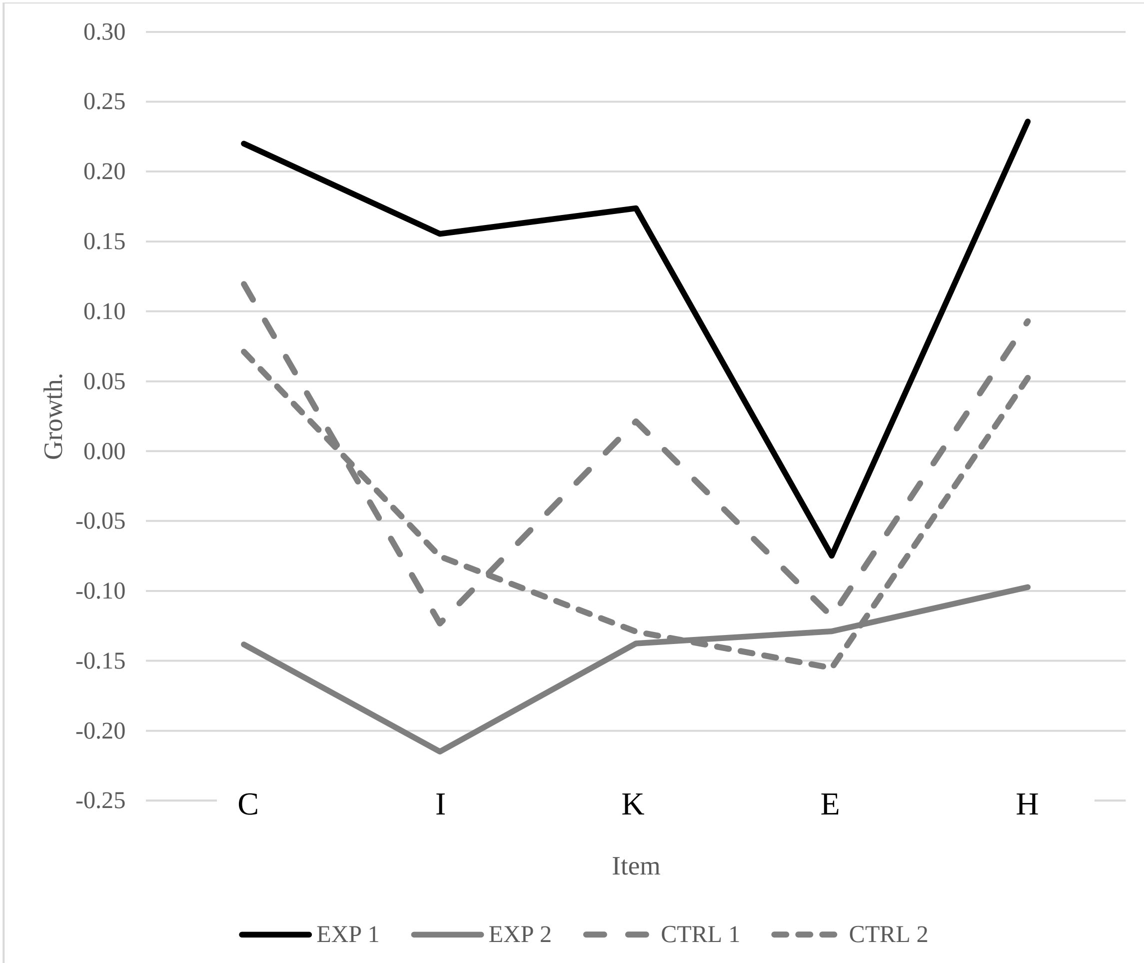


Figure 4.4: ENGR141 Per-Section CATME BARS Item Growth (3rd-1st Administration)

4.3.2 Kruskal-Wallace Testing

In ENGR131, the mean ranks for each item at each administration between the experimental and control cohorts were tested for statistically significant differences at $\alpha=0.05$ using the Kruskal-Wallace test. At the first administration, only item K was found to be statistically significantly different, being higher in the experimental sections ($p=0.015$). However at the third administration only item C was found to be different ($p=0.045$), being higher in the control sections. It is noted that if the atypical performance of ENGR131's EXP1 is removed from the data set, both K at administration 1 and C at administration 3 are no longer statistically significantly different. This finding reinforce the impression from examining the results tables that the differences in CATME ratings between the experimental and control cohort are not substantial across the semester in ENGR131.

The differences in growth between the experimental and control cohorts in ENGR131 was found to be significant for four of the five items (the exception is item I, Interacting with Teammates) and for the overall per-student average of the scale, with the highest p-value being equal to 0.01. For all items and the overall scale, the control cohort's growth values were statistically significantly higher than those of the experimental section. However, when the atypical EXP1 section is removed, only item K retains statistical significance (p=0.033), and for this item the control section started with lower ratings. The lower ratings are not still statistically significant with EXP1's data removed from the data set (p=0.062) but it remains plausible that the control sections gained growth in this area versus the experimental sections more easily due to lower initial ratings.

In ENGR141, testing for differences between the experimental and control sections at the first and third administration, the results are not significant with the exception of item K (Keeping the Team on Track) in the first CATME administration. In that case, the mean rank of the experimental cohort is found to be statistically significantly lower than the control cohort with p=0.001. Growth is not significantly different between the experimental and control cohorts for any item.

However, when CATME BARS growth was tested for differences between individual sections, all items except E were significantly different, with the highest p-value being 0.032. This indicates that there is more difference between sections than across the experimental and control conditions, as would be expected based on the results previously discussed and shown in Figure 4.4. When the results from the low-performing EXP2 section are removed, EXP1 is found to have statistically significantly higher growth in item K (Keeping the Team on Track) ($p=0.002$) and to have statistically significantly higher overall (all items averaged) growth than the control cohort ($p=0.016$). However, with only one experimental section, section level effects cannot be distinguished from the experimental effects.

4.4 Research Question 4 Results

Research question 4: To what extent did the improvement in the targeted metacognitive capabilities or psychological safety correlate with improved individual student teaming performance?

The results of testing for correlation (Spearman Rho) between individual students' responses on the metacognitive frequency survey and individual students' change in CATME BARS between the first and final CATME administration are given in Table 4.15. It can be seen that only no statistically significant correlations were found.

Table 4.15: Spearman Rho Correlation between Metacognitive Frequency and CATME BARS Ratings Growth

Item	ENGR131 Admin 1 Correlation	ENGR131 Admin 3 Correlation	ENGR141 Admin 1 Correlation	ENGR141 Admin 3 Correlation
C	-0.011	0.065	0.047	0.046
I	0.017	0.028	-0.089	0.039
K	0.067	0.037	0.104	0.057
E	0.005	0.077	0.99	0.121
H	0.055	0.026	0.075	0.066
Overall	0.024	0.062	0.057	0.086
N	509	386	247	236

Note: * is statistically significant at $\alpha=0.05$, ** is statistically significant at $\alpha=0.01$

The results of testing for correlation (Spearman Rho) between individual students' scores from the psychological safety scale and individual students' change in CATME BARS ratings between the first and final CATME administration are given in Table 4.16. It can be seen that higher psychological safety in the first ENGR131 administration had a statistically significant negative correlation with CATME items C (p=0.016), I (p=0.011), H (p=0.043) and the overall (average of other ratings) rating (p=0.044).

Positive correlations between high psychological safety on the second administration of the instrument and CATME BARS growth on item E (Expecting Quality) were found for both ENGR131 (p=0.019) and ENGR141 (p=0.038). In ENGR141, high psychological safety was also found to be positively correlated with C (p=0.004), I (p=0.045), and the overall CATME average rating (p=0.016).

Table 4.16: Psychological Safety Spearman Rho Correlation with Change in CATME BARS Ratings between First and Final Administration of CATME BARS

	ENGR131		ENGR141	
Item	**Admin 1**	**Admin 2**	**Admin 1**	**Admin 2**
C	-0.118*	0.026	-0.006	0.213**
I	-0.123*	-0.029	-0.013	0.148*
K	-0.02	0.083	-0.037	-0.031
E	-0.025	0.123*	0.047	0.153*
H	-0.099*	0.011	0.483	0.119
Overall	-0.098*	0.059	-0.036	0.177*
N	423	363	228	184

Note: * is statistically significant at α=0.05, ** is statistically significant at α=0.01

4.5 Case Study Results

This section presents short case studies exploring individual teams' psychological safety and other relevant data across the semester. These cases do not directly address any research question, but do provide specific examples of the behavior of some teams participating in the study for illustration and discussion. Cases include teams with high growth in psychological safety, consistently high psychological safety across the semester, high initial psychological safety that decreased sharply over the semester, and consistently low psychological safety. Cases in these areas are presented in turn.

4.5.1 Cases with High Psychological Safety Growth

Four teams were selected with high growth in psychological safety across the semester, two in ENGR141 and two in ENGR131. In each course, one case was taken from the experimental cohort and one from the control, though it is not possible to draw general conclusions about the experimental interventions from these cases.

4.5.1.1 Case 1 – ENGR131 Experimental

For the team with high growth in psychological safety from an ENGR131 experimental section, the average increase in psychological safety was 0.75 on a seven-point scale (1 is low safety, 7 is very high, and 4 is neutral), representing more than 10% of the scale in average growth. The team was all male and possessed one international student (mainland Chinese), one African-American student, and two white domestic students. Three students initially reported moderate psychological safety (just above neutral) while the fourth student reported very high psychological safety. Over the term, of the three students who initially reported moderate psychological safety, two later reported psychological safety more than one point higher than before, while the third increased only marginally.

One notable area of growth in psychological safety was the ability to bring up tough issues in the team, where ratings of 5, 2, 6, and 5 went to 5, 5, 7, 7. One area notable for its consistency was the team not rejecting others for being different, which went from 6, 5, 7, 7 to 6, 7, 7, 7. This team was apparently able to integrate all members, despite some demographic differences. However, CATME teaming performance growth was marginal, with three members being rated slightly worse on the final administration and the fourth only marginally improved. The improved member was the African-American student, who primarily was rated more highly in H, having relevant knowledge and skills (by a full 1 point on the five-point scale), at the end of the term.

Teaming motivation overall held steady, but this was accomplished by one student's teaming motivation dropping sharply (a white student) while another's (the Chinese student's) rose an equivalent amount. Team satisfaction did not appreciably change, and the students earned A's and B's. CATME text responses were variations on generic pleasure with the team, such as "Our group is still working well together! Especially after hearing stories from some other groups, I fell (sic) pretty lucky that I ended up with who I did."

Overall, it appears that while some team members may have had some initial concerns about psychological safety on the team, they found they were able to raise tough issues and clearly did not reject any members' contributions. They did not appear to suffer any particular difficulties or challenges as a group, and their teaming performance, motivation, and satisfaction largely held steady, leading to only marginal gains and losses in teaming performance as measured by CATME. Overall, the growth in psychological safety shown by this team could plausibly be attributed to time spent working together without negative experiences – a virtuous cycle of nothing going wrong breeding a greater sense of safety.

4.5.1.2 Case 2 – ENGR131 Control

For the team with high growth in psychological safety from an ENGR131 control section, the average increase in psychological safety was 1.28 on a seven-point scale (1 is low safety, 7 is very high, and 4 is neutral), representing more than 15% of the scale in average growth. The team was two men and two women and possessed one international student (Indian subcontinent), and three white domestic students. The same three students completed the initial and final psychological safety surveys. All three students initially reported psychological safety hovering around neutral. All three students later reported large gains in psychological safety, in two cases more than an entire point on average across the scale.

One notable area of growth in psychological safety was whether students thought it was likely that the team would hold mistakes against someone, rising from ratings of 4, 4, 6 to 7, 7, 6. Ratings on the expectation that the team could bring up and handle tough issues also increased, from 5, 4, 5, to 7, 6, 6. However, taking a risk on the team appears to have been a trouble point, with ratings of 3, 4, 3 only rising to 3, 4, 4, neutral at best, and the team disagreed sharply on whether one team member might actively undermine another, with ratings of 6, 4, 3 (initial), and 7, 1, 2 (final). The team appears to have had confidence in some of their group processes, but not in the actions or reactions of at least some individual members. It does seem incongruous to see multiple ratings implying that team members were actively working against each other alongside high confidence that mistakes would not be held against team members.

CATME teaming performance growth was marginal, with gains in C, I and K being cancelled about by losses in E and H for the three domestic students. Ratings for the international student were initially very low but matched those of the rest of the team by the end of the term, leading to substantial gains in teaming performance. Teaming motivation overall was high and steady across the board. Team satisfaction varied sharply between students, ranging from 2 to 4 on the five-point scale. All of the students earned A's, aside from the international student who failed based on excessive absences (she otherwise would have earned a B).

CATME text responses from the first administration show that the team was not working together well yet, including comments like "We work more as individuals than as a team" and "The only big problem have is that the team isn't really meeting up that much. Also there haven't been many classes with all team members present." By the second CATME administration, the situation seems to have improved but earlier conflict is also mentioned: "Our team is improving greatly. We have all members participating efficiently. We are now able to get more work done in a shorter amount of time due to less conflict." By the third administration, several team members report being pleased with the team, in comments like "I am proud of the way the team completed its assignments over the course" and "I am very happy with the way me and my team have synchronized our efforts to complete the tasks on time."

Overall, it appears that absences by team members may have contributed to the initially low measures of psychological safety – it may be difficult to trust someone to treat others well who skips obligations and potentially does not do their share of the work. As attendance by all members became more regular confidence in at least some aspects of psychological safety increased. Further evidence for teammates working against each other or undermining each other was not observed outside of the psychological safety scale. No driving event or mechanism to bind the team together and promote psychological safety in specific areas was observed, but it seems plausible that the "conflict" mentioned in early CATME feedback served as a turning point to unify the team. Once unified, the team worked together successfully and developed high confidence in team processes relating to psychological safety.

4.5.1.3 Case 3 – ENGR141 Experimental

Growth in average psychological safety reported by members of this team was 0.86 on the seven-point scale (1 is low safety, 7 is very high, and 4 is neutral), representing more than 10% of the scale in average growth. The team was two men and two women, three white domestic students and one African-American student. The students reported gains in psychological safety that varied in magnitude across the team, from a low of 0.4 to a high of 1.3.

One notable area of disagreement on psychological safety in the team on the first administration was whether or not it was safe to take a risk while working in the team, with ratings of 7, 6, 4, and 2. The rating of 2 was given by the African-American student. However, by the end of the term the ratings had improved to 7, 7, 5, 5, above neutral for all members of the team. Other aspects of psychological safety show more limited growth.

CATME teaming performance growth was moderate, with gains in C and I being cancelled out by losses in K, E, and H for most team members. This may reflect the team becoming more comfortable together in part by lowering expectations of the team and its members as the term progressed. Teaming motivation decreased somewhat, but consistently across team members and items. Team satisfaction held constant at around 4 on the five point scale for all team members. The students earned two A's, a B, and a C in the course.

One CATME text responses from the first administration suggests the team may have bonded socially but not professionally: "I feel that my team has bonded extremely well and we have become close friends. However, I feel that at times this makes us unproductive and unwilling to work on the task at hand." At the next CATME administration another team member echoed this sentiment: "I feel that my team has grown very close, but this might not always be a positive because we are not as formal and businesslike in our meetings." No useful comments were submitted with the third CATME administration.

Overall, this team's growth in psychological safety appears to have been driven by the successful inclusion of an initially-cautious student and the strong social bonding of the team. The team developed trust in the conduct of its members, but did not focus the efforts of those members towards class goals. The mixed results on other aspects of team performance such as CATME (where Keeping the Team on Track and Expecting Quality dropped conspicuously across the team) and final grades may reflect the socially-friendly-but-not-professionally-effective nature of the team mentioned in the comments.

4.5.1.4 Case 4 – ENGR141 Control

Growth in average psychological safety reported by members of this team was 0.82 on the seven-point scale (1 is low safety, 7 is very high, and 4 is neutral), representing more than 10% of the scale in average growth. The team was two men and two women, all white domestic students. The students reported gains in psychological safety that varied in magnitude across the team, from a low of 0.4 to a high of 1.3.

One notable area of disagreement on psychological safety in the team on the first administration was whether or not members of the team might be rejected for being different, with ratings of 6, 4, 3, and 5, which continued to be of concern to some team members, with final ratings of 5, 6, 4, and 5. Other aspects of psychological safety showed relatively uniform growth.

CATME teaming performance growth was good, with gains principally in C, I, and H with some minor losses in K and E for most team members. This may reflect the team becoming more comfortable together in part by lowering expectations of the team while giving more credit to individual members as the term progressed. Teaming motivation decreased by about 10% of scale, consistently across team members and items. Team satisfaction rose sharply, from 3.75 to 4.8 on average, on a five-point scale. The students earned two A's, a B, and a C in the course.

CATME text responses for the first administration include "Teaming is tricky" and "For the most part, our team has worked together fine" but responses grew more enthusiastic by the final administration with "Very glad I got placed with this team. At first I was a little hesitant but grew used to them over time" and "My team works together very well and we've become good friends as well" both being collected then.

Overall, while some areas of potential tension are apparent throughout the term on the psychological safety instrument, it appears that the team as a whole managed to improve their interactions, contribute work evenly, and perhaps lowered their expectations for the team's performance as psychological safety and team satisfaction rose.

4.5.2 Cases with Consistently High Psychological Safety

Two teams were selected with consistently high psychological safety across the semester, one in ENGR141 and one in ENGR131. Both cases happened to come from the experimental cohort, but it is not possible to draw general conclusions about the experimental interventions from these cases.

4.5.2.1 Case 5 - ENGR131 Experimental

This team reported psychological safety averaging around 6 on the 7-point scale for both the initial and final measurements. The team was all male and possessed one international student (Indian subcontinent) and three white domestic students. No items on the psychological safety instrument appeared to differ from the others in terms of ratings. However, CATME teaming performance grew noticeably across the term, with all but one student improving on average by more than 0.5 on the five point scale. E, Expecting Quality, showed by far the most growth, all students growing more than 0.5 and two students growing 1.5 points on the five point scale. Teaming motivation and teaming satisfaction were both high and consistent. Every member of the team earned an A in the course. In CATME comments, members of the team consistently expressed satisfaction with the team, with text such as “Overall, I fell (sic) that the team functions well together and does not have any problems doing tasks or deciding roles when solving problems.”

Overall, the high psychological safety on this team appears to be only one of several high markers of team quality present. Teaming motivation, satisfaction, and performance both in CATME and in final grades were very high. This team may be a case where students of comparable (high) academic and teaming preparation and motivation were placed together. It is possible that the team benefitted from the experimental materials, but no evidence to support this possibility was observed.

4.5.2.2 Case 6 - ENGR141 Experimental

This team reported psychological safety averaging around 6.5 on the 7-point scale for both the initial and final measurements. The team was two men and two women, all white domestic students. Over the term, the team's views diverged on whether members were able to bring up tough issues, going from 6, 7, 6, 7 to 6, 6, 5, 2. It seems likely that some important issues arose during the term that the team was not able to bring up. However, the team built on strength in their views on whether the team made good use of each individual's unique skills and abilities, going from 6, 5, 6, 7, to 6, 7, 7, 7, so it seems likely that all members felt they were able to contribute effectively.

CATME teaming performance dropped across the board, from an average of -0.1 to -0.45 points on average, on the five point scale. The largest and most consistent drops were in I, Interacting with Teammates, suggesting that the team may have been able to work together and retain high psychological safety as interactions deteriorated. Some ratings of I on the final CATME administration were 2's on the five-point scale, indicating some instances of poor performance. Teaming motivation and teaming satisfaction both fell by about 15% of scale. Team members earned two A and two B course final grades.

In CATME comments, one comment at the first administration was very enthusiastic: "I do not know what system you used to put together the teams, but I cannot imagine a better match. My team communicates praise and critiques very openly, there is an almost perfect delegation of team member roles to match personalities, and our personalities, while different, highlight the good in each other rather than causing conflict. I am extremely satisfied with my placement and hope that our team is not changed." At the third administration, team members held differing opinions about team roles, with conflicting comments collected. The first was "It was nice to see the team roles change for project 3 as each member found their strengths in programming, organizing, and building the robot" and the second was "As project 3 began to peak, individuals became very settled in their respective roles. This confined the learning to one area of the project so that individuals constructing the robot were clueless at coding and the individuals writing the report weren't very involved in building…I think that the roles may have had too much emphasis. I was disappointed by the lack of input I was able to make on the construction of the robot…" This comment would seem to put in context the team reports of using each individual's skills well, suggesting that while skills were used well it might not always have been ideal for learning. This may have been one of the important issues that team members were not able to bring up.

Overall, this team appears to have improved enough in some areas of psychological safety to offset erosion in other areas of psychological safety as deteriorating interactions between teammates and unresolved issues in the team were allowed to endure.

4.5.3 Cases with High Initial Psychological Safety that Decreases

Two teams were selected with large negative growth in psychological safety across the semester, one in ENGR141 and one in ENGR131. Both cases happened to come from the experimental cohort, but it is not possible to draw general conclusions about the experimental interventions from these cases.

4.5.3.1 Case 7 - ENGR131 Experimental

This team showed an average drop in psychological safety across the semester of -1.9 on the seven point scale, starting from an initial value above 6. The team was all male and possessed two Hispanic students and two white students. Only three students completed the first and final psychological safety surveys, with the missing student being different across measurements, making the comparison of initial and final psychological safety in this team less robust. However, few teams with initially high psychological safety had large drops in psychological safety across the term and the presence of extreme disagreements between raters on psychological safety in this team merits further review.

Team members differed sharply about the state of team psychological safety. In the final measurement of the ability of team members to bring up tough issues for discussion, team members gave ratings of 7, 4, and 1, being the highest possible value, the neutral value, and the lowest possible value. The prompt on whether one teammate would undermine another is also divided in the final measurement, at 5, 6, and 1. It appears that one or more team members may have been oblivious to serious issues in the team.

CATME teaming performance was reported as strong (most team members received ratings of 4 in most categories) but ratings decreased across the term, with all four students showing negative growth on average across the scale (-0.7, -0.7, -0.7, -0.2 on a five-point scale). Average teaming motivation overall dropped by 0.6 on a seven-point scale, with drops coming roughly evenly from each member of the team. Counterintuitively, team satisfaction actually increased by 0.7, from 4 to almost the top of the five-point scale. Every member of the team earned an A in the class.

Some CATME text responses showed disagreement about the state of the team. At the first administration, the comment “Our team works well together” was collected alongside “I feel that one of the teammates does not care much for taking part of the assignments. Moreover…many times he would spawn an irrelevant conversation in the middle of a meeting or in class when we are in the middle of working on an assignment. I must admit I have not talked to him about it nor has this been spoken of as a team.” However, no useful further comments were submitted by any member of the team.

Overall, it appears that there were hidden disagreements or ruptures in this team that may not ever have been brought to the surface. Some members of the team may not even have been aware of the feelings of other team members. The strong disagreement over whether it was possible to raise tough issues in the team certainly points in this direction. However, the overall lessening in CATME scores does not appear to be due to a subset of raters, and the general decline in teaming motivation suggests a more general unhappiness with the function of the team. However, the broad increase in teaming satisfaction to almost the top of the scale would seem to suggest that team members were happy to have been placed together. It is possible that this satisfaction was driven by high grades in the face of other troubles, or that high ratings were given on this instrument to conceal issues on the team. While there is seemingly contradictory information here, there were decreases in both psychological safety and teaming skills performance across the term, which suggests that some form of trouble was present in the team.

4.5.3.2 Case 8 - ENGR141 Control

This team showed an average drop in psychological safety across the semester of -1.57 on the seven point scale, starting from an initial value above 6. The team was composed of two men and two women, all domestic, two white, one African-American, and one multiracial. The average psychological safety reported for each team member declined across the term by a minimum of 1.2, 20% of scale.

The largest drops were observed in the categories of whether the team might reject someone for being different (4, 7, 7, 7 to 5, 6, 5, 4), whether members of the team might undermine each other (6, 7, 6, 7 to 4, 5, 6, 5), and whether the team made use of each member's unique skills (7, 6, 6, 5 to 4, 5, 5, 3). It seems likely that more than one incident or ongoing problem would be needed to induce all of these changes.

CATME teaming performance declined across the term, with all four students showing negative growth on average (-0.5, -0.1, -0.5, -0.6 on a five-point scale). It is notable that on the third administration, two team members consistently rate the other two as performing poorly in C, Contributing to the Team's Work. Average teaming motivation overall dropped by 0.9 on a seven-point scale, with drops coming roughly evenly from each member of the team and item. The low number of responses to the team satisfaction questions makes the overall values meaningless as a method of comparison. The team earned two C's and two D's.

An early CATME text response showed good feelings about the team "Knowing this is confidential, I still would like to say that I think I have a great, well-functioning team." However, by the second administration problems were apparent to team members, with the comments "I feel like while our team has not taken the time to air grievances about the team out with one another, that would be helpful to all of us" and "It's very difficult working with my team…My interactions with my team are good at some points, but always very, very distracted. Beyond the normal level of distraction a team would normally have." By the third administration, at least one team member believes that the team has not been contributing evenly to tasks: "…it would have been nice if more of the team contributed to the major projects instead of just having two people do most of the work." This comment echoes the low ratings for contributing in CATME earlier.

Overall, it appears that likely that this team as a whole may have been underprepared for the rigor of the course or perhaps engineering school in general. With no single member of the team earning even a B, the under-contributing team members may have been overwhelmed by the material of their coursework. Failure to perform well as a team early in the term would likely increase stress and conflict, increasing the likelihood of team members taking actions that might undermine psychological safety.

Strong psychological safety early in the term was undone by the team largely disintegrating under the stress of the semester and the work to be performed.

4.5.4 Case with Consistently Low Psychological Safety

While teams did submit surveys with average psychological safeties of neutral or below, very few were the product of multiple raters who completed both the initial and final surveys. Most teams with low initial psychological safety increased their scores across the term. In the case of ENGR131, no team with consistently low psychological safety was identified. One team in ENGR141 was identified with low and decreasing average psychological safety across the team and is reviewed below.

4.5.4.1 Case 9 – ENGR141 Experimental

This team showed an average drop in psychological safety across the semester of -0.68 on the seven point scale, starting from an initial value of 4.6, just above the neutral rating. The team was composed of three women and one man, all domestic, three white, and one Asian-American. The initial average psychological safety was so low primarily due to the rating of a single member of the team, as overall initial averages were 5.6, 5.7, 5.1, and 1.9. With the exception of the rating of 1.9, the other values are not extraordinary. However, at the end of the term, average ratings were 3.4, 3.6, 4.6, and 4, just above and below the neutral mark and among the lowest recorded across all the teams. Note that while the initially low-rating member greatly improved their score, all other members of the team decreased.

Changes in psychological safety were most pronounced in members being able to bring up tough issues (7, 7, 5, 2 to 2, 1, 4, 5) and whether team members might undermine each other (2, 7, 6, 1 to 6, 3, 3, 2). There appears to be consensus that the team was not strong at raising tough issues. There were large differences in opinion on whether team members might be undermining each other – as three of the four students marked below neutral on that item on the final administration of the instrument, it seems likely that one or more students were in fact undermining other members of the team during the term.

CATME teaming performance declined across the term, with three of the four students showing negative growth on average (0.2, -0.3, -0.35, -0.6 on a five-point scale). Drops in performance were concentrated in C, I, and E, indicating that unequal contributions, poor interactions, and lower expected quality became apparent across the term. Average teaming motivation overall dropped by a seemingly low 0.3 on a seven-point scale, with drops coming roughly evenly from each member of the team and item. It may be that team members were able to separate their unhappiness with the current teaming situation from their desire to learn more about teaming. Average team satisfaction dropped from 3.2 to 2.6 on a five point scale across the semester. Members of the team earned an A, two C's, and a D in the course.

Early CATME comments reflect problems in the team related to skill level: "The range of skill level makes it difficult for the group to work cohesively" and "I can get easily frustrated with my team…I think that part of that comes from my frustration with not fully understanding the material". At the second administration, specific skills and course grades are mentioned "I have severe worries about completing Project 3 on time, if at all. This is due to the balance of skills on our team: most of our team is very good at presentations and technical writing, while few can code or build the robot quickly and reasonably. Also, we've been discussing Engr 141 grades a lot recently, and the team dynamic has certainly changed because of it." At the third administration, fewer useful comments were submitted, one being "[student name removed] and I are very different which is strenuous." The comments examined show consistent concerns about differences in personality and skill. Skill is potentially related to academic preparation in this instance, which seems congruous with the course grades earned.

Overall, it appears that differentials in academic preparation and personality may have contributed to power struggles on this team. Given that some of these issues were identified in comments from the first CATME administration, some members of the team may have observed issues quite early in the term. One team member felt very strongly that there were serious issues in psychological safety early in the term; by the end of the term that student's outlook had improved somewhat, but the rest of the team lowered their assessment of psychological safety considerably, retaining the low overall score observed. The problems do not appear to have been resolved across the term, and may have simply become more visible.

4.5.5 Case Study Coda

These cases, while not generalizable, provide interesting examples and counterexamples of how psychological safety can change (or stay the same) as teams generally work together well or poorly and succeed or fail at their tasks. For example psychological safety was seen to rise while team interactions either improved (Case 3, Case 4) or declined (Case 1). Psychological safety can be high for teams raising or lowering their expectations for quality work (Case 2 and Case 4, respectively). In these cases, unique combinations of people and circumstances came together to produce unique outcomes, showing some of the richness that is abstracted out at the level of quantitative analysis employed in the rest of this dissertation.

CHAPTER 5. DISCUSSION & CONCLUSIONS

5.1 Research Question 1 Discussion and Conclusions

Research Question 1: To what extent did the experimental intervention promote the target aspects of metacognition in FYE students in the experimental group in comparison to the control group?

Metacognitive growth relating to teaming skills and processes was expected to be observed in both the experimental and control cohorts of this study. The existing teaming instruction, support, and learning experiences in both ENGR131 and ENGR141 are non-trivial and are intended to support student teaming skills growth. Prior research has shown growth in student teaming skills performance in ENGR131 (Jimenez-Useche et al., 2015). Veenman (2006, p. 8) states that "metacognitive knowledge and skills… become more sophisticated and academically oriented whenever formal educational requires the explicit utilization of a metacognitive repertoire." While the specific metacognitive skills of interest have not previously been widely tested for growth in ENGR131 or ENGR141, it would be reasonable to expect at least some students to exhibit development in these areas due to the existing (control) instruction. Then, due to the additional development opportunities afforded the experimental sections, additional metacognitive development was expected to be observed in the experimental sections.

In ENGR131, greater growth in the experimental sections is observed, but the scale of the difference is educationally marginal and not statistically significant. Differences between the sections are statistically significant, and do not appear to be driven by sample characteristics. This points to instructor differences or sampling procedures across sections as potentially dominant factors. The much higher than expected rates of negative metacognitive frequency reported growth shown in Table 4.5 (more than twice as high as 'no change' and two-thirds as high as positive growth) suggest that the instrument itself has poor test-retest reliability or that the data collection procedure may not have adequately controlled the working activities during the sampling period. Differences in working period task characteristics as a potential source of error were not carefully considered during study design, but are potentially quite damaging to the credibility of the data gathered. It was assumed that selection of similar activities would be sufficient to control differences in this area, but the difference between a period of strong team interaction and a period where the team was simply working individually on team tasks at the same table might be sufficient to lower the need for teaming metacognition, and thus measured teaming metacognition. This is one potential explanation for the high proportion of negative growth observed.

In ENGR141, further evidence of unsatisfactory instrumentation appears, as the control sections were found to provide statistically significantly better acquisition of metacognitive skills at the end of the term, despite results at the first administration not being statistically differentiable. It is more plausible that data collection procedures and/or the instrument itself are flawed than that the intervention is actively harmful to student metacognitive skills gain. It is also possible that factors such as the effects of section faculty, GTA, time-of-day, and starting student composition influenced the results. However, the student sample information in Table 3.1 suggests baseline comparability in student cohorts per-class. The potential effects of section faculty and TA's are more difficult to estimate with the given data.

Additional alternative potential explanations for the lower growth in the experimental cohort include Type 1 error or students rating themselves more harshly in experimental sections due to higher expectations for teaming metacognition performance later in the term. Students consistently engaging with exercises that highlight metacognition may cease being 'unskilled and unaware' about metacognition, while the control sections continued in blissful ignorance. This effect was described in the landmark paper by Kruger and Dunning (1999). This mechanism may have been present in the pilot study (Rynearson & Hynes, 2015), but results there were also potentially influenced by clear initial differences between the experimental and control cohorts. Collecting data on how good students estimated themselves to be at metacognition at points throughout the term might have allowed a quantification of the effects of ignorance and knowledge, but no way to eliminate the effect of more knowledge of metacognition on the study itself was seen. It is possible that results in both ENGR131 and ENGR141 are depressed by this effect, but enough growth occurred in ENGR131 to offset that depression and still achieve higher ratings than the control cohort.

Overall, the conclusions reached with respect to teaming metacognition are that the study's measurements and analysis do not provide clear evidence of experimental efficacy, and further that too much evidence exists that the instrumentation or collection procedures may be flawed to place much credence in the measurements taken. Potential weaknesses in the instrument and experimental procedures need to be resolved before future use.

5.2 Research Question 2 Discussion and Conclusions

Research Question 2: To what extent did the experimental intervention promote psychological safety of student teams in the experimental group in comparison to the control group?

At the beginning of the study, expectations of what would be observed with respect to the promotion of psychological safety in ENGR131 and ENGR141 were limited. The extent to which teams in these courses were psychologically safe for students had not previously been investigated. It was expected that the experimental sections, subject to interventions designed to promote psychological safety, would exhibit a higher level of psychological safety across the semester.

Contrary to those expectations, it is notable that near zero or negative growth (averaged across the seven-item scale) was reported for all experimental conditions except the ENGR131 control cohort. However, the results show that the experimental cohorts' ratings of psychological safety for the first administration of the scale were higher than the control cohort's ratings. In ENGR131, this difference was statistically significant. This may be an effect of the intervention's altered Code of Cooperation assignment for experimental sections raising psychological safety early in the term, and having that safety hold steady or be eroded by actual team working events over time. This would act to reduce apparent growth in the experimental cohorts.

Negative growth in psychological safety does not seem unreasonable in ENGR131 and ENGR14 as teams are put under stress and interactions between students become more fraught. Initial positive impressions and accord could fray as stakes and disagreements arise. Case studies 7, 8, and 9 in Chapter 4 provide some examples of teams whose psychological safety declined over the term for various reasons. The case studies in general illustrate the diversity of possible paths for increasing and decreasing psychological safety. The lack of growth in the experimental section from the higher starting point may also represent something of a psychological safety ‘ceiling’ where conditions in ENGR131 and ENGR141 teams are conducive to growth in psychological safety up a point, beyond which growth for at least some teams is unlikely.

Higher psychological safety at any time point is generally desirable (recall that it is associated with learning from teammates and as a team in general – not just in teaming skills), so if the experimental intervention is capable of raising psychological safety early in the term, potentially speeding team formation and moving student teams more rapidly towards effective working and learning conditions, it may well be worthwhile to employ it more generally in courses with team-based learning aspects. This outcome may be what is observed in the ENGR131 results. The ENGR141 experimental results show more (but not statistically significant) negative growth, which is worth consideration.

There are a number of potential causes for negative growth. First is that the intervention was actually creating a false or poorly-supported appearance of psychological safety that was not robust against teaming challenges. Second, it is possible that ENGR141’s higher challenge level places more stress on teams, leading to more conflict and more negative growth in psychological safety. Case 8, in Chapter 4, appears to be an example of this process. Third, the population or instruction of ENGR141 may affect outcomes in psychological safety in some way. For instance, the Honors students may be more prone to initial higher psychological safety due to higher self-efficacy.

Overall, the study results suggest the experimental intervention can raise psychological safety early in the term, but the effects over time were comparable to the control sections. The case studies in Chapter 4 examine some downstream occurrences from initially high and low psychological safety, but broad quantitative analysis to determine if there are common downstream effects of high initial psychological safety remains a future work.

Finally, finding that overall psychological safety ratings were adequate-to-good for most students, while not central to this study, is positive news for both ENGR131 and ENGR141. ENGR131 and ENGR141 faculty may wish to consider continuing to assess psychological safety in their courses, potentially using teams with low psychological safety as a warning sign for required intervention. The psychological safety instrument employed in this study is easy to administer, quick, and appears to have found results that make theoretical sense in the classroom environment. It appears to be suitable for continued use.

5.3 Research Question 3 Discussion and Conclusions

Research Question 3: In the event that the intervention succeeds in supporting more metacognitive skills growth and/or psychological safety in the experimental than the *control groups, to what extent did the teaming performance of the experimental group then improve beyond that of the control group?*

Teaming skills performance growth was expected to be observed in both the experimental and control cohorts of this study. The existing teaming instruction, support, and learning experiences in both ENGR131 and ENGR141 are non-trivial and are intended to support student teaming skills growth. As discussed, previous research has shown growth in student teaming skills performance in ENGR131. It was further expected that the intervention, promoting psychological safety and metacognitive *awareness* of opportunities to practice teaming skills, would result in teaming skills growth in the experimental sections beyond that observed in the control sections.

In ENGR131, growth was observed for both the experimental and control cohorts with the exception of item H (Having relevant Knowledge, Skills, and Abilities). Against expectations, more growth was observed in the control sections than the experimental sections and for most items this difference was statistically significant. However, of the three experimental sections, only Experimental Section 1 actually showed substantial negative growth. Average growth values in the other two experimental sections both fall between the average growth values for the control sections. This suggests that factors not directly related to the intervention led to the negative growth observed in Experimental Section 1. When the seemingly atypical results of the first experimental section are removed from the analysis, the experimental and control cohorts are no longer statistically significantly different.

In ENGR141, growth shown in both the experimental and control sections was limited in scale and not statistically significant. However, with one experimental section clearly obtaining the best growth results, one experimental section clearly obtaining the worst results, and both control sections falling into the middle between them, it is difficult to say that the intervention had clear results on teaming performance in ENGR141.

It can be noted that Experimental Section 1 likely represented a best-case-scenario for fidelity of study implementation methods as the author taught that section. However, the fidelity of implementation in Experimental Section 2 is not known to have been poor. It can be stated that between-section differences (which were statistically significant) were more notable in ENGR141 than differences between the experimental and control cohorts, and that on average Experimental Section 1 substantially outperformed the control sections while Experimental Section 2 substantially underperformed them. The extent to which these results reflect on the intervention and not confounding factors is unclear. The experimental intervention does not appear to have raised the standard deviation of CATME measurements in comparison to the control section, which would be expected if the intervention itself prompted more extreme measurements. It seems most likely that the observed differences reflect on the raw chance of team composition and each instructor's efforts to monitor and support strong teaming in their classes. While the efforts to support teaming beyond standard course practices in Experimental Section 2 are not clear, in Experimental Section 1, the instructor gave personalized written feedback on CATME results to several dozen students over the term. Such feedback included encouragements, advice, and warnings which may have prompted members of the class to pursue higher teaming performance. With the effort to collect data on instructor practices abandoned due to broad non-compliance, it is difficult to contextualize these results. Future investigations may benefit from data on instructor practices in implementing teaming learning activities.

Overall, no robust evidence of the efficacy of the intervention with respect to growth in CATME performance was observed. However, as it is not clear to what extent the intervention delivered on the intended growth in teaming metacognition and especially the *awareness* step of the teaming decision making model employed in this study. As psychological safety was intended to bolster the later *motivation* step, gains in psychological safety in the experimental sections are not necessarily intended to provide gains in teaming practice, and therefore performance, without growth in the earlier *awareness* step that may not have been present. A more detailed study of the effects of changes in early-term psychological safety can influence teaming results (beyond the illustrative case studies) remains a future work.

5.4 Research Question 4 Discussion and Conclusions

Research Question 4: To what extent did the improvement in the targeted metacognitive capabilities or psychological safety correlate with improved individual student teaming performance?

The experimental intervention was designed on the premise that increasing teaming metacognition and increasing psychological safety would both lead to increased teaming skills performance. The intervention's effects on metacognition, psychological safety, and teaming skills performance have previously been discussed at the level of experimental and control cohorts and sections. However, it makes sense that individual students with higher teaming metacognition and psychological safety should see benefits to teaming performance even if the higher teaming metacognition and psychological safety arise in a control section. Therefore, positive correlations could be expected between higher teaming metacognition and teaming skills performance and between higher psychological safety and teaming skills performance.

However, no statistically significant correlations were found between on either administration of the metacognitive frequency survey across ENGR131 and ENGR141. This finding aligns with the earlier conclusion that the instrumentation and experimental procedures employed in assessing metacognition were unsatisfactory. The alternative is that the theorized relationship between teaming metacognition and teaming skills growth is not detectable in the sample. Given the known difficulties in assessing metacognition, the former possibility seems more probable.

Examining psychological safety, higher psychological safety was expected to result in greater teaming skills growth as previously discussed. The results from ENGR131, the only statistically significant results for the first administration of the instrument, suggest that high psychological safety early in the term may have a negative relationship with teaming skills growth for CATME items C, I, and H, leading to a statistically significant negative correlation between psychological safety and growth in CATME ratings. If the intervention successfully raises psychological safety early in the term with meaningful negative effects on teaming growth, that would be both unexpected and a serious problem with the intervention. The negative growth might be related to the previously mentioned overconfidence or the ‘unskilled and unaware’ effect found by Kruger and Dunning (1999) – teams blithely rating themselves highly on both CATME and psychological safety early in the term, then being disabused of their notions of personal teaming skills excellence later in the term as unforeseen negative events occurred. No method to conclusively test this hypothesis is seen with existing data, however this sequence of events could potentially be examined with interview or short-answer data focused on events or occurrences that changed a student's opinion about another team member or the team itself, in either a positive or negative direction. No similar effect was observed in ENGR141.

The positive correlations between psychological safety on the final administration of and CATME growth in ENGR141 (and to a lesser extent 131) align more closely with expectations. Students on teams with high psychological safety late in the term seem likely to have been in a team environment with either consistently high or increasing psychological safety, allowing more safe space for the practice of teaming skills, leading to improvement in teaming skills. However, it is also possible that these correlations are due instead to a relationship in the negative direction – teams that started with high psychological safety which fell due to negative teaming events during the term also seem likely to be teams where negative CATME growth would be expected. It could be that one, the other, or a combination of these scenarios explains these results. The case studies in Chapter 4 have examples that appear to align with each scenario. Further analyses of these relationships, including examining the data on a per-team basis, remain potential future works.

Finally, it is noted that given the number of correlation tests performed and the standard for significance selected (α=0.05) one or more of the results found to be significant could represent Type 1 error. However, the fact that the number of significant correlations in some areas far exceeds the number that would be expected due to Type 1 error suggests the majority of such results are legitimate.

5.5 Summary of Conclusions

In summary, efforts to address the first research question were not successful due to unsatisfactory instrumentation and study procedures. Assessing metacognition is notoriously difficult, and the methods employed in this study do not seem to have been successful, judging by the inconsistent and implausible results collected.

The intervention does seem to have been successful in raising early-term psychological safety in the ENGR131 audience. As the altered Code of Cooperation assignments and other aspects of the intervention relating to psychological safety are relatively simple and consume little time to implement, further investigation is merited to determine if these results are reproducible. Theory suggests that psychological safety is almost universally good for learning in teams, but the correlation results and case studies imply that high psychological safety early in the term may sometimes represent blindness to potential problems rather than psychological safety rooted in strong team processes and norms. It is important to determine if the early-term psychological safety fostered by the intervention primarily represents one or the other of these outcomes. Whether teams with strong psychological safety and teams with a strong impression of psychological safety can be differentiated in a scaleable way is an interesting question with potential implications for practice.

With unclear and potentially null results for the intervention in terms of metacognitive development, which provides the foundation for the predicted growth in teaming skills via *awareness* and the other steps of the teaming decision making process, it is not surprising that the experimental sections do not perform significantly better in CATME skills growth. It is noted that the correlation results show teams with high psychological safety at the end of the term tending to report more growth in CATME skills. This clearly aligns with theory, supports the suitability of the psychological safety instrument for classroom use, and demonstrates that psychological safety and CATME skills acquisition can be related under at least some circumstances.

5.6 Study Limitations

This study's strengths and limitations are tightly entwined. In general, by developing interventions and assessments based upon theory and putting them into practice, the study obtained large-scale results from authentic FYE environments. However, substantial control over the administration of the intervention was lost and a number of reasonable and compelling ideas for the intervention were either curtailed or omitted to create an intervention of appropriate scale.

In more detail, key limitations existed in the experimental design, experimental administration, and experimental assessment. In experimental design, it should be noted that this study is only quasi-experimental – it is a typical and unfortunate limitation of educational research that students and faculty cannot be randomly assigned to suit experimental conditions. While the sections selected avoided listed learning communities and made an effort to select an appropriate mix of times-of-day and faculty experience levels, numerous potentially confounding factors including faculty interest in teaming, faculty capabilities in providing teaming feedback, section GTA conduct, and to some extent differences in student preparation and predilection for learning in the target areas may have affected the results. Also, in order to get permission to collect data in ENGR131 and ENGR141, the intervention needed to fit into available class time. For this reason, some potential aspects of the intervention such as the regulatory checklist (RC) were omitted from the study, as discussed in Chapter 3. While the intervention represents the synthesis of a number of promising ideas and methods for the promotion of teaming metacognition, other promising methods remain untested in engineering environments at this time.

In administration, it proved very difficult to get sufficient information from some participating faculty about their conduct and implementation of the intervention or various assessments. A plan to collect information from faculty on these items was abandoned after it became apparent that only the most diligent faculty, putting the most care into implementation, were putting in the effort to report on their practice. As mentioned in Chapter 3, it occurred twice that redo or make-up administrations of various instruments were required after participating faculty missed an assessment entirely or implemented it in a fundamentally flawed way, leading to unusable results. It should be assumed that some aspects of the intervention were applied in unintended ways or omitted by some faculty, some of the time. At this time, there does not appear to be a way to quantify these divergences. This adds another source of variability between the experimental conditions and specific sections and limits the study's ability to definitively report on both methods and results. However, administrative difficulties are expected in this type of study and the substantial response rates to the various instruments are strong evidence that participating faculty did not abandon the study outright.

Finally, in assessment, it must be noted again that disagreement exists in the metacognition literature on the best practices for assessing metacognition and that no pre-existing validated instruments targeting the metacognitive processes of interest, especially at scale, were or are available. Instruments and methods were developed based upon best practices for both 'online' and 'offline' assessment of metacognition (note that only the 'offline' assessment was discussed in this study – the metacognitive frequency survey), but these instruments have not undergone rigorous development and validation. The difficulty in robustly assessing metacognition on the large scale is a significant limitation of this study.

Another limitation in assessment was the employment of the CATME BARS to measure teaming skills performance. While administratively essential to this effort due to its integration into both ENGR131 and ENGR141 and ability to scale, the balance struck by the CATME BARS between usability and detailed data collection potentially limited this study's ability to differentiate the effects of the intervention on teaming skills performance using the analytic methods selected for the study. Briefly, this is because to limit the number of responses required of each student rater at each administration (limiting survey fatigue), there are only five possible ratings for each student. This can result in students getting the same rating across the semester if their growth is limited, making statistical difference more difficult to detect with the methods employed in this study. It is possible that more sophisticated analytic tools could potentially be employed in future works to circumvent some of these difficulties.

5.1 Future Works

The results and limitations of this study suggest a number of directions for potential future works. First, opportunities exist to examine data collected as a part of this study acquired with an eye towards further research, such as how demographic factors may influence results in the areas of interest. Demographic questions were not of primary concern in this dissertation, but further analysis drawing in more data and potentially employing more sophisticated quantitative models could provide additional insight into new and current research questions without further data collection. Such methods could allow for examining the effects of factors at the level of the team or clustering teams similar in some factors to explore common trends.

Second, as ENGR131 in particular is a hotbed of educational research, opportunities may exist to extend the data set collected for this study by combining it with one or more other extant data sets. Some discussion along these lines took place during study planning; it is known that potentially complimentary data was collected in several sections of ENGR131. Such additional data, including a deeper look into student demographic, gender and sexual orientation information, could be used to pursue new research questions or to facilitate new approaches to current research questions.

Third, the psychological safety results of this study invite further analysis to understand the relationship between psychological safety and other factors of interest, especially teaming skills performance. While the case studies provide some examples of how psychological safety and teaming skills performance (among other factors) might relate in a given team, the collected data could be turned to new research questions relating to psychological safety. One example would be working to identify ways that psychological safety can both support and potentially undermine teaming skills acquisition, as this study's results suggest may be occurring. It is noted that the research question on psychological safety in this dissertation was limited to whether the intervention was effective in promoting psychological safety.

Fourth, more sophisticated analytic techniques may allow pursuit of new research questions. Examining the relationships between multiple teaming-relevant factors simultaneously, while beyond the scope of this work, could enhance understanding and potentially practice for engineering student teams.

Finally, the mixed results of some assessments of this study with respect to the underlying theory suggest that more development-focused research efforts may be appropriate for future inquiries, especially in the area of metacognition. Some theories and examples incorporated in this study are primarily from non-educational contexts (such as the 'sales funnel') or have primarily been used educationally in non-engineering and often quite structured, lab-centric contexts, while this study attempted integration of theory and implementation in an authentic environment. Conclusive evidence about some hypotheses may not be possible to gather without better measurement of metacognition or other processes in the context of interest. Identification or development of more sensitive data collection procedures or instruments for metacognition and other areas of interest, likely on a smaller scale than this study, could lay a stronger foundation for future pursuit of the important research questions approached here. One question not considered in this study is whether students have different levels of ability to discern social cues (which seems likely) and what effects this may have on the types of metacognitive intervention attempted in this study. It may be also be fruitful to assess at what point students are departing the teaming decision making ‘sales funnel’, and why, to inform the design of future interventions.

REFERENCES

REFERENCES

ABET. (2013). Criteria for accrediting engineering programs. Baltimore, MD: Engineering Accreditation Commission.

Adams, E. A. (2014). *Improving learning productivity and teamwork skills in freshman engineer-ing students through conative understanding.* University of Illinois at Chicago.

Adams, R. S., Turns, J., & Atman, C. J. (2003). Educating effective engineering designers: The role of reflective practice. *Design studies*, *24*(3), 275-294.

Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General*, *136*(4), 569.

Barron, B. (2003). When smart groups fail. *The journal of the learning sciences*, *12*(3), 307-359.

Boiarsky, C. (2004). Teaching engineering students to communicate effectively: A metacognitive approach. *International Journal of Engineering Education*, *20*(2), 251-260.

Borkowski, J. G. (1985). Signs of intelligence: Strategy generalization and metacognition. *The growth of reflection in children*, *7*, 105-144.

Borkowski, J. G., Carr, M., & Pressley, M. (1987). “Spontaneous” strategy use: Perspectives from metacognitive theory. *Intelligence, 11*(1), 61-75.

Bradley, B. H., Postlethwaite, B. E., Klotz, A. C., Hamdani, M. R., & Brown, K. G. (2012). Reaping the benefits of task conflict in teams: The critical role of team psychological safety climate. *Journal of Applied Psychology, 97*(1), 151-158. doi: 10.1037/a0024200

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). How people learn: Brain, mind, experience, and school (expanded edition): Washington, DC: National Academy Press.

Breed, B., Mentz, E., & Van der Westhuizen, G. (2014). A metacognitive approach to pair programming: Influence on metacognitive awareness. *Electronic Journal of Research in Educational Psychology, 12*(1).

Brown, A. L. (1978). Knowing when, where, and how to remember: A problem of metacognition.

Brown, A. L. (1981). Metacognition: The development of selective attention strategies for learning from texts. *Directions in reading: Research and instruction*, 21-43.

Carmeli, A., & Gittell, J. H. (2009). High-quality relationships, psychological safety, and learning from failures in work organizations. *Journal of Organizational Behavior, 30*(6), 709-729.

Case, J., Gunstone, R., & Lewis, A. (2001). Students' metacognitive development in an innovative second year chemical engineering course. *Research in Science Education, 31*(3), 313-335.

Chi, M. T. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science, 1*(1), 73-105.

Coleman, E. B. (1998). Using explanatory knowledge during collaborative problem solving in science. *Journal of the Learning Sciences, 7*(3-4), 387-427.

Crawley, E. F. (2001). The cdio syllabus: A statement of goals for undergraduate engineering education. Boston, MA: MIT

Cunningham, P., Matusovich, H. M., Hunter, D. A. N., & McCord, R. E. (2015). *Teaching metacognition: Helping engineering students take ownership of their own learning* Paper presented at the IEEE Frontiers in Education Conference, El Paso, Texas.

Davey, B., & McBride, S. (1986). Effects of question-generation training on reading comprehension. *Journal of Educational Psychology, 78*(4), 256-262. doi: 10.1037/0022-0663.78.4.256

Donald, J. G. (2002). *Learning to think: Disciplinary perspectives* (1st Edition ed.). San Francisco: Jossey-Bass.

Duhigg, C. (2016, February 25, 2016). What google learned from its quest to build the perfect team. *The New York Times Magazine*.

Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative science quarterly, 44*(2), 350-383.

Edmondson, A. (2002). *Managing the risk of learning: Psychological safety in work teams*: Citeseer.

Edmondson, A., & Lei, Z. (2014). Psychological safety: The history, renaissance, and future of an interpersonal construct. *Annual Review of Organizational Psychology and Organizational Behavior, 1*(1), 23-43. doi: 10.1146/annurev-orgpsych-031413-091305

Engineers, P. U. O. o. F. (2016). Frequently asked questions. Retrieved 4/08/2016, 2016, from https://www.purdue.edu/futureengineers/What%20To%20Know/Common%20Questions/Frequently%20Asked%20Questions.html

Felder, R. M., & Brent, R. (2001). Effective strategies for cooperative learning. *Journal of Cooperation & Collaboration in College Teaching, 10*(2), 69-75.

Felps, W., Mitchell, T. R., & Byington, E. (2006). How, when, and why bad apples spoil the barrel: Negative group members and dysfunctional groups. *Research in organizational behavior*, *27*, 175-222.

Flavell, J. H. (1970). Developmental studies of mediated memory. *Advances in child development and behavior, 5*, 181-211.

Flavell, J. H. (1976). Metacognitive aspects of problem solving. *The nature of intelligence, 12*, 231-235.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American psychologist, 34*(10), 906.

Georghiades, P. (2004). From the general to the situated: Three decades of metacognition. *International Journal of Science Education, 26*(3), 365-383.

Gilbert, J. K. (2005). *Visualization in science education* (Vol. 1): Springer.

Gourgey, A. F. (2001). Metacognition in basic skills instruction. In H. J. Hartman (Ed.), *Metacognition in learning and instruction* (pp. 3-16): Springer.

Hacker, D. J., Dunlosky, J., & Graesser, A. C. (1998). *Metacognition in educational theory and practice*: Routledge.

Hansen, E. J. (2012). *Idea-based learning: A course design process to promote conceptual understanding*: Stylus Publishing, LLC.

Hrivnak, G. A. (2013). Catme smarter teamwork. *Academy of Management Learning & Education, 12*(4), 679-681. doi: 10.5465/amle.2013.0302

Hsiung, C. M. (2010). Identification of dysfunctional cooperative learning teams based on students' academic achievement. *Journal of Engineering Education, 99*(1), 45-54.

Jamieson, L. H., & Lohmann, J. R. (2012). Innovation with impact: Creating a culture for scholarly and systematic innovation in engineering education. *American Society for Engineering Education, Washington*.

Jeong, H., & Chi, M. T. (2007). Knowledge convergence and collaborative learning. *Instructional Science, 35*(4), 287-315.

Jimenez-Useche, I. C., Ohland, M. W., & Hoffmann, S. R. (2015). *Multicultural dynamics in first-year engineering teams in the U.S.* Paper presented at the ASEE Annual Conference & Exposition, Seattle, WA.

Johnson, D. W. (1991). *Cooperative learning: Increasing college faculty instructional productivity. Ashe-eric higher education report no. 4, 1991*: ERIC.

Johnson, D. W., & Johnson, R. T. (2009). An educational psychology success story: Social interdependence theory and cooperative learning. *Educational researcher, 38*(5), 365-379.

Johnson, D. W., Johnson, R. T., & Smith, K. A. (1998). Cooperative learning returns to college what evidence is there that it works? *Change: the magazine of higher learning, 30*(4), 26-35.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of personality and social psychology*, *77*(6), 1121.

Kuhn, D. (1999). A developmental model of critical thinking. *Educational researcher, 28*(2), 16-46.

Kuhn, D. (2000a). Metacognitive development. *Current directions in psychological science*, *9*(5), 178-181.

Kuhn, D. (2000b). Theory of mind, metacognition, and reasoning: A life-span perspective *Children's reasoning and the mind* (pp. 301-326).

Langfred, C. W., & Moye, N. (2014). Does conflict help or hinder creativity in teams? An examination of conflict's effects on creative processes and creative outcomes. *International Journal of Business and Management*, *9*(6), 30.

Layton, R. A., Loughry, M. L., Ohland, M. W., & Ricco, G. D. (2010). Design and validation of a web-based system for assigning members to teams using instructor-specified criteria. *Advances in Engineering Education, 2*(1), 1-28.

Livingston, J. A. (2003). Metacognition: An overview. Washington, DC: United States Department of Education.

Loughry, M. L., Ohland, M. W., & Woehr, D. J. (2014). Assessing teamwork skills for assurance of learning using catme team tools. *Journal of Marketing Education, 36*(1), 5-19.

Malone, D. (2011). Empirical evidence of the fairness and quality of peer evaluations. *Academy of Educational Leadership Journal, 15*(2), 129-140.

Martin, R., Maytham, B., Case, J., & Fraser, D. (2005). Engineering graduates' perceptions of how well they were prepared for work in industry. *European Journal of Engineering Education, 30*(2), 167-180.

May, E., & Strong, D. S. (2011). Is engineering education delivering what industry requires. *Proceedings of the Canadian Engineering Education Association*.

McDonald, J. H. (2014). *Handbook of biological statistics* (3rd ed.). Baltimore, Maryland: Sparky House Publishing.

McMasters, J., & Matsch, L. (1996). Desired attributes of an engineering graduate - an industry perspective *Advanced measurement and ground testing conference*: American Institute of Aeronautics and Astronautics.

Millard, E. (2012). 5 reasons flipped classrooms work. *University Business, 15*(11), 26-29.

Mithaug, D. E., Mithaug, D. K., Agrain, M., Martin, J. E., & Wehmeyer, M. L. (2007). *Self-instruction pedagogy: How to teach self-determined learning*: Charles C Thomas Publisher.

Nessel, D. D., & Graham, J. M. (2006). *Thinking strategies for student achievement: Improving learning across the curriculum, k-12*. Thousand Oaks, Calif. : Corwin Press

Newell, J., Dahm, K., Harvey, R., & Newell, H. (2004). Developing metacognitive engineering teams. *Chemical Engineering Education, 38*(4), 316-320.

Oakley, B. A., Felder, R. M., Brent, R., & Elhajj, I. (2004). Turning student groups into effective teams. *Journal of student centered learning, 2*(1), 9-34.

Oakley, B. A., Hanna, D. M., Kuzmyn, Z., & Felder, R. M. (2007). Best practices involving teamwork in the classroom: Results from a survey of 6435 engineering student respondents. *Education, IEEE Transactions on, 50*(3), 266-272.

Ohland, M. W., Loughry, M. L., Woehr, D. J., Bullard, L. G., Felder, R. M., Finelli, C. J., . . . Schmucker, D. G. (2012). The comprehensive assessment of team member effectiveness: Development of a behaviorally anchored rating scale for self- and peer evaluation. *Academy of Management Learning & Education, 11*(4), 609-630. doi: 10.5465/amle.2010.0177

Ostafichuk, P., Hodgson, A., Sophie Bartek, & Naylor, C. (2010). *Teaching team dynamics: Experiences in second-year mechanical engineering design*. Paper presented at the 6th international CDIO conference, École Polytechnique, Montréal.

Pelz, D. C., & Andrews, F. M. (1966). Scientists in organizations: Productive climates for research and development.

Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of educational psychology, 82*(1), 33.

Prince, M. J. (2004). Does active learning work? A review of the research. *Journal of engineering education, 93*(3), 223-231.

Pung, C., & Farris, J. (2011). *Assessment of the catme peer evaluation tool effectiveness*. Paper presented at the ASEE 2011 Annual Conference & Exposition, Vancouver, BC.

Reid, K., Hertenstein, T., Fennel, G. T., Spingola, E. M., & Reeping, D. (2013). *Development of a first-year engineering course classification scheme*. Paper presented at the 120th ASEE Annual Conference & Exposition, Atlanta, Georgia.

Rynearson, L., & Hynes, M. (2015). *Design and impact of a classroom intervention to support teaming metacognition*. Paper presented at the Frontiers in Education 2015, El Paso, Texas.

Schraw, G. (2001). Promoting general metacognitive awareness *Metacognition in learning and instruction* (pp. 3-16): Springer.

Schwartz, D. L. (1995). The emergence of abstract representations in dyad problem solving. *The Journal of the Learning Sciences, 4*(3), 321-354.

Sears, D. A., & Reagin, J. M. (2013). Individual versus collaborative problem solving: Divergent outcomes depending on task complexity. *Instructional Science, 41*(6), 1153-1172. doi: 10.1007/s11251-013-9271-8

Seat, E., & Lord, S. M. (1999). Enabling effective engineering teams: A program for teaching interaction skills*. *Journal of Engineering Education, 88*(4), 385-390.

Seat, E., Parsons, J. R., & Poppen, W. A. (2001). Enabling engineering performance skills: A program to teach communication, leadership, and teamwork*. *Journal of Engineering Education, 90*(1), 7-12. doi: 10.1002/j.2168-9830.2001.tb00561.x

Slavin, R. E. (1996). Research on cooperative learning and achievement: What we know, what we need to know. *Contemporary educational psychology, 21*(1), 43-69.

Smith, K. A., Sheppard, S. D., Johnson, D. W., & Johnson, R. T. (2005). Pedagogies of engagement: Classroom-based practices. *Journal of Engineering Education, 94*(1), 87-101. doi: 10.1002/j.2168-9830.2005.tb00831.x

Spearman, C. (1923). *The nature of" intelligence" and the principles of cognition*. Oxford, England: Macmillan.

Stephens, C. S. (2001). *A meta-analysis of research on student team effectiveness: A proposed application of phased interventions*. Paper presented at the International Conference on Informatics Education & Research, New Orleans, LA.

Svarovsky, G. N., & Shaffer, D. W. (2006). *Design meetings and design notebooks as tools for reflection in the engineering design course.* Paper presented at the Frontiers in Education Conference, 36th Annual, San Diego, California.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive science*, *12*(2), 257-285.

Tarricone, P. (Producer). (2011a, 6/1/2015). Conceptual framework of metacognition. Retrieved from http://media.routledgeweb.com/pp/common/sample-chapters/9781841698694-figure.pdf

Tarricone, P. (2011b). *The taxonomy of metacognition*. New York, NY: Psychology Press.

Thompson, V. A., Prowse Turner, J. A., & Pennycook, G. (2011). Intuition, reason, and metacognition. *Cognitive psychology, 63*(3), 107-140.

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of educational Research, 68*(3), 249-276.

Vanasupa, L., Stolk, J., & Herter, R. J. (2009). The four-domain development diagram: A guide for holistic design of effective learning experiences for the twenty-first century engineer. *Journal of Engineering Education, 98*(1), 67-81.

Veenman, M. V. J., Van Hout-Wolters, B. H. A. M., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. *Metacognition and Learning, 1*(1), 3-14. doi: 10.1007/s11409-006-6893-0

Wankat, P. C., & Oreovicz, F. S. (2014). *Teaching engineering* (2nd Edition ed.). West Lafayette, Indiana: Purdue University press.

Weimer, M. (2013). Assessing team members. *The Teaching Professor, 27*(April 2013). http://info.catme.org/wp-content/uploads/Teaching-Professor-April-3013.pdf

Wheldon, T. (2013). Reducing time to degree by cutting credit creep *Capitol Research*. Lexington, KY: The Council of State Governments.

Wijaya, B. S. (2015). The development of hierarchy of effects model in advertising. *International Research Journal of Business Studies, 5*(1).

Zohar, A., & Barzilai, S. (2013). A review of research on metacognition in science education: Current and future directions. *Studies in Science Education, 49*(2), 121-169. doi: 10.1080/03057267.2013.847261

APPENDICES

Appendix A Surveys

PURDUE UNIVERSITY

This survey will help us improve the class. You should complete this survey only when directed to by your instructor. Your responses will not affect your grade but you should complete the survey in class when asked to. The survey will not be available later.

During the team activity preceding this survey, how frequently did the following situations occur?

I monitored the progress my team made overall and tried to change things if the progress was not adequate.

- Never
- Rarely
- Sometimes
- Often
- All the time

I monitored the progress an individual team member was making on a given task.

- Never
- Rarely
- Sometimes
- Often
- All the time

I noticed that I gave a teammate timely, specific, and constructive feedback .

- Never
- Rarely
- Sometimes
- Often
- All the time

I noticed that I asked for and showed an interest in a teammate's ideas and contributions.

- Never
- Rarely
- Sometimes
- Often
- All the time

I noticed that I made sure that my teammates stayed informed and understood each other.

- Never
- Rarely
- Sometimes
- Often
- All the time

I noticed that I provided encouragement or enthusiasm to the team.

- Never
- Rarely
- Sometimes
- Often
- All the time

I noticed that I asked for or respectfully received feedback from a teammate.

- Never
- Rarely
- Sometimes
- Often
- All the time

>>

Appendix Figure A.1: Metacognitive Frequency Survey

PURDUE
UNIVERSITY

The following questions ask about your impressions of aspects of your class teaming experience. Remember there are no right or wrong answers, just answer as accurately as possible based on what you believe to be true. Use the scale below to answer the questions. If you think the statement is very inaccurate for you, circle 1. If the statement is very accurate for you, circle 7. If the statement is moderately accurate for you, find the number between 1 and 7 that best describes you.

Your responses on this survey will not affect your course grades, but accurate answers will help us understand how to improve our courses.

If you make a mistake on this team, it is often held against you.

- 1 - Very Inaccurate
- 2
- 3
- 4
- 5
- 6
- 7 - Very Accurate

Members of this team are able to bring up problems and tough issues.

- 1 - Very Inaccurate
- 2
- 3
- 4
- 5
- 6
- 7 - Very Accurate

People on this team sometimes reject others for being different.

- 1 - Very Inaccurate
- 2
- 3
- 4
- 5
- 6
- 7 - Very Accurate

It is safe to take a risk on this team.

- 1 - Very Inaccurate
- 2
- 3
- 4
- 5
- 6
- 7 - Very Accurate

It is difficult to ask other members of this team for help.

- 1 - Very Inaccurate
- 2
- 3
- 4
- 5
- 6
- 7 - Very Accurate

No one on this team would deliberately act in a way that undermines my efforts.

- 1 - Very Inaccurate
- 2
- 3
- 4
- 5
- 6
- 7 - Very Accurate

Working with members of this team, my unique skills and talents are valued and utilized.

- 1 - Very Inaccurate
- 2
- 3
- 4
- 5
- 6
- 7 - Very Accurate

>>

Appendix Figure A.2: Psychological Safety Survey

The survey questions listed below are rated on a 7-point scale from 1, not at all true of me, to 7, very true of me.

1. I prefer class work that is challenging so I can learn new things.
2. It is important for me to learn the teaming skills taught in this class.
3. I like what I am learning about teaming skills in this class.
4. I think I will be able to use the teaming skills I learn in this class in other classes or later in my career.
5. I often choose paper topics I will learn something from even if they require more work.
6. Even when I do poorly on a test I try to learn from my mistakes.
7. I think that the teaming skills I am learning in this class are useful for me to know.
8. I think that what we are learning in this class about teaming skills is interesting.
9. Understanding teaming skills is important to me.

Appendix Figure A.3: Teaming Motivation Survey

Appendix B Forms

Teaming Strategy Evaluation Matrix

Name: ___________________________   Team Number: __________

Directions

1) Before beginning, review the Cues, Strategy, Why to Use, and How to Use columns.
2) Work to recognize cues in the team working environment while the team is working.
3) **Each** time you recognize a cue while the team is working, put a checkmark in the 'Times Cued' Column.
4) When you recognize a cue, try to employ an appropriate corresponding strategy.
5) **Each** time that you employ a strategy, place a checkmark in the 'Times Used' column.
6) Refer to the contents of the sheet as necessary to recall the cues and strategies.

Cues	Times Cued	Strategy	Why to Use	How to Use	Times Used
-Teammate shows you work -Teammate tells you about work -Teammate makes a plan or suggestion		Give Feedback	-Show appreciation -Help them learn -Help you learn -Discuss integration of work into project	-Listen carefully -Identify good parts first -Ask questions to understand -Critique work, not people -Build on good parts	
-Work being planned -Beginning to do work -Not feeling pride at work -Feeling disappointed -Trying to 'just be done'		Expect Quality	-Motivate learning -Team pride -Self-esteem -More satisfying to do good work than poor work	-Discuss team quality goals -Try hard and be enthusiastic -Work on progress, not perfection -Work to get better at tasks -Learn from previous mistakes	
-Relationship with team getting worse -Confusion -Anger, rudeness -Teammate talking too much or not enough -Power struggles -Not following Code of Cooperation		Interact Like a Pro	-Maintain working relationships -Improve learning opportunities -Better grades -Build trust -More fun than working on a team with poor interactions	-Diagnose causes of team problem before taking action -Constructive communication ('Will what I say improve the situation?') -Discuss important issues -Admit when you're wrong -Give benefit of doubt -Refer to Code of Cooperation -Update Code of Cooperation	
-Lack of progress -Frustration -Boredom -Discussion of unimportant issues -Team is distracted		Keep Team on Track	-Save time -Greater accomplishments -Greater learning -More satisfying than sitting around	-Work on key problems first -Only work all together on the same item if crucial or difficult -Make, monitor, update plans -Plan to finish early -Respect teammate's time	
-Waiting for teammates to finish their work -Teammates do most of the work -You tend to work on same kind of task		Contribute to Team's Work	-Greater personal accomplishments -Greater learning -Grow as an engineer -Try new things -Help others learn	-Make your work a credit to yourself and your team -Look for opportunities to try new types of work -Offer to assist others -Don't waste time, take action	

Appendix Figure B.4: Sample SM Worksheet

Team Code of Cooperation **Fall 2015**

Part 1: Names & Signatures

1. Complete the two tables below.
2. Once you complete and print the Code of Cooperation, have each team member enter his/her initials in the appropriate blue shaded column.

ENGR 131 Section Number	
Team Number	

Team Members

Note: *Your initials in the blue shaded column below indicate your approval of this Code of Cooperation.*

Name	**Initials for Version 1**	**Initials Version 2**

Part 2: Individual Guidelines

1. Review the individual guidelines provided in the table below. **These guidelines must remain in your Code of Cooperation!**
2. **Add at least one** (but not more than 3) **additional individual guidelines in the table below.**
    - Write each guideline so it completes the sentence "I agree to…"

I AGREE TO…

<table>
<tr><td>1</td><td colspan="2">Complete all assignments on time.</td></tr>
<tr><td>2</td><td colspan="2">Constructively criticize ideas, not individuals.</td></tr>
<tr><td>3</td><td colspan="2">Resolve conflicts promptly and constructively.</td></tr>
<tr><td>4</td><td colspan="2">Attend all team meetings, be on time, and be prepared.</td></tr>
<tr><td>5</td><td colspan="2">Encourage team members and allow everyone to participate.</td></tr>
<tr><td>6</td><td colspan="2">Take responsibility for the team's goals, progress, and success.</td></tr>
<tr><td>7</td><td colspan="2">Be an active listener and show respect for the contributions of other team members.</td></tr>
<tr><td></td><td>Complete this column for
Version 1</td><td>Complete this column for
Version 2
(add any new or revised individual guidelines)</td></tr>
<tr><td>8</td><td></td><td></td></tr>
<tr><td>9</td><td></td><td></td></tr>
<tr><td>10</td><td></td><td></td></tr>
</table>

Part 3: Team Guidelines

1. Create **at least 5** (but not more than 10) team guidelines. These guidelines should address topics such as the following; also see the example guidelines below:
   - How team roles will rotate
   - How meeting times will be determined and communicated
   - How the team will accomplish and communicate its work
   - How the team will ensure team assignments are turned in on time
2. Type each guideline into the table below.
   - Write each guideline so it completes the sentence "Our team agrees to…"

Example team guidelines:

- Have a pre-determined agenda (list of discussion topics) developed before each meeting.
- Meet on a weekly basis at a set location and time that works for all team members.
- Put cell phones on quiet and do not have other distracters, such as Facebook, open during team meetings.

Team Guidelines

OUR TEAM AGREES TO…

	Complete this column for Version 1	**Complete this column for Version 2** ***(add any new or revised team guidelines)***
1		
2		
3		
4		
5		
6		

7		
8		
9		
10		

Part 4: Psychological Safety

1. Create **at least 5** (but not more than 10) guidelines for increasing team psychological safety. These guidelines should address topics such as the following; also see the example guidelines below. How to ensure…
    - That honest mistakes are not held against individuals
    - That it is relatively easy and automatic that tough issues are brought up
    - That no members of the team are rejected for being different
    - That team members can take reasonable risks for educational reasons, such as opting to work on a part of a project that requires them to learn new skills
    - That is easy to ask other team members for help even if asking 'looks' bad
    - That the unique talents and characteristics of team members are valued and used
    - That no member of the team is the 'boss' or orders other team members around
2. Type each guideline into the table below.
    - Write each guideline so it completes the sentence "Our team agrees to…"

Example psychological safety guidelines:

- Select team working tasks according to what team member wish to learn rather than what they are already good at.
- Open each team meeting with a discussion of current problems or tough issues. Each team member will be individually asked each time to bring up any issues they are aware of and a list of issues from all team members made prior to the beginning of discussion.

Team Guidelines

OUR TEAM AGREES TO…

	Complete this column for Version 1	**Complete this column for Version 2** *(add any new or revised team guidelines)*
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		

Appendix Figure B.5: ENGR131 Experimental Code of Cooperation Modification

ENGR141 - Code of Cooperation Development Assignment

Background

Your team's Code of Cooperation sets norms and expectations, forming the 'rules' by which members of your team interact and work together. The Code of Cooperation is also a resource when conflict arises on the team – this document can determine whose conduct is not as agreed and what consequences that team member should face. Teams should regularly consult and update their Code of Cooperation as new circumstances arise.

Preparing a Code of Cooperation

A Code of Cooperation requires careful forethought. The decisions made in preparing the Code in this class are made very early in the term, before the team comes to discover often meaningful differences in preferred working and communication styles. Simply, team members don't know what problems they're likely to have within the team in the future. Many ENGR141 teams therefore prepare superficial Codes that assume there will not be major problems on their teams, assume that problems will be easily resolved by unanimous consent, or state that various methods will be used to resolve problems without giving sufficient detail in describing the exact methods to be used. When conflict arises, these Codes are not sufficient to resolve the problems and get the team back onto a constructive working path.

In preparing a Code of Cooperation, it is recommended that teams brainstorm a long list of possible issues, circumstances, and contingencies that could potentially happen on their team. What happens if someone is late to a meeting? How late? What if they miss a meeting? What if we agree that if they miss a meeting they have to bring food to the next one but they miss the next one also? While your teammates are likely pleasant and professional individuals, the first term of engineering school is very stressful for many students. In preparing the Code your team should consider a range of potential circumstances, from the everyday to the extreme. Your Code should in detail specify how team members and the team itself ought to act and also what the next steps or consequences will be if those expectations are not met.

Assignment Requirements

1. Bullet points or paragraphs are acceptable; use a format for your Code that fits your overall aims and your team's preferences.
2. Include at least ten individual norms and five group norms. These can include later items on this list. These numbers are the minimum for completion and do not reflect the work usually required to create an appropriate Code of Cooperation – keep going until your Code is comprehensive. You may wish to review the slides on teaming for potential topics for inclusion in your Code. The order, ranking, and/or labeling of items should make sense and make it easy for the reader to quickly comprehend your Code.

3. Include a detailed plan for the operations of your team. A few sample questions to be addressed include: when are meetings, how are meeting times communicated, how do team roles rotate? One common problem in ENGR141 is teams submitting work late due to miscommunication. What procedures could guarantee work being submitted on time?
4. Include a detailed plan to construct a psychologically safe environment on your team. What specific norms, procedures, actions, or consequences will guarantee your team is a psychologically safe working environment where all team members will learn and contribute as valued equals? Recall that creative, high-performing teams are usually highly psychologically safe.
5. Include detailed procedures to ensure that team member conduct is in accordance with your written Code. This usually involves a set of consequences for various failures or violations that escalates in severity. Include a step or stage where the team knows that they are not able to solve their problems alone and require the intervention of the instructor. This should be a serious step after one or more preliminary actions taken to resolve problems in the team. However, the course instructors are a key resource for teams in serious conflict or confusion and the Code should specify how the team knows when and how to access this resource.
6. Include a detailed plan on how your Code of Cooperation can be updated. A few sample questions include: how do you know when this needs to happen, what is the process, and who needs to agree? Teams in conflict should strongly consider revisions to their Code and an established method makes it much more likely this will happen.

Psychological Safety

As discussed in class, psychological safety has been demonstrated to lead to higher team performance, more pleasant working environments, and greater creativity and innovation. In a class context, psychological safety is important to allow students to try new things and learn new skills. Therefore, as students and engineers learning how to construct a psychologically safe team environment is a powerful and relevant skill to develop. In preparing the aspects of your Code addressing psychological safety, consider what specific, actionable, detailed methods you could use to ensure that:

- That honest mistakes are not held against individuals
- That it is relatively easy and automatic that problems and tough issues are brought up
- That no members of the team are rejected for being different
- That team members can take reasonable risks for educational reasons, such as opting to work on a part of a project that requires them to learn new skills
- That is easy to ask other team members for help even if asking ‘looks’ bad
- That the unique talents and characteristics of team members are valued and utilized
- That no member of the team is the ‘boss’ or orders other team members around

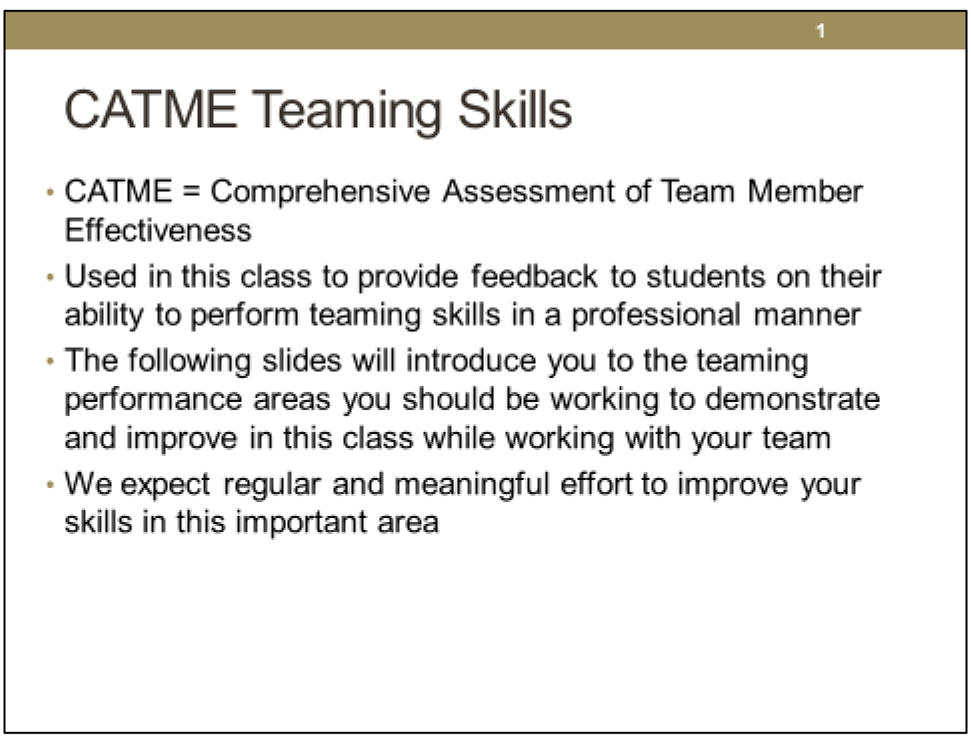
Some teams in the past have identified creative and fun phrases, methods, habits, and other team idiosyncrasies aimed at prioritizing psychological safety that were also fun/funny/contributed to team spirit. The Code does not need to be dreary – but it does need to be effective.

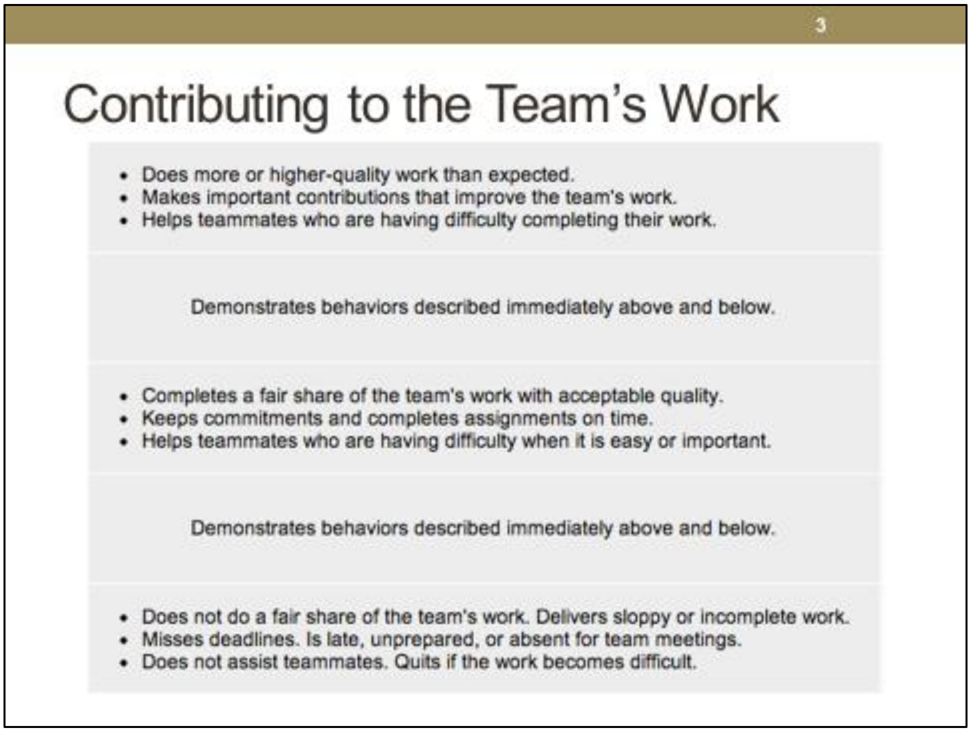
Notice
The instructors of ENGR141 have established minimum requirements for a Code of Cooperation above. At the instructor's discretion they may require teams who turn in Codes of Cooperation deemed inadequate (even those meeting the requirements above) to expand and reform their Codes.

Submission
Submit your Code of Cooperation by the due date via BlackBoard Learn. Use the file name `CoC_teamXX.docx`) where *XX* is your team number.

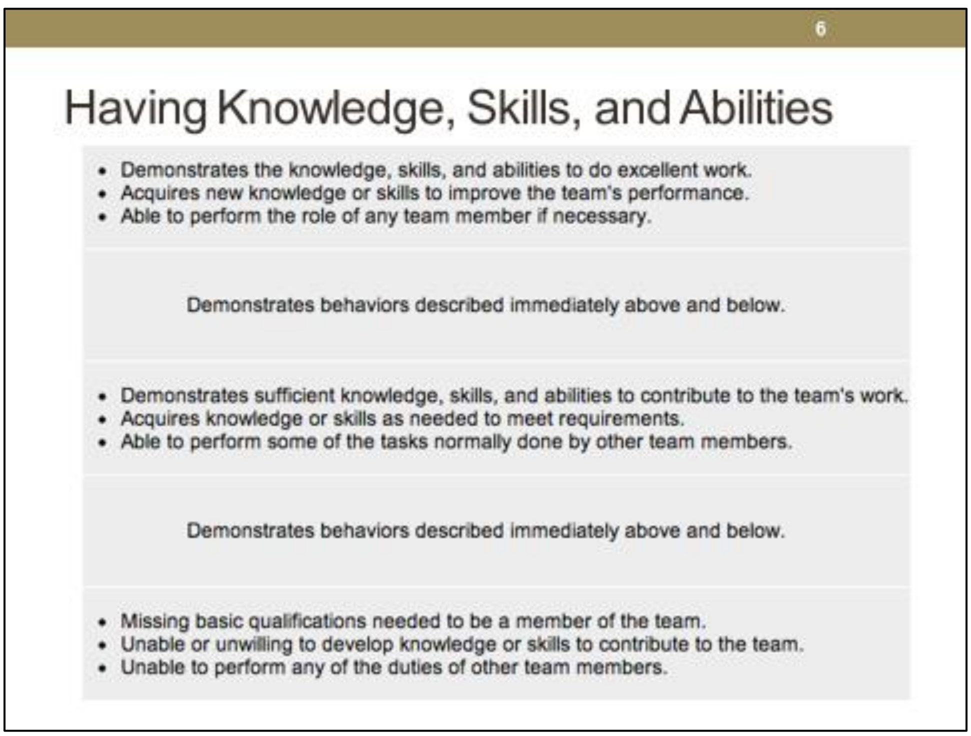
Appendix Figure B.6: ENGR141 Code of Cooperation Assignment

Appendix C Sample Slides




3
Contributing to the Team's Work
• Does more or higher-quality work than expected.
• Makes important contributions that improve the team's work.
• Helps teammates who are having difficulty completing their work.
Demonstrates behaviors described immediately above and below.
• Completes a fair share of the team's work with acceptable quality.
• Keeps commitments and completes assignments on time.
• Helps teammates who are having difficulty when it is easy or important.
Demonstrates behaviors described immediately above and below.
• Does not do a fair share of the team's work. Delivers sloppy or incomplete work.
• Misses deadlines. Is late, unprepared, or absent for team meetings.
• Does not assist teammates. Quits if the work becomes difficult.


4
Expecting Quality
• Motivates the team to do excellent work.
• Cares that the team does outstanding work, even if there is no additional reward.
• Believes that the team can do excellent work.
Demonstrates behaviors described immediately above and below.
• Encourages the team to do good work that meets all requirements.
• Wants the team to perform well enough to earn all available rewards.
• Believes that the team can fully meet its responsibilities.
Demonstrates behaviors described immediately above and below.
• Satisfied even if the team does not meet assigned standards.
• Wants the team to avoid work, even if it hurts the team.
• Doubts that the team can meet its requirements.



Appendix Figure C.7: ENGR131 & ENGR141 Discussion of CATME Skills

Introducing Teaming SEM

- As discussed, teaming is important for engineering
- Today we will introduce an additional tool that can help you improve your teaming abilities
- This tool is called a **S**trategy **E**valuation **M**atrix (SEM)
- SEM's can assist learners in taking more active control of their work, improving performance
- By noticing specific events or 'cues' in the team environment, you can select and practice better teaming strategies
- This SEM emphasizes the CATME skills

Using an SEM - Example

Cues	Strategy	Why to Use	How to Use
-Teammate shows you work -Teammate tells you about work -Teammate makes a plan or suggestion	Give Feedback	-Show appreciation -Help them learn -Help you learn -Discuss integration of work into project	-Listen carefully -Identify good parts first -Ask questions to understand -Critique work, not people -Build on good parts

3

SEM Working Period

- Please review the information on the sheet
- During the ten minute working period, the goal is to become more sensitive to the cues and use the strategies
- After the working period begins, put a checkmark in the 'Times Cued' column each time you notice a cue
- Make a checkmark in the 'Times Used' column each time you employ a listed strategy or similar
- Worksheets will be collected at the end of the working period

4

Continue to Use the SEM

- We recommend you continue to use and reference the Teaming SEM during team working periods both inside and outside of class
- A version of the SEM has been posted to BBL

Appendix Figure C.8: ENGR131 & ENGR141 Introduction of Teaming SM Sample Slides

1

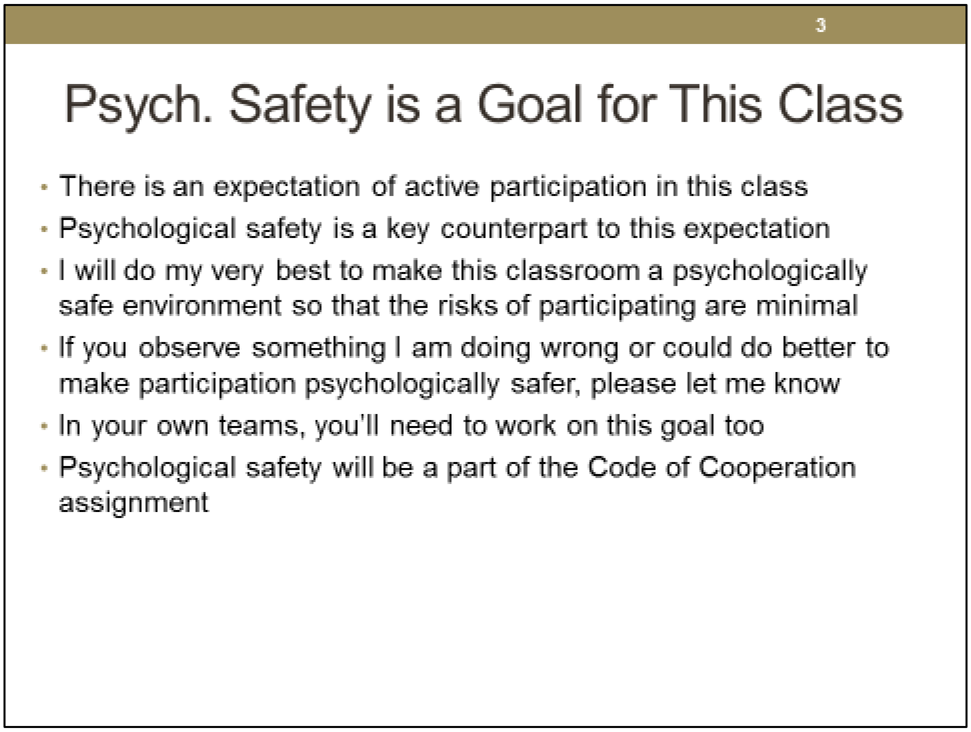
Introducing Psychological Safety

- Research on effective teams have shown that one characteristic usually observed in creative, high-performing teams is ‘psychological safety’
- Psychological safety is "a shared belief that the team is safe for interpersonal risk taking" (Edmondson 1999)
  - Risks like:
  - Giving new ideas that may or may not be good
  - Trying to accomplish tasks you may not be good at
  - Speaking up about tough problems or potential problems
  - Admitting mistakes
- Psychologically safe environments lead to better communication about important team and work issues because team members know they won’t be punished for trying to contribute

2

Some Indications of Psychological Safety

- Honest mistakes are not held against individuals
- No members of the team are rejected for being different
- Team members can take reasonable risks - in education this might include responding to questions where the answer is uncertain or trying out new skills on course projects
- It is easy to ask other team members for help even if asking ‘looks’ bad



Appendix Figure C.9: ENGR131 & ENGR141 Psychological Safety Introduction & Norm Setting Sample Slides

1

PE 06: Teaming Journal #1

- **What?**
  - In this assignment, you will reflect on your teaming experience in ENGR 131 to date.
- **Why?**
  - Reflecting on your teaming experience can help you become a better team member and ultimately help your team become more effective.
- **How?**
  - In Blackboard, click **Assignments** and locate the assignment **PE 6: Teaming Journal #1**.
  - Complete the assignment as directed.
    - ***Time = 10 minutes***

2

Teaming Reflection #1

- We previously reviewed the CATME teaming skills and used the Teaming Strategy Evaluation Matrix tool in class
- Recalling these discussions and the reflection you just completed, individually consider:
  - How much have you improved your teaming skills so far?
  - Do you notice various opportunities to employ teaming skills more often? Why or why not?
  - If you had to pick only one thing to do to improve your teaming skills, what would it be?
  - ***Time = 2 minutes***

3

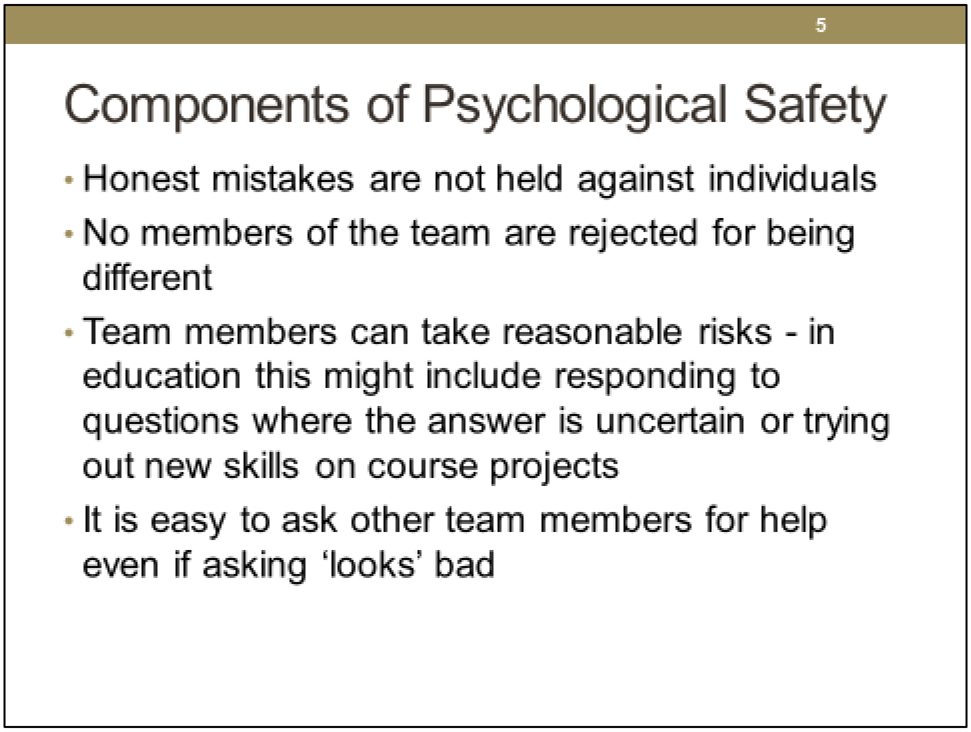
CATME Skills

- Interacting with Teammates
  - Listening, being polite, building understanding, being enthusiastic
- Contributing to the Team's Work
- Expecting Quality
  - High standards for work done by the team
- Keeping the Team on Track
- Having Knowledge, Skills, and Abilities
  - Including willingness to learn new skills

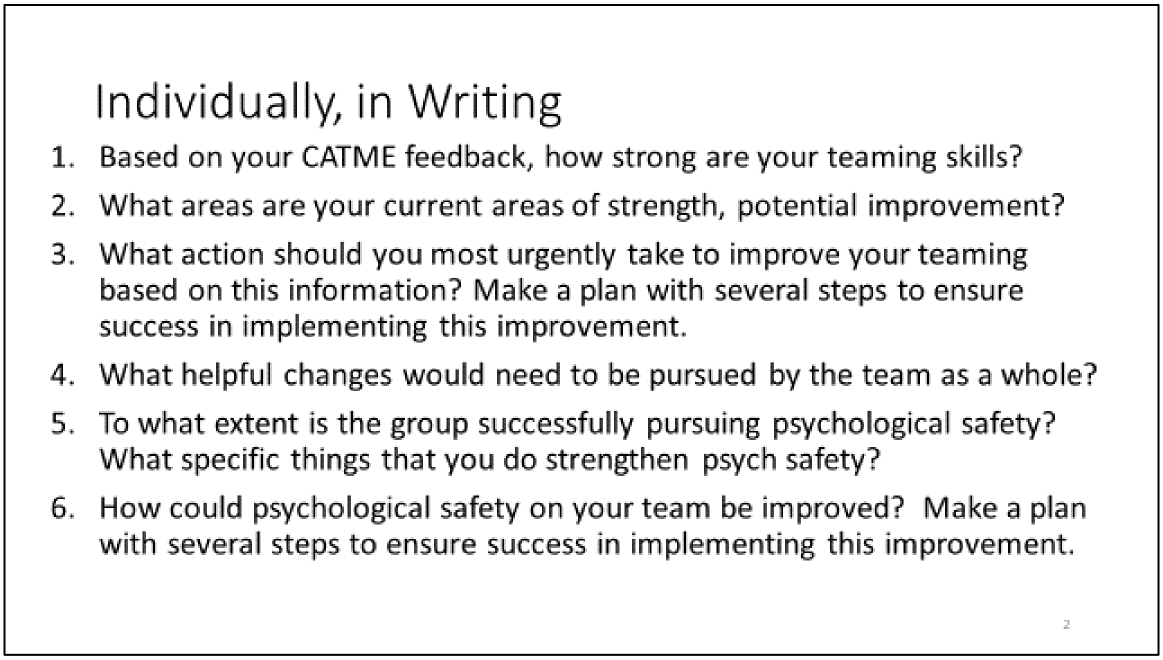
4

Teaming Reflection #2

- We previously reviewed the idea of psychological safety in class
- Recalling these discussions and the reflection you just completed, individually consider:
  - How strong is psychological safety on your team?
  - What is the weakest part of psychological safety on your team?
  - If you had to pick only one thing to do to improve your team's psychological safety, what would it be?
  - ***Time = 2 minutes***



Appendix Figure C.10: ENGR131 Week 7 Teaming and Psych Safety Reflection Slides



Appendix Figure C.11: ENGR141 Week 7 Teaming and Psych Safety Reflection Slides

1

Teaming Skills Monitoring Reflection

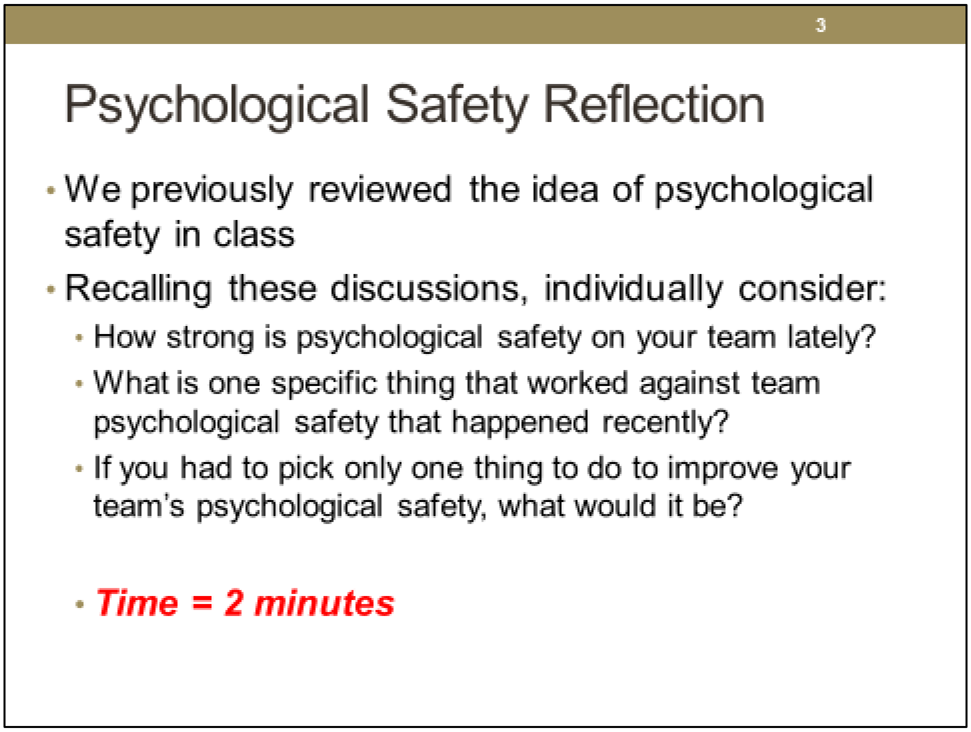
- We have discussed teaming skills and the Strategy Evaluation Matrix as a method to improve them
- Recalling this, individually consider:
  - Do you notice opportunities to use various teaming skills more often now than earlier in the term? Why or why not?
  - What is the most important teaming skill for you to get better at using regularly?
  - What is one practical thing you can do to increase your performance on a teaming skill?

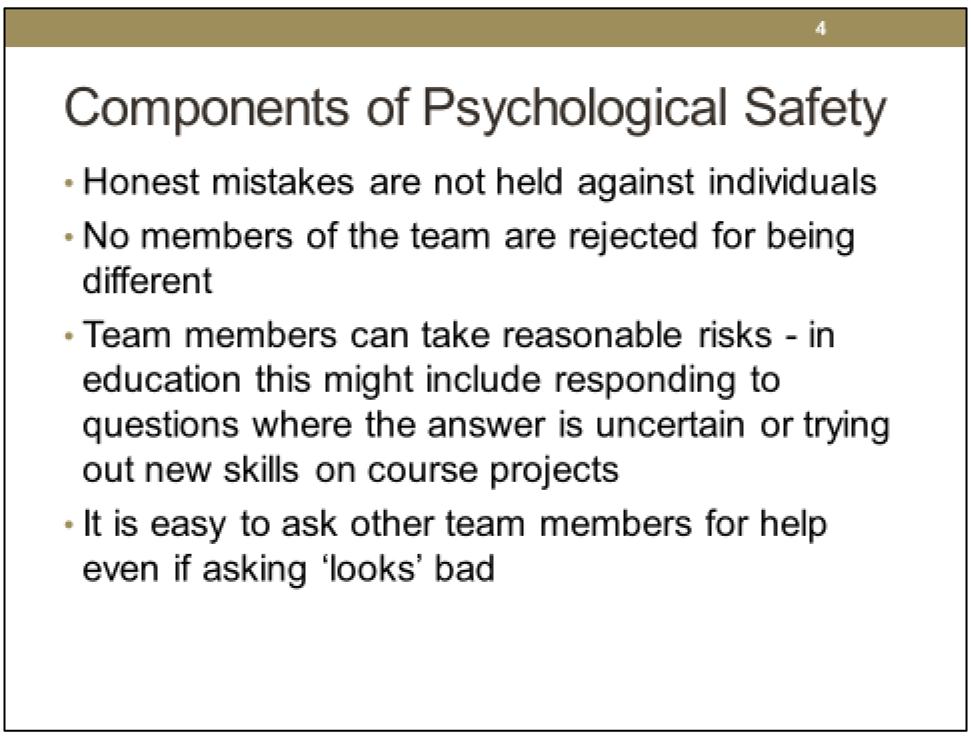
  - ***Time = 2 minutes***

2

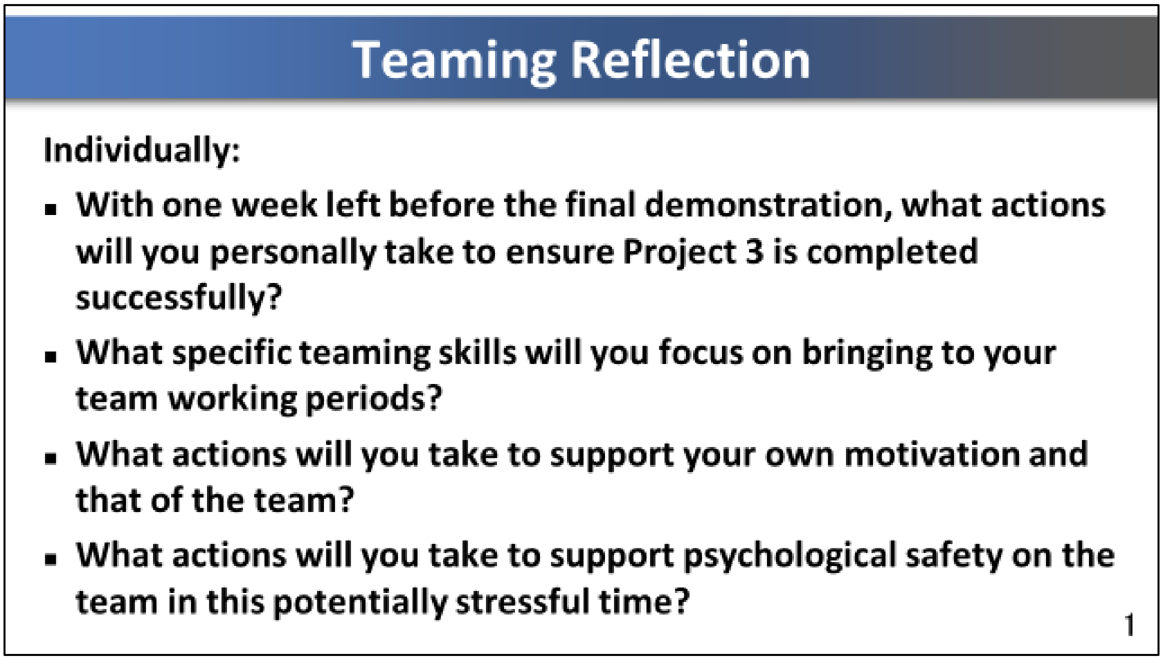
CATME Skills

- Interacting with Teammates
  - Listening, being polite, building understanding, being enthusiastic
- Contributing to the Team's Work
- Expecting Quality
  - High standards for work done by the team
- Keeping the Team on Track
- Having Knowledge, Skills, and Abilities
  - Including willingness to learn new skills





Appendix Figure C.12: ENGR131 Week 13 & ENGR141 Week 10 Teaming and Psych Safety Reflection Slides



Appendix Figure C.13: ENGR141 Week 15 Teaming and Psych Safety Reflection Slides

VITA

VITA

LEE K. RYNEARSON

EDUCATION

PhD in Engineering Education, Expected August 2016
Purdue University, West Lafayette, Indiana

Bachelor of Science & Master of Engineering in Mechanical Engineering, May 2008 [Dual-Degree]
The Rochester Institute of Technology, Rochester, New York

Doctoral Dissertation Title Promoting Teaming Metacognition
Committee Heidi Diefes-Dux (chair), Matthew Ohland, Morgan Hynes, Edward Berger

PROFESSIONAL TEACHING EXPERIENCE

Instructor of Engineering, January 2016 - Present
Campbell University, Buies Creek, North Carolina

Contributing to the development and implementation of the Campbell University School of Engineering in numerous capacities. Responsible for Foundations of Engineering course sequence.

Future Faculty Fellow, August 2014 – December 2016
Purdue University, West Lafayette, Indiana

As instructor of record for two honors First-Year Engineering courses (approx. 60 students per section), prepared and delivered lectures and lead in-class activities in engineering design, teaming, statistics, and other topics. Proposed substantial reforms to class methods and topics, collaborated with other course faculty to improve course curriculum & content. Designed and prepared all-new lectures and materials, including units on the mathematical modeling of physical systems, globally aware design, and applied brainstorming. Redesigned multiple existing lecture-based units into active in-class projects meeting original learning objectives while challenging teams of students to put technical content to use alongside design, project management, and technical communication skills.

Lead Graduate Teaching Assistant 2013-2014, Graduate Teaching Assistant 2012-2013
Purdue University, West Lafayette, Indiana

Lead a team of more than twenty undergraduates and four graduate students in the administration of an honors First-Year Engineering program while also lecturing on specific topics and supporting office hours. Completely reorganized course operations, training specialized materials development and validation teams and empowering graduate students to take ownership over elements of the course. Co-conceived, championed, and oversaw a transition from paper-based grading to a custom digital grading system, using automation to reduce grading errors, save time, and automatically provide instructional insights with learning analytics. Also developed numerous online learning modules, presented all class lectures on programming in LabVIEW, Python 3.3, MATLAB, and C, and supported office hours.

Project Education Center Research Associate, 2008 – 2012
Kanazawa Institute of Technology, Kanazawa, Japan

Provided classroom instruction on engineering design, teaming, estimation, and other topics in English in primary and supporting roles, collaborated in research and publication in the field of engineering education, developed new lecture and textbook materials, reviewed English-language research in engineering education on topics of interest to the Project Education Center and prepared condensed reports, provided language and editing support for faculty across dozens of research papers and other projects, and lead the effort improve a survey instrument used as an official benchmark of student skills growth by the Institute.

INDUSTRIAL EXPERIENCE

Technical Intern, June - November 2007
DuPont Engineering Mechanics, Wilmington, Delaware

Developed finite element models of new materials and novel structures under the guidance of experienced researchers. Worked with uncertain data and materials characterization to derive and communicate engineering understanding. Provided engineering simulation services for DuPont plant operations.

Technical Intern, June - August 2005, 2006
BAE Systems, Nashua, New Hampshire

Designed and built mechanical elements of prototype electronic warfare and countermeasures systems, from establishing needs through final assembly. Assisted in calibration of tools and field testing. Operated under demanding requirements with minimal oversight.

PROFESSIONAL AWARDS

Purdue University Graduate Teaching Excellence Award, 2015 (Purdue’s highest graduate teaching award)
Purdue University Teaching Academy Graduate Teaching Award, 2014
Purdue University College of Engineering Magoon Award for Excellence in Teaching, 2013
NSF Graduate Research Fellowship Honorable Mention, 2012

PROFESSIONAL QUALIFICATIONS

Purdue University Advanced Graduate Teaching Certificate, 2015
Fundamentals of Engineering Exam, New York State, 2008

WORKSHOPS & PROFESSIONAL DEVELOPMENT
Living with the Lab Faculty Training Workshop, Louisiana Tech University, 2016
Effective Teaching (presented by Drs. Felder & Brent) 2015
Improving First-Year Engineering Student Retention, Success, and Time to Graduation, ASEE 2015
New Paradigms on Teaching Engineering Ethics, ASEE 2015
Purdue Center for Instructional Excellence IMPACT Program Audit 2014
Making Academic Change Happen (MACH) 2014
Teaching Undergraduates for Learning Investment Program (TULIP) 2012 – 2016

PROFESSIONAL ENGAGEMENT
Purdue Student American Society for Engineering Education (ASEE) Chapter President Fall 2015-2016
Engineering Education Graduate Student Association Professional Development Chair Fall 2015-2016
Co-Presenter, Purdue University Teaching Assistant Orientation 2015
Purdue Student ASEE Chapter Classroom and Instructional Practices Chair 2014-2015
Reviewer, ASEE Conference & Exposition, First-Year Programs Division 2014 & 2015, Computers in Education Division 2015
Reviewer, Frontiers in Education (FIE) 2015
Reviewer, FYEE 2015
American Society for Engineering Education (ASEE) Member 2010 – Present
Reviewer, ASEE Conference & Exposition, First-Year Programs Division

PUBLICATIONS
Rynearson & Reazin (2015) Automation in Undergraduate Classes: Using Technology to Improve Grading Efficiency, Reliability, and Transparency in Larger Classes. *Transactions in STEM Education*, Volume 1, Issue 1, 42-53.
Rynearson & Hynes (2015) Design and Impact of a Classroom Intervention to Support Teaming Metacognition. *Frontiers in Education Conference 2015*. El Paso, TX.
Rynearson (2015) Balancing Authenticity with Scaffolding and Adding Alignment: Initial Reform of an FYE Student Team Design Project. *7th Annual First Year Engineering Experience*. Blacksburg, Virginia.
Rynearson & Reazin (2015) Automation in Undergraduate Classes: Using Technology to Improve Grading Efficiency, Reliability, and Transparency in Larger Classes. *ASEE Annual Conference & Exposition 2015*. Seattle, WA.
Rynearson (2014) Distinctive Academic Programs as a School Choice Factor. *ASEE Annual Conference & Exposition 2014.* Indianapolis, IN.
Koch, Borden, Berger, Brautigam, Culbertson, **Rynearson**, Siemens, & Wang (2012) Taking Action: A Proposal for an Analytic Solution to Increase Gateway Course

Success. *Instructional Development Center Publications.* http://docs.lib.purdue.edu/idcpubs/4

Rynearson & Matsuishi (2011) Improving A University-Wide Survey for Assessing Growth in Student Personal, Interpersonal, and Technical Skills. *Japan Society for Eng. Ed. Annual Conference 2011*. Sapporo, Japan.

Takechi & **Rynearson** (2010) Understanding Student Performance Diversity Using Clustering Techniques. *Joint IGIP-SEFI Annual Conference 2010*. Trnava, Slovakia.

Matsuishi, Nakamura, & **Rynearson** (2010) Efficient Skills Enhancement Through Continuing Education. *Joint IGIP-SEFI Annual Conference 2010*. Trnava, Slovakia.

Furukawa, Matsuishi, Matsumoto, Takemata, Yamakawa, & **Rynearson** (2010) Ways of Using and Characteristics of Two Team Contribution Peer-Evaluation Methods. *INNOVATIONS 2010: World Innovations in Engineering Education and Research.* International Network for Eng. Education & Research.

Ferro, Matsuishi, Liao, Tseng, Chang, Stamper, Sanders, Xu, White, **Rynearson**, Furukawa, Yamakawa, & Marui (2009) International Design Course with Institutions from Three Countries: First Year Review. *Japan Society for Engineering Education Annual Conference 2009.* Nagoya, Japan.

Nakamura, Matsuishi, & **Rynearson** (2009) Instructional Design to Enhance and Sustain Motivation for Learning in an Engineering *Course. IGIP Annual Conference 2009.* Austria.

*This journal publication is a re-printing of the ASEE conference paper of the same name.

COMPETITIVE GRANTS

Purdue College of Engineering Travel Grant for FYEE 2015, $500
Purdue Graduate Student Government Travel Grant for ASEE 2015, $1000
Purdue Graduate Student Government Travel Grant for FIE 2015, $1000
Purdue Graduate Student Government Grant for student ASEE chapter events, 2015 $4000
Purdue Graduate Student Government Grant for student ASEE chapter events, 2014 $1000
Purdue Graduate Student Government Travel Grant for ASEE 2014, $500
NSF Change Agent Grant for MACH Workshop 2014, $2,000

NOTABLE COURSEWORK

Educational Methods in Engineering
Assessment Methods in Engineering Education
Content, Assessment, and Pedagogy
Collaborative Learning
Program Evaluation
Leadership, Policy, and Change in STEM Education

ONLINE PROFILE

LinkedIn: https://www.linkedin.com/in/leerynearson